Proceedings of the 45th Annual

Symposium on Frequency Control

1991

Sponsored by the



The Institute of Electrical and Electronics Engineers, Inc.

Ultrasonics, Ferroelectrics, and Frequency Control Society

with the participation of personnel of the U. S. Army Electronics Technology and Devices Laboratory.

IEEE Catalog No. 91CH2965-2

Library of Congress No. 87-654207

29-31 May 1991

Los Angeles Airport Marriott Hotel
Los Angeles, California

Best
Available
Copy

ADDITIONAL NOTE TO DTIC for CITATIONS for the Proceedings of the Annual Frequency Control Symposia

Suggested Descriptors:

accelerometers
aging
Allan variance
atomic clocks
atomic frequency standards
atomic resonators
cesium
cesium standards
clocks
crystals
etching
filters
frequency control
frequency standards
global positioning system
GPS
hydrogen masers
noise(electrical and electromagnetic)
oscillators
phase noise
quartz
quartz crystals
quartz oscillators
quartz resonators
radiation effects
resonators
rubidium
rubidium standards
sensors
stability
surface acoustic wave devices
synthesizers
time
timekeeping
time transfer
transducers
vibration effects

Abstract: Main headings in the Table(s) of Contents are good to use for Abstract data.

Abstracting is permitted with the credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 21 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for non-commercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 345 E. 47th Street, New York, NY 10017. All rights reserved. Copyright © 1991 by the Institute of Electrical and Electronics Engineers, Inc.

45TH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SYMPOSIUM CHAIRMEN

General Chairman.....Raymond L. Filler

Technical Program Chairman.....Thomas E. Parker

Local Arrangements Chairman.....Victor Reinhardt

Editorial Chairman.....John R. Vig

Publicity Chairman.....S. Clark Wardrip

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Materials & Resonators: Robert Kinsman, Motorola, Inc.

Oscillators & Circuits: Michael Driscoll, Westinghouse Systems Center

Atomic Standards & Timing: Richard Sydnor, Jet Propulsion Laboratory

TECHNICAL PROGRAM COMMITTEE

David Allan, NIST
Arthur Ballato, U.S. Army LABCOM
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, NRC, Canada
Jan Brown, Fischer Controls Int'l., Inc.
Mary Chiu, Johns Hopkins University, APL
Leonard Cutler, Hewlett-Packard Company
Michael Driscoll, Westinghouse Systems Center
Errol EerNisse, Quartztronics, Inc.
Raymond Filler, U.S. Army LABCOM
Marvin Frerking, Rockwell International
Satoru Fujishima, Murata Mfg. Co., Japan
Jean-Jacques Gagnepain, CNRS-LPMO, France
Edward Garber, TRW
Michael Garvey, Frequency & Time Systems, Inc.
Helmut Hellwig, Air Force Office of Scientific Research
James Pon-Wei Hou, Quartz Freq. Tech. Ltd., Taiwan
Charles Jensik, Piezo Crystal Company
Tom Jerse, Hewlett Packard Company
Gary Johnson, Sawyer Research Products, Inc.
Robert Kinsman, Motorola, Inc.
William Klepczynski, U.S. Naval Observatory
Jack Kusters, Hewlett Packard Company
Theodore Lukaszek, U.S. Army LABCOM
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Joel Martin, Oklahoma State University
Thrygve Meeker, Consultant
Thomas Parker, Raytheon Research Division
Victor Reinhardt, Hughes Aircraft Company
Gerald Roberts, General Electric Company
Tadashi Shiosaki, Kyoto University, Japan
Robert Smythe, Piezo Technology, Inc.
Samuel Stein, Ball Corporation
Dan Stevens, AT&T Bell Laboratories
Richard Sydnor, Jet Propulsion Laboratory
John Vig, U.S. Army LABCOM
S. Clark Wardrip, Bendix Field Engineering Corp.
Rolf Weglein, Consultant
Joseph White, Naval Research Laboratory
Roger Williamson, STC Technology, United Kingdom
Gernot Winkler, U.S. Naval Observatory
Nicholas Yannoni, U.S. Air Force/RADC

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
Thomas E. Parker, Raytheon Research Division

PIEZOELECTRIC SENSORS & NONLINEAR EFFECTS
Robert Kinsman, Motorola, Inc.

RF AND MICROWAVE OSCILLATORS
Rolf Weglein, Consultant

RESONATOR PROCESSING
Charles Adams, Hewlett-Packard Company

OPTICAL FREQUENCY STANDARDS
Lute Maleki, Jet Propulsion Laboratory

BAW RESONATORS AND SAW DEVICES
Theodore Lukaszek, U.S. Army LABCOM

ADVANCED MICROWAVE ATOMIC FREQUENCY STANDARDS
Andrea DeMarchi, Politecnico di Torino, Italy

OVENIZED AND TEMPERATURE COMPENSATED OSCILLATORS
John R. Vig, U.S. Army LABCOM

FREQUENCY STANDARDS INCORPORATING LASERS
William Klepczynski, U.S. Naval Observatory

CRYSTAL GROWTH AND EVALUATION
Gary Johnson, Sawyer Research Products, Inc.

HIGH TEMPERATURE SUPERCONDUCTING RESONATORS
Leonard Cutler, Hewlett-Packard Company

POSTER SESSION
Grant Moulton, Hewlett-Packard Company

RESONATOR DESIGN
Arthur Ballato, U.S. Army LABCOM

SUPERCONDUCTING RESONATORS AND OSCILLATORS
Daniel Oates, MIT Lincoln Laboratory

ACCELERATION SENSITIVITY & FILTERS
Raymond L. Filler, U.S. Army LABCOM

PHASE NOISE & FREQUENCY STANDARDS
Gary Montress, Raytheon Research Division

OSCILLATOR CIRCUIT ANALYSIS
Michael M. Driscoll, Westinghouse Electric Company

NOISE, CLOCKS & FREQUENCY STANDARDS
Samuel R. Stein, Ball Communications Systems Division

TABLE OF CONTENTS

PLENARY SESSION — **PAGE**

Award Presentations 1

PIEZOELECTRIC CRYSTALS

Movement of Dislocations in Quartz 2
R.A. Murray, US Army LABCOM

An Analytical Study of Electrodiffusion (Sweeping) of Synthetic Quartz Crystals..... 9
J. Asahara, K. Nagai, K. Hamaguchi, H. Sone & S. Taki, Toyo Communication Equipment Co., Ltd., Japan

Thermal Expansion of Alpha Quartz..... 22
J.A. Kosinski, J.G. Gualtieri & A. Ballato, US Army LABCOM

Influence of Convective Flows on the Growth of Synthetic Quartz Crystals..... 29
V.A. Klipov & N.N. Shmakov, Research Department of FONON, USSR

Sweeping and Irradiation Effects on Hydroxyl Defects in Crystalline Quartz 37
H. Bahadur, National Physical Laboratory, India

Assessment of Material Perfection and Observation of Vibration Modes in Lithium Tantalate by X-Ray Topography 58
A. Zarka, B. Capelle & Y. Zheng, Université P. & M. Curie; J. Détaint, J. Schwartzel & C. Joly, CNET; D. Cochet-Muchy, Crismatec, France

Evaluation of Nonlinear Elastic Coefficient Causing Frequency Shifts in AT-Cut Resonators..... 72
Y. Hirose & Y. Tsuzuki, Yokohama National University, Japan

RESONATORS - BAW, SAW, THIN FILM, FILTERS

The Aging of Bulk Acoustic Wave Resonators, Filters and Oscillators **(Invited)** 77
J.R. Vig, US Army LABCOM & T.R. Meeker, Consultant

Influence of Surface Catalytic Reactions on Crystal Unit Aging..... 102
P.I. Aseev, V.B. Grouzinenko & A. Rodichev, Scientific-Research Institute "Phonon", USSR

Bulk Wave Quartz Resonators Fabricated by a Hollow Cathode RF Plasma Etching Technique..... 106
S. Schreiter, Technical University of Chemnitz, Germany & D.Y. Yankov, Institute of Solid State Physics, Bulgaria

Automated X-Ray Sorting Machine for Round Quartz Blanks 114
H. Bradaczek, Freie Universität Berlin, Germany

A Very Low Profile, Shock-Resistant, Surface-Mountable Quartz Crystal Unit..... 117
R.C. Smythe, Piezo Technology Inc.; J. Greenstein, Coleman Research Corporation; J. Lipeles, International Dynamics Corporation

Suppression of the Unwanted Response of an SC-Cut Crystal Unit 124
M. Koyama, S. Mizuno & M. Okazaki, Nihon Dempa Kogyo Co., Ltd. Japan

TT-Cut Torsional Quartz Crystal Resonator 130
H. Kawashima & M. Nakazato, Seiko Electronic Components Ltd., Japan

Thickness-Shear Mode Shapes & Mass-Frequency Influence Surface of a Circular and Electroded AT-Cut Quartz Resonator 137
Y-K.Yong, J.T. Stewart, Rutgers University; J. Détaint, CNET; A. Zarka, N. Capelle & Y. Zheng, Université P. & M. Curie, France

A Simple Model for Quartz Resonator Low Level Drive Sensitivity 148
L. Dworsky & R. Kinsman, Motorola

Two-Dimensional Equations for Guided EM Waves in Dielectric Plates Surrounded by Free Space 156
P.C.Y. Lee & J.S. Yang, Princeton University

Optimization of AT Berlinite & Quartz Thickness Shear Devices for VHF Filter Applications 166
J. Détaint, H. Carru, J. Schwartzel, C. Joly, CNET; B. Capelle, A. Zarka & Y. Zheng, Université P. & M. Curie; E. Philippot, Université du Languedoc, France

Monolithic Filters Using Strong Piezoelectrics 181
S.A. Sakharov, I.M. Larionov & V.A. Issaev, Scientific-Research Inst., "PHONON", USSR

Inflection Temperature of Contoured AT-Cut Resonators 184
J. Dutka, Crystek Corp., J. Erasmus, Consultant & J.H. Sherman, Jr., Consultant

Load Resonant Measurements of Quartz Crystals 191
D. Rose, Saunders & Associates, Inc.

Fundamental Properties of Thin Film Resonators 201
K.M. Lakin, TFR Technologies, Inc.

High Frequency Oscillators Using Cointegrated BAW Thin-Film Piezoelectrics with Microwave BJTS 207
S.G. Burns, R.J. Weber & S.D. Braymen, Iowa State University

Piezoelectric Resonators Based on Single Crystals of Strong Piezoelectrics 212
V.B. Grouzinenko & V.V. Bezdelkin, Scientific-Research Institute "PHONON", USSR

A Comparison of Predicted and Measured Properties of Doubly-Rotated Lithium Tetraborate Resonators 217
A. Ballato, J. Kosinski & T. Lukaszek, US Army LABCOM; S. Mallikarjun, Monmouth College

An Analysis of the Influence of Design Parameters on the Resonant Frequency and Q-Factor of Surface Transverse Wave (STW) Resonators 222
E. Bigler, CNRS, France; B.A. Auld, Stanford University; E. Ritz & E. Sang, Tektronix

1 GHz Voltage Controlled Oscillator Using a SSBW/BGW Combined Mode Resonator Filter 230
I.D. Avramov, Institute of Solid State Physics, Bulgaria

A High Performance Waveguide-Coupled Resonator SAW VCO 239
J.W. Andersen, Sawtek Inc.

Grating Assisted SAW Directional Couplers 247
G. Golan, I. Charuvy, M. Zilberstein, A. Seidman & N. Croitoru, Tel Aviv University, Israel

PIEZOELECTRIC SENSORS

Quartz Resonator Sensors in Extreme Environments (Invited) 254
E.P. EerNisse & R.W. Ward, Quartztronics, Inc.

Piezoelectric Vibratory Gyroscope Using Flexural Vibration of a Triangular Bar 261
S. Fujishima, T. Nakamura & K. Fujimoto, Murata Mfg. Co., Ltd., Japan

Solder Bond Applications in a Piezoelectric Sensor Assembly 266
P.T. Vianco & J.A. Rejent, Sandia National Laboratories

ACCELERATION SENSITIVITY

Acceleration Effects in Crystal Filters -- A Tutorial 281
R.C. Smythe, Piezo Technology Inc.

The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Due to its Thickness Asymmetry 289
H.F. Tiersten & Y.S. Zhou, Rensselaer Polytechnic Institute

An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips 298
Y.S. Zhou & H.F. Tiersten, Rensselaer Polytechnic Institute

Frequency Shift Calculations for Quartz Resonators 309
L.D. Clayton & E.P. EerNisse, Quartztronics Inc.

SAW Oscillators with Low Vibration Sensitivity 321
T.E. Parker, J.A. Greer & G.K. Montress, Raytheon Research Division

A Low Noise Vibration Isolated Airborne Radar Synthesizer 330
A. Vulcan & M. Bloch, Frequency Electronics, Inc.

CRYSTAL OSCILLATORS

Computation of Oscillator Open-Loop Bode Plots 336
D.E. Phillips, Rockwell International Corporation

High Precision Nonlinear Computer Modelling Technique for Quartz Crystal Oscillators 341
R. Brendel, F. Djian, CNRS & E. Robert, CNES, France

Oscillator Insights Based on Circuit Q 352
R.L. Clark, Penn State University

Analysis of Start-Up Characteristics of Crystal Oscillators 360
J-Q. Lu, Sony LSI Design Corporation & Y. Tsuzuki, Yokohama National University, Japan

Oscillator Design Using S Parameters and a Predetermined Source or Load 364
R.J. Weber, Iowa State University

Universal, Computer Facilitated, Steady State Oscillator, Closed Loop Analysis Theory and Some UHF and Microwave Applications 368
B. Parzen, Consultant

Quartz Crystal Oscillators with Direct Resonator Heating 384
B. Long & G. Weaver, Piezo Crystal Company

Performance Tests on an MCXO Combining ASIC and Hybrid Construction 393
A. Benjaminson, General Technical Services, Inc. & B. Rose, Q-Tech Corporation

Frequency-Temperature Considerations for Digital Temperature Compensation 398
R.L. Filler, US Army LABCOM

Digital TCXO Using Delta Modulation 405
M. Watanabe, Nihon Dempa Kogyo Co., Ltd.; Y. Sakuta & Y. Sekine, Nihon University, Japan

Temperature Compensated Crystal Oscillator Employing New Shape GT Cut Quartz Crystal Resonator 410
H. Kawashima & K. Sunaga, Seiko Electronic Components Ltd., Japan

Voltage Controlled Temperature Compensated Crystal Oscillator Using 2-Port Crystal Resonator 418
Y. Ueno, Nihon Dempa Kogyo Co. Ltd. & H. Shimizu, University of Electro-Communications, Japan

BVA-Type Quartz Oscillator for Spacecraft 426
J.R. Norton, Johns Hopkins University

A Space Oscillator with Cylindrical Oven and Symmetry 431
M. Mourey & R.J. Besson, Ecole Nationale Supérieure de Mécanique et des Microtechniques, France

A Space-Qualified Frequency Synthesizer 442
M.J. Reinhart, Johns Hopkins University

High-Stability Quartz Oscillators on Internally-Heated Quartz Resonators with AT and SC Cuts 447
Y.L. Vorokhovsky & B.G. Drakhlis, MORION Mfg. Co., USSR

SUPERCONDUCTING AND LOW TEMPERATURE RESONATORS AND OSCILLATORS

High-Temperature Superconducting Resonators **(Invited)** 452
J.N. Hollenhorst, R.C. Taber, L.S. Cutler, Hewlett-Packard Laboratories; T.L. Bagwell, Hewlett-Packard; N. Newman, Conductus, Inc.

Superconducting Thin-Film $YBa_2Cu_3O_{7-X}$ Resonators and Filters 460
D.E. Oates, W.G. Lyons & A.C. Anderson, Massachusetts Institute of Technology

A High-T_c Superconducting Resonator for a Compact Hydrogen Maser 467
D.B. Opie, H.E. Schone, College of William & Mary; M. Hein, G. Müller, H. Piel, H.-P. Schneider, University of Wuppertal, Germany; V. Folen, A. Frank, W.M. Golding & S. Wolf, Naval Research Laboratory

High-Tc Superconducting Microstrip Resonators and Filters on $LaAlO_3$ 477
J.H. Takemoto, C.M. Jackson, R. Hu, H.M. Manasevit, C. Petite-Hall, J.F. Burch, D.C. St. John, A. Lee & K.P. Daly, TRW Space and Technololgy Group

Superconducting Stabilized X-Band Oscillators 482
R. Cadotte, A. Rachlin, R. Babbit & W. Wilber, US Army LABCOM

Combined Stability of Hydrogen Maser and Superconducting Cavity Maser Oscillator 491
R.T. Wang & G.J. Dick, Jet Propulsion Laboratory

Analysis of a Low Noise Tunable Oscillator Based on a Tunable Sapphire Loaded Superconducting Cavity 495
M.E. Tobar & D.G. Blair, University of Western Australia

LASERS AND THEIR APPLICATIONS, OPTICALLY PUMPED STANDARDS

Laser Frequency Stability Requirements for Coherent Space Communications 500
C-C. Chen & M.Z. Win, Jet Propulsion Laboratory

Stabilization of Oscillator Phase Using a Fiber-Optic Delay-Line 508
R.T. Logan, Jr., L. Maleki & M. Shadaram, Jet Propulsion Laboratory

Prototype of an Optically Pumped Cesium Beam Frequency Standard 513
K. Hisadome & M. Kihara, NTT Transmission Systems Laboratories, Japan

Miniature Optically Pumped Cesium Standards 521
L.L. Lewis, Ball Communications Systems Division

Single Ion Optical Frequency Standard **(Invited)** 534
J.C. Bergquist, W.M. Itano, D.J. Wineland, F. Diedrich, F.Elsner & M.G. Raizen, NIST

Intensity and Frequency Noise in Semiconductor Lasers **(Invited)** 539
K.J. Vahala, California Institute of Technology

A Survey of Optical Techniques with an Emphasis on Frequency Control (Phase Stability) Aspects **(Invited)** 540
H.T.M. Wang, Hughes Research Laboratories

ATOMIC FREQUENCY STANDARDS AND PRECISION TIMING

Frequency Pulling in Cesium Beam Frequency Standards Due to Delta M = $\pm$ 1 (SIGMA) Transitions 544
L.S. Cutler, C.A. Flory & R.P. Giffard, Hewlett-Packard Laboratories; A. DeMarchi, Politecnico di Torino, Italy

A New High-Performance Cesium Beam Tube Compensated for Ramsey Pulling 554
L.F. Mueller, D.B. Dull & J.A. Kusters, Hewlett-Packard Company

Prospects in Improvement of Accuracy Performances of Passive Atomic Frequency Standards 562
A.G. Gevorkyan & V.S. Zholnerov, Leningrad Scientific Research Radiotechnical Institute, USSR

Compact Rectangular-Cylindrical Cavity for Rubidium Frequency Standard 567
E. Eltsufin, A. Stern & S. Fel, Time & Frequency Ltd., Israel

Ultra-Stable Hg^+ Trapped Ion Frequency Standard 572
J.D. Prestage, R.L. Tjoelker, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Hydrogen Maser Performance at the United States Naval Observatory and the Naval Research Laboratory 582
E. Powers, A. Gifford, Naval Research Laboratory & P. Wheeler, United States Naval Observatory

Barometric Pressure-Induced Frequency Offsets in Hydrogen Masers 586
R.A. Dragonette & J.J. Suter, Johns Hopkins University

Atomic Frequency Standards for Satellite Radionavigation Systems 591
Y.G. Gouzhva, A.G. Gevorkyan & V.V. Korniyenko, Leningrad Scientific Research Radiotechnical Institute, USSR

Deep Space Tracking and Frequency Standards **(Invited)** 594
J.S. Border & E.R. Kursinski, Jet Propulsion Laboratory

Interim Report on the Geographic Dependency and Latitude Effects Study 608
R. Beard, A. Gifford & S. Stebbins, U.S. Naval Research Laboratory; S. Rasmussen, Dept. of Defense; T. Bartholomew, The Analytical Sciences Corporation

Automatic Hierarchy Control Over Stratum 1 Redundant Time & Frequency Systems 626
E. Peled, M. Zelitzki & A. Nemesh, Time & Frequeny Ltd., Israel

NOISE, FREQUENCY STABILITY, LOW NOISE SOURCES

Phase Noise Modeling in Frequency Dividers **(Invited)** 629
W.F. Egan, ESL, Inc.

High Spectral Purity Frequency Sources Using Low Noise Regenerative Frequency Dividers 636
J. Groslambert, M. Olivier, CNRS, France; E. Rubiola, Politechnico di Torino, Italy

Effects of Cavity Q and Coupling on Oscillator Noise Measurements 640
C. Cook, Johns Hopkins University

Ultralinear Small-Angle Phase Modulator 645
J. Lowe & F.L. Walls, NIST

Stochastic Resonance: A Signal + Noise in a Two State System **(Invited)** 649
F. Moss, University of Missouri at St. Louis

About Sum of Signals and its Possible Use in Time & Frequency Systems 659
A. Stern, Time & Frequency Ltd., Israel

A Frequency-Domain View of Time-Domain Characterization of Clocks and Time and Frequency Distribution Systems 667
D.W. Allan, M.A. Weiss & J.L. Jespersen, NIST

Status of Frequency and Timing Reference Signal Transmission by Fiber Optics 679
G.F. Lutes & R.T. Logan, Jet Propulsion Laboratory

Determination of the Limiting Factors in the Absolute Phase Noise of an L-Band Dielectric Resonator Oscillator 687
M. Mizan, R.C. McGowan, T. Lukaszek & A. Ballato, US Army LABCOM

Ultra-Low-Noise, 8.3 GHz Dielectric Resonator Oscillator 693
M. Mizan & R. McGowan, US Army LABCOM

LATE SUBMISSION PAPERS

Cooled, Ultra-High Q, Sapphire Dielectric Resonators for Low Noise, Microwave Signal Generation 700
M.M. Driscoll, J.T. Haynes, S.S. Horwitz, R.A. Jelen, Westinghouse Electronics Systems Group; R.W. Weinert, J.R. Gavaler, J. Talvacchio, G.R. Wagner, Westinghouse Science & Technology Center; K.A. Zaki & X-P. Liang, University of Maryland

Frequency and Timing Requirements for Space Exploration **(Invited)** 707
L. Maleki, Jet Propulsion Laboratory

A Wide Range (550-700MHz) Monolithic Timing Recovery Circuit 712
E.M. Frymoyer & B. Lai, Hewlett Packard Company

AUTHOR INDEX 716

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL 718

PROCEEDINGS AVAILABILITY INFORMATION 721

1991 AWARD WINNERS

The Cady Award

The Cady Award was presented to Robert C. Smythe, Piezo Technology, Inc., "for contributions to the development of single and dual mode quartz resonators for filter applications." The award was presented by Harry F. Tiersten, Rensselaer Polytechnic Institute.

The Rabi Award

The Rabi Award was presented to Andrea DeMarchi, Politecnico di Torino, "for contributions to significant improvements in the accuracy and stability of cesium beam frequency standards." The award was presented by David W. Allan, National Institute of Standards and Technology.

The Sawyer Award

The Sawyer Award was presented to John G. Gualtieri, U. S. Army Electronics Technology and Devices Laboratory, "for significant contributions in further understanding alpha quartz as regards to sweeping, etch channels, irradiation, and developing high yields in photolithographic processing." The award was presented by Joel J. Martin, Oklahoma State University.



Andrea DeMarchi, Rabi Award winner; Robert C. Smythe, Cady Award winner; and John G. Gualtieri , Sawyer Award winner; after the award presentations.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

MOVEMENT OF DISLOCATIONS IN QUARTZ

R. A. Murray

US Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5601

Abstract

Dislocations in quartz crystals have been known to cause problems in the fabrication of resonators by the formation of etch channels. In addition, it is suspected that they contribute to acceleration sensitivity, thermal hysteresis, and possibly aging. The behavior of dislocations in quartz is also of interest to the fields of geophysics, seismology, and plate tectonics because it affects the underground movement of rock. This paper is a review of the literature, including geophysical work, and its implications for frequency control.

Introduction

Fundamental studies on dislocations in quartz have not been a high priority in the frequency control community. Early work indicated that dislocation densities found in cultured quartz did not adversely affect resonator performance, and since then, most of the work has been focused on how dislocations affect the processing of resonators. Etch channels and pits, which result when quartz containing decorated dislocations is chemically etched or polished, decrease the mechanical strength of the resonators and lower the yields in manufacturing processes. However, as the requirements for resonator performance becomes more stringent, the presence of dislocations in resonators is likely to become more significant. Recent work has shown that etch channels dramatically reduce the Q of high frequency bulk wave resonators. It has also been shown that the vibrational mode shape can affect the acceleration sensitivity of resonators, and X-ray topographs clearly show that mode shapes are distorted by dislocations. Thermal hysteresis and aging may also be affected by the presence of dislocations.

The properties of dislocations in quartz are also of interest to researchers in the fields of geology, seismology, and plate tectonics. When quartz is subjected to high stress and temperature, it flows (plastically deforms) through a mechanism involving the movement of dislocations. It has been reported that the ease with which these dislocations move is linked to the amount of water in the crystal. This effect has been studied extensively over the past twenty-five years and a large body of literature has been published in the geophysical and mineralogical journals.

The current interest in UHF resonators, low acceleration sensitivity resonators, low hysteresis and low aging resonators, and the relatively untapped geophysical work make this an ideal time for a review paper on dislocations. In this paper we discuss historical background and current theories of dislocations in quartz. Some of the geophysical findings may be used to help explain results reported in the frequency control community.

Observing Dislocations

A dislocation is defined as a linear lattice defect. An edge dislocation is formed by inserting an extra plane (or planes) of atoms into an existing lattice. A screw dislocation is generated by displacing the lattice on one side of the dislocation line relative to the other.[1] The Burgers vector states how much and in what direction a dislocation distorts the crystal lattice. The Burgers vector of an edge dislocation is normal to the dislocation line while for a screw dislocation it is parallel.[1]

There are many ways to observe dislocations in quartz, some direct and some indirect. The most common technique in the frequency control community is to use etching in order to produce etch channels.[2,3,4] In this method a chemical etchant, usually HF, ammonium bifluoride, or a buffered oxide etchant (i.e. various mixtures of $NH_4F{\cdot}HF$ and HF) is used to etch the quartz.

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

The etch channel density is then counted (see standard EIA-477-A[5]). This method is indirect and insufficient in that the etch channels only form if the dislocations are decorated with impurities,[4,6] so not all of the dislocations show up as etch channels. Sweeping moves the impurities away from the dislocation sites, and for this reason, etching will not work at all for swept quartz. Another technique, used by Moriya and Ogawa[7], is light scattering tomography. This method scans the quartz sample with a low power (in this case a 3 mW) He-Ne laser. The scattered light is then recorded on film. This apparatus is sensitive enough to observe both decorated and undecorated dislocations in quartz samples and it is non-destructive. Other methods include transmission electron microscopy (TEM)[8,9,10] and X-ray topography.[3,11] A recent improvement on standard X-ray topography is the use of synchrotron X-rays to attain very high resolution and high speed (1 ns) pictures of dislocations and strain fields.[12,13,14] This is especially useful for looking at the effects of dislocations on active resonators. If the quartz has been deformed sufficiently, as happens in some of the geophysical experiments, the bands of dislocations can be seen with a polarizing microscope.[15]

Impact of Dislocations on Resonators

The questions of real interest for the frequency control community is how dislocations affect device fabrication and performance. First, consider the fabrication process. As mentioned earlier, when quartz containing decorated dislocations is etched, channels form. In HF, etch channels about 1 micron in diameter propagate from the surface at more than 100 micron/hour. This is much faster than the bulk etching rate.[11] One way to reduce the number of etch channels that form during chemical polishing and etching is to use swept quartz. However, this does not reduce the dislocation density. Hanson[16] took X-ray topographs before and after sweeping quartz samples and no decrease was seen in the dislocations.

The next concern is the effect of dislocations on performance. A resonator property that is often measured and reported is Q. In the past it was commonly accepted that dislocations did not affect resonator Q. In 1982, Iwasaki and Kurashige[11] fabricated 5 MHz AT-cut resonators and found no correlation between dislocations and mechanical Q for quartz blanks with dislocation densities up to $10^3/cm^2$. In a more extensive study, Meeker and Miller[17] also looked for a correlation between the dislocation density and the performance characteristics of 8 MHz fundamental AT-cut resonators cut from r-bars. They found that not only were the Q's independent of the dislocation density, but so were the resistance and inductance of the resonators. However, they did correlate the grown-in strain associated with dislocations to abnormal f-T behavior. That is, the dislocations were responsible for apparent angle shifts in resonators of up to several minutes of arc. In a recent article, Kent[18] correlated Q degradation with etch channels in high frequency (45 to 140 MHz) bulk wave resonators which are chemically etched as part of the fabrication process. In these very thin resonators a single etch channel in the active area reduced the Q by 46%. For these resonators, Kent states that low etch channel quartz (either as grown or swept) is necessary to achieve consistent Q values. Since the dislocation density is unaffected by the sweeping process, it would appear that for these resonators it is the presence of the etch channels rather than the dislocations that degrades the Q.

The apparent angle shifts in resonators, found by Meeker and Miller[17] to be correlated with dislocations, suggests that dislocations can change the material constants in quartz. In support of this hypothesis, James[19] stated that he found significant differences in the elastic moduli in quartz of different grades. The difference between the grades was the level of Al impurities and the dislocation densities ($10/cm^2$ for the higher grade vs. 10^3 to $10^4/cm^2$ for the lower). Such changes in the elastic moduli would be likely to show up as deviations in the f-T characteristics (apparent angle offsets).

Recent attempts to make resonators less sensitive to acceleration have not been uniformly successful. One reason for this seems to be that the vibrational modes are not the correct shape or in the correct place to cancel the acceleration frequency offset.[20] Several groups have studied the mode shapes in active resonators using standard and stroboscopic x-ray topography.[13,14,21] They have found that the presence of dislocations in the active regions of the resonators distorts the mode shape, enhances coupling among modes, and interferes with mode trapping. This means that very low dislocation quartz is necessary for fabricating acceleration insensitive resonators.

Finally, consider the problems of hysteresis and aging. If dislocations change their positions during the life of the device, this could contribute to aging. There is no direct evidence that this happens[22], but Glüer et al.[12] saw apparent dislocation motion in stroboscopic X-ray topographs of a vibrating AT-cut resonator. They dismissed this as an illusion caused by the vibration

because they didn't believe the dislocations could be mobile at the stress levels present. In a related vein, Stephenson[23] used electric potentials at elevated temperatures (near the alpha-beta transition) to induce dislocations in nearly perfect quartz. He states that this type of accelerated aging can anticipate a type of defect that might show up, in time, in resonators. If this happens gradually over months (and years) it could contribute to aging. Could dislocation movement also be a factor in hysteresis? As the temperature is increased and decreased, the dislocations may move to different potential minima, changing the crystal constants. Beaussier[24] has put forth the proposition that slight plastic deformations, associated with the movement of dislocations, can cause an intrinsic hysteresis in bulk quartz.

Characteristics of Dislocations in Quartz

In cultured quartz it has been observed that dislocations usually originate in the seed or at inclusions within the growth zones. It has also been observed that these dislocations propagate nearly perpendicularly to the growth front in each of the growth regions (+x, -x, s and z).[11] The angular distribution of dislocations in the growth zones was used by Alter and Voigt[25] to show how the directions change when dislocations pass growth boundaries. They cut blanks from -x, +x, z, r, and R growth sectors and used them as seeds. The degree to which the dislocations would deviate from the growth front normal was predicted by an elastic energy minimization per length model. The dislocations on the +x seeds were nearly parallel to the seed. Meeker and Miller[17] found two types of dislocations in cultured r bars. The first was uniform and slow etching. After these dislocations had been etched into etch channels, there was no disruption of f-T curves. The second type were clustered, etched faster, and changed the f-T curves. The dislocations normal found in the z region of quartz bars are thought to be edge type and those in the +x and -x regions are screw type.[11,26] Given that Meeker and Miller used r bars, and not z cut seeds, the two types of dislocations they observed may be edge and screw type. It is not clear, however, why these dislocations should behave so differently. Iwasaki[3] proposed two atomic models for the edge dislocations in quartz z growth but the Burgers vectors of the models do not fit those determined recently by Epelboin and Patel.[27] X-ray topographs of Barns[26] were examined by Epelboin and Patel using a computer simulation method. This allowed them to determine that two different Burgers vectors are present in the Z region dislocation. These are

$$[1\bar{2}10],\quad [\bar{2}110]\ .$$

The first Burgers vector is the most prevalent. An example of an edge dislocation with this vector is shown in Fig. 1. Note that the O atoms have not been drawn, only the Si.

Over the years various methods have been tried to reduce the dislocation density in quartz. Doherty et al.[28] attempted to grow high purity quartz. Their aim was low Al content and low dislocation density. By using very pure fused silica as the nutrient, gold lined vessels, and growing very slowly, (0.16 to 0.35 mm/side/day) the Al content was kept well below 1 ppm (0.01 to 0.2 ppm Al) and the number of dislocations introduced during growth was reduced. To reduce the number of dislocations originating from the seed, the seeds were cut from high quality natural quartz which had been inspected by X-ray topography for dislocations. Using these methods, the dislocation density of the grown bars was kept to less than $10/cm^2$. This material also had radiation induced offsets of less than 10^{-13} Hz/rad. Armington et al.[29,30,31] have made an extensive study of ways to reduce the number of inclusions and dislocations in cultured

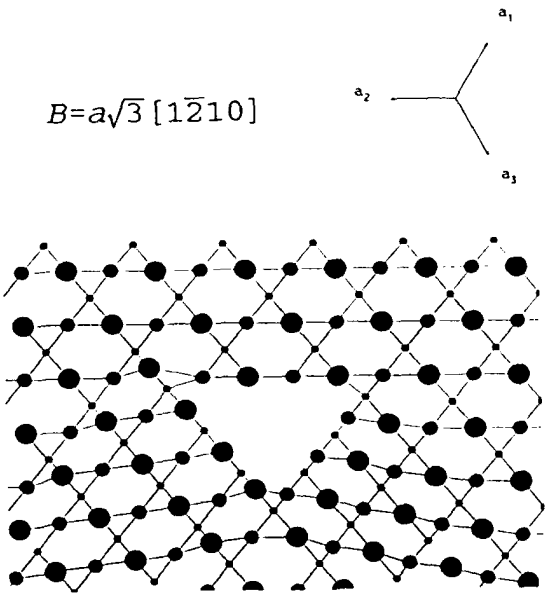


Figure 1 Edge dislocation in quartz. Note that only the Si atoms have been drawn, no O atoms are shown.

quartz. Some of the methods which did not improve the quartz quality were sweeping the seeds before growth, doping the seeds to change the lattice dimensions, and annealing the quartz bars. In fact, the annealing seemed to increase the strain around the inclusions. One method that did decrease the dislocation density in the grown bars was growing on seeds cut from the +x region. Armington's data was inconclusive on the effect of silver liners on dislocations. Irvine et al.[32] at Sawyer Research also had mixed results for silver liners. The typical Sawyer quartz bar grown in a silver liner had 86 dislocations/cm^2, showing that simply using inert liners does not automatically abolish dislocations. Another aspect of dislocation effects was detected when Armington analyzed the material swept from quartz. They found, among other things, Al (this had also been reported by Gualtieri and Vig[33]). Armington speculated that the Al was swept along dislocations since it is improbable that it would be moving through the c (z) channels.

The properties of quartz, both mono-crystalline and as aggregates, have long been of interest to geologists and geophysicists. This is partly because of the abundance of this material in the earth and partly because of its unusual properties. One of the more important aspects of these is how quartz behaves at high stress levels. The long term strengths and flow properties of

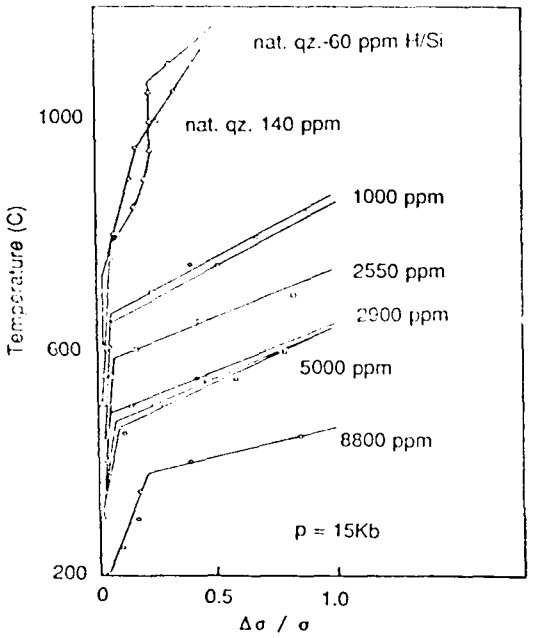


Figure 2 Fractional stress drop vs. temperature for natural and cultured quartz.

rock is of major importance to geophysicists because it is critical to the understanding of e.g. earthquake seismology.[34] The goal of geophysical research on the high temperature deformation of rocks is to determine the stress and strain history of the Earth's crust and mantle. This will allow accurate modeling of rock movement.[35] When quartz is permanently deformed by stress, it is done so through the movement of dislocations. Geophysicists are interested in using dislocation density to infer the stresses experienced by the samples while underground.

Natural quartz is unusual in the sense that the combination of high strength combined with low elastic constants means that very large elastic deformation of single crystals is possible.[8] With dry natural quartz, there is a change in the response to an applied strain at around 700 degrees C.[34] At this temperature there is a change that allows plastic deformation. This yield temperature is much lower for cultured (and for natural quartz that has water added during the experiment).[8,15] These yield temperatures are dependent on the amount of water in the crystals (see Fig. 2). For cultured quartz the yield temperature ranges from 650 degrees C for 1000 ppm H/Si to 380 degrees C for 9000 ppm H/Si (here the H content was calculated from the 3 μm IR absorption band).[8] In the plastic deformation of the quartz, this movement is accomplished by dislocation propagated slip.[8,15] That is, the movement of dislocations allows bulk plastic flow. The slip planes in quartz, which correspond to preferred directions for dislocation movement, have been determined.[2,8] Tullis[35] notes that there is a critical weakening pressure in the same sense that there is a yield temperature. After quartz is plastically deformed, high dislocation densities appear. Hobbs examined cultured quartz that had been permanently deformed at 300 to 400 degrees C using TEM and found large numbers of tangled dislocation lines.[36]

One of the first theories which considered the role of water in dislocation movement was the Frank-Griggs model[8]. In the Frank-Griggs model (Fig. 3), the effect of water weakening is explained by the hydrolyzing of O-Si bonds.[8] This supposedly makes the movement of the dislocations easier because the O-H bonds are easier to break than O-Si. Recently, McLaren et al.[10] looked at the problem of water weakening using TEM. They found that when wet quartz is heated, water bubbles expand and generate dislocation loops. These loops can, given time, connect several water bubbles. They claim that the OH terminations do not make the dislocations glide easier. Instead, when the quartz is heated, the bubbles expand (see Fig. 4) and generate the dislocations needed for

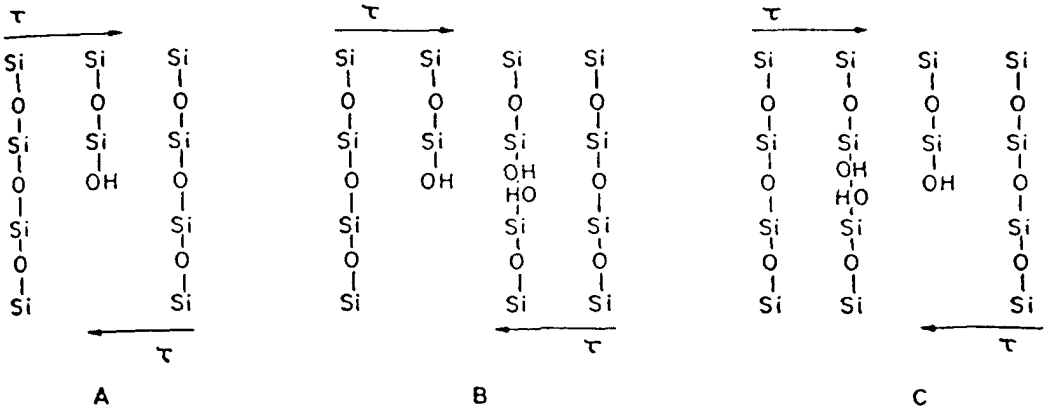


Figure 3 Frank-Griggs mechanism of dislocation movement. (A) Hydrolyzed edge dislocation with anhydrous neighbors. (B) Neighboring Si-O-Si bond hydrolyzed by water migration. (C) Dislocation moves by exchange of hydrogen bond.

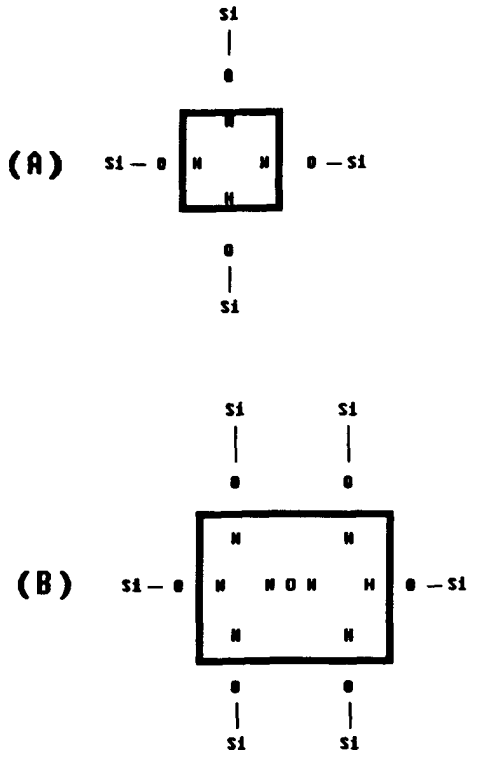


Figure 4 McLaren model of dislocation formation. (A) Bubble in quartz proir to heating. (B) Expansion of bubble through the incorporation of water.

dislocation propagated slip to occur. This explains the incubation time (the delay between the time the sample was loaded and when the first signs of strain are detected) found in some experiments. This also explains the results Armington et al.[31] reported after annealing cultured bars. As mentioned earlier, Armington et al. found that the strain around the inclusions was increased after annealing, and in the same paper they stated that most of the dislocations in their quartz were voids, partially filled with liquid. These may correspond to the bubbles seen by McLaren. Brice[37] has conducted experiments in which the alpha (the infra-red absorption coefficient at 3500 cm^{-1}, proportional to H content) is shown to be correlated to dislocation density. That is, the dislocation density can be estimated from alpha (Q_{IR}). Samples with higher alpha'a (lower Q's) have higher dislocation densities. This can be fit to the curve

$$\log_{10} N_D = 5.00 \pm 0.48 + 2.5 \log_{10} \alpha$$

where N_D is the dislocation density (cm^{-2}) and α is in cm^{-1}. Furthermore, quartz from different manufacturers yield different lines. Brice feels that this difference in the alpha vs. dislocation density plots for quartz from different suppliers is more supportive of the Griggs-Frank model as opposed to the McLaren model.

Summary

It has been shown in the geophysical work that dislocations definitely can move in quartz. However, all of that work was done at high temperatures and pressures. It remains to be seen whether or not dislocations can be formed, or are mobile under the conditions present during the fabrication and operation of resonators. One area that was not covered in the publications reviewed here was the possibility of reducing the dislocation density through post growth pressure treatment. The possibility of moving dislocations out of the active areas in resonator blanks is very appealing and deserves to be investigated.

In parallel with these experimental studies, a theoretical basis for the behavior of dislocations in quartz needs to be developed. Heggie, et al.[38-42] have published several papers studying the dislocation lines in alpha quartz using atomic models with Keating type interatomic potentials. While this is useful for determining which dislocations are possible, it is essentially a static process. Pontikis[43] recently made the point that computer molecular dynamics (MD) techniques are very well suited for examining the movement of dislocations in crystals. With MD simulations it would be possible to better interpret the dislocation experiments suggested in the last paragraph. Also, it is possible to perform simulations of P-T conditions that would be difficult and costly to do in experiments.

Acknowledgements

The author wishes to thank John Vig for suggesting this study and for his help in editing this paper.

References

[1] Derek Hull, Introduction to Dislocations, Pergamon Press Ltd., New York, 1965, pp. 13-21.

[2] M.W. Wegner and J.M. Christie, Phys. Chem. Minerals, v. 9, pp. 67-78, 1983.

[3] Fumiko Iwasaki, J. of Cryst. Growth, v. 39, pp. 291-298, 1977.

[4] John R. Vig, John W. LeBus, and Raymond L. Filler, Proc. 31st ASFC, pp. 131-143, 1977.

[5] Electronic Industries Association, Engineering Dept., 2001 Pennsylvania Ave., N.W., Washington, D.C. 20006.

[6] Gary R. Johnson and Robert A. Irvine, Proc. 41st ASFC, pp. 175-182, 1987.

[7] Kazuo Moriya and Tomoya Ogawa, Phil. Mag. A, v. 41, pp. 191-200, 1980.

[8] David Griggs, Geophys. J. R. Astr. Soc., v. 14, pp. 19-31, 1967.

[9] A.C. McLaren, R.F. Cook, S.T. Hyde, and R.C. Tobin, Phys. Chem. Minerals, v. 9, pp. 79-94, 1983.

[10] A.C. McLaren, J.D. Fitz Gerald, and J. Gerretsen, Phys. Chem. Minerals, v. 16, pp. 465-482, 1989.

[11] Fumiko Iwasaki and Masakazu Kurashige, Ferroelectrics, v. 43, pp. 43-50, 1982.

[12] C.C. Glüer, W. Graeff, and H. Möller, Nuclear Instruments and Methods, v. 208, pp. 701-704, 1983.

[13] A. Zarka, B. Capelle, J. Detaint, and J. Schwartzel, Proc. 41st ASFC, pp. 236-240, 1987.

[14] A. Zarka, B. Capelle, Y. Zheng, J. Detaint, and J. Schwartzel, Proc. 42nd ASFC, pp. 85-92, 1988.

[15] James D. Blacic, Tectonophysics, v. 27, pp. 271-294, 1975.

[16] W. Hanson, Proc. 41st ASFC, pp. 228-235, 1987.

[17] T.R. Meeker and A.J. Miller, Proc. 34th ASFC, pp. 85-92, 1980.

[18] John W. Kent, Proc. 12th Piezoelectric Devices Confrence & Exhibition, v.1, p. 1, 1990.

[19] B.J. James, Proc. 42nd ASFC, pp. 146-154, 1988.

[20] Errol. P. EerNisse, Roger W. Ward, and O. Lew Wood, Proc. 43rd ASFC, pp. 388-395, 1989.

[21] R.A. Young, R.B. Belser, A.L. Bennett, W.H. Hicklin, J.C. Meaders, and C.E. Wagner, Proc. 19th ASFC, pp. 23-28, 1965.

[22] John A. Kusters and John R. Vig, Proc. 44th ASFC, pp. 165-175, 1990.

[23] J.D. Stephenson, Phys. Stat. Sol. A, v. 106, pp. 441-449, 1988.

[24] J. Beaussier, 2nd European Frequency and Time Forum, pp. 807-819, 1988.

[25] U. Alter and G. Voigt, Crystal Res. and Technol., v. 19, pp. 1619-1623, 1984.

[26] R.L. Barns, P.E. Freeland, E.D. Kolb, R.A. Laudise, and J.R. Patel, J. of Cryst. Growth, v. 43, pp. 676-686, 1978.

[27] Y. Epelboin, J.R. Patel, J. Appl. Phys., v. 53, no. 1, pp. 271-275, 1982.

[28] S.P. Doherty, S.E. Morris, D.C. Andrews, D.F. Croxall, Radiation Effects, v. 74, pp. 145-150, 1983.

[29] Alton F. Armington, John J. Larkin, John J. O'Connor, J. Emery Cormier, and Jane A. Horrigan, Proc. 37th ASFC, pp. 177-180, 1983.

[30] Alton F. Armington and Joseph F. Balascio, Proc. 38th ASFC, pp. 3-6, 1984.

[31] Alton F. Armington, Jane A. Horrigan, M.T. Harris, and Joseph F. Balascio, Proc. 41st ASFC, pp. 213-215, 1987.

[32] R. Irvine, J. Foise, E. Leeson, and Gary Johnson, Proc. 42nd ASFC, pp. 127-137, 1988.

[33] John G. Gualtieri and John R. Vig, Proc. 38th ASFC, pp. 42-49, 1984.

[34] Neville L. Carter, Rev. Geophys. and Space Phys., v. 14, no. 3, pp. 301-360, 1976.

[35] J.A. Tullis, Reviews of Geophysics and Space Physics, v. 17, no. 6, pp. 1137-1154, 1979.

[36] B.E. Hobbs, Tectonophysics, v. 6, no.5, pp. 353-401, 1968.

[37] J.C. Brice, Reviews of Modern Physics, v. 57, no. 1, pp. 105-146, 1985.

[38] M. Heggie and M. Nylén, Phil. Mag. B, v. 50, no. 5, pp. 543-555, 1984.

[39] M. Heggie, R. Jones, and M. Nylén, Phil. Mag. B, v. 51, no. 6, pp. 573-580, 1985.

[40] M. Heggie and M. Nylén, Phil. Mag. B, v. 51, no. 6, pp. L69-L72, 1985.

[41] M. Heggie, and R. Jones, Phil. Mag. A, v. 53, no. 5, pp. L65-L70, 1986.

[42] M. Heggie and R. Jones, Revue Phys. Appl., v. 23, p. 670, 1988.

[43] Vassilis Pontikis, Physics World, pp. 25-28, 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN ANALYTICAL STUDY OF ELECTRODIFFUSION (SWEEPING) OF SYNTHETIC QUARTZ CRYSTALS

J.ASAHARA, K.NAGAI, K.HAMAGUCHI, H.SONE, AND S.TAKI

Toyo Communication Equipment Co.,Ltd.
2-1-1, Koyato, Samukawa, Kouza-gun, Kanagawa Pref.,
253-01 Japan

ABSTRACT

It was observed that Ag ions moved from the anode to cathode with higher mobility than those of Au and that the residual Ag was closely corresponded to the substitutional alminum impurity. The entire Ag is used up on alminum for charge compensation to form the association Al-Ag^{+}. The diffusion of Ag in an air atomosphere has the same etch channel reducing effect as the diffusion of Au.

The line defects observed by X-ray topography before and after sweeping were not changed by sweeping under any conditions. It was confirmed the total number of etch pits and channels roughly the same as a number of line defects detected by X-ray topography and that electrodiffusion, though it reduced the etch channel density, did not change the total number of line defects.

In comparing the results of impurity analysis of the samples before and after electrodiffusion, it was seen that Na^{+} and Li^{+} were easily diffused to the outer surface of the crystal, while the concentrations of Al^{3+}, K^{+}, Ca^{++}, and Mg^{++} were not changed. The impurity analysis of the unswept samples showed that Ca^{++} and Mg^{++} significantly decreased with the formation of etch channels without any change in the concentrations of Na^{+}, Li^{+}, K^{+}, and Al^{3+}. Therefore, it appears Ca^{++} and Mg^{++} are localised to the highly distorted lattice surrounding a line defect and are preferrentially removed during etch channel formation.

INTRODUCTION

It is well known that electrodiffusion (sweeping) reduce the formation of etch channels. This effect, is influenced by the electrodiffusion conditions, including the electric field strength, electrode type and atmosphere.

The etch channel formation conditions differs according to whether electrodiffusion takes place in a vacuum or air atmosphere. Vacuum electrodiffusion produces etch channel free elements.[1] If done in air, however, the impurities originally present in the crystal, diffuse to the cathode and the airbone impurities are taken up by the crystal from the anode.[2]

Electrodiffusion performed in a nitrogen atmospere leads to the same result as vacuum electrodiffusion, in that the synthetic crystal with a high etch channel density is converted to an etch channel-free crystal.[3] If electrodes of Cr and Au evaporated metallization are applied, etch channels are not form even when electrodiffusion is carried out in an air atmosphere.

Extensive research was carried out on etch channels in quartz crystal.[5]-[12] Etch channels are formed as a result of the presence of line defects in the quartz crystal and of impurities segregating out along such line defects. These impurities have a highly selective solubility vis a vis the etchant. The mechanism by which etch channels are reduced through electrodiffusion, however, has not been fully expained . The present study was undertaken with a view to making this mechanism clear. To establish the relationship between the reduction in etch channel density and the electrodiffusion conditions, the latter were systematically altered in this investigation.

CH2965-2/91/0000-009 $1.00 © 1991 IEEE

The diffusion of silver (Ag) into quartz crystals in an air atomosphere has the same etch channel-reducing effect as the diffusion of gold (Au). The diffused Ag forms a layered deposition on the cathode.

The Au and Ag diffusion time and amount play a decisive role in the process of reducing the quartz crystal's etch channel density. Diffused Ag, in paticular, imparts an electrical charge to Al to form Al-Ag^{+}.

The total number of etch pits and etch channels was roughly identical with the number of line defects detectable by X-ray topography. There were no observable changes in the number of dislocations before and after electrodiffusion.

It is possible to detect calcium and magnesium in the vicinity of the line defects giving rise to the formation of etch channels. The amount of calcium impurities is correlated to the etch channel density. These impurities are not swept by electrodiffusion suggesting that the line defects responsible for etch channel formation are Ca^{++} and Mg^{++}-containing, high-strain defects, with the strain energy reducing the activation energy required for the etchant. Electrodiffusion relieves this strainand thus prevents the formation of etch channels. Au and Ag difusion and use of a vacuum atmosphere inhibit the ingress of impurities and are therefore effective in achieving strain relaxation.

EXPERIMENTAL

Fig.1 shows the experimental equipment and Table I an example of the electrodiffusion conditions. Both surfaces of the specimen were metallized with the electrode material by vapor deposition. The electrodiffusion conditions were systematically altered to observe the resulting diffusion process and etch channel formation. The high-voltage supply used in the experiments was a RHR10PN60 unit manufactured by SPELLMANN and the ammeter a Digital Multi-Meter Unit 6552 by Yokogawa. A constant voltage was applied from the inception of heating until the specified temperature was reached. The electric field was applied after the specimen was heated to, and maintained at, this specified temperature until its temperature dropped to ambient temperature. The current was recorded at regular intervals.

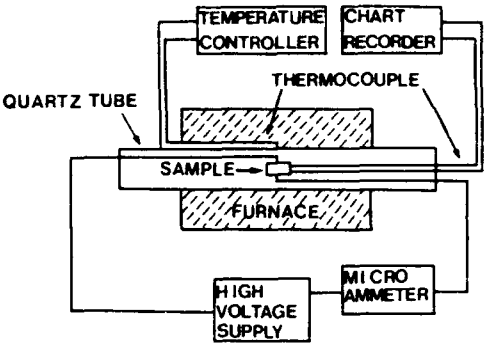


FIG. 1. Outline of Experimental Equipment

TABLE I. Electrodiffusion Conditions

Temperature	: 500 - 560 ℃
Electric Field Strength	: 1000 - 2000 V/cm in Z axis direction
Atmosphere	: Air
Electrodiffusion Time	: Time until current reaches a constant value

The specimen was etched with hydrofluoric acid at 100℃ in a Teflon pressure vessel. For observation, a microscope with 50-fold magnification was used. To determine the presence of Al-OH^{+} before and after electrodiffusion, the specimen was cooled to 78K for mesurement with an infrared spectrometer (JASCO A-702). The specimen's line defect were assessed by X-ray topography.

ICP atomic emmission spectrometry was used to analyzed the impurities. Due to the poor sensitivity for potassium, however, flameless atomic absorption analysis was used to analyze this element. The specimen was ground using a glove box, in a cleanroom controlled to strict cleanliness standards. In order to prevent contamination due to the grinding operation, it was not ground to a fine powder but totally dissolved by the pressur·zed dissolving method, using high-purity HF. To avoid error, the acidity of the specimen prepared in the cleanroom is the same as that of the measurement line standard. The detection limit for Al is 10ppb and for Na, K, Li, Ca, and Mg are 1ppb. The accuracy of the analysis is therefore ±5% in either case.

RESULTS

Figs. 2 and 3 show the current density values measured in 24 hour electrodiffusion experiments carried out with Au and Ag as the electrode materials. Au and Ag diffusion have differnt current density vs. time profiles.

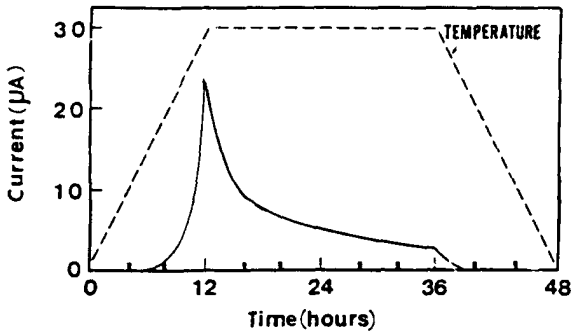


FIG. 2. Changes in Temperature and Current Density during Au Diffusion vs. Time

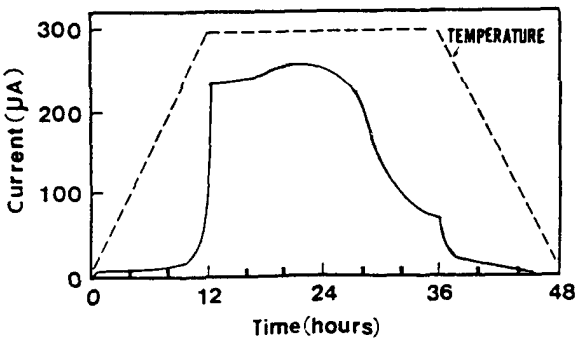


FIG. 3. Changes in Temperature and Current Density during Ag Diffusion vs. Time

In Au diffusion, the current density rises with rise in temperature. At the hold temperature, the impurities inside the crystal are swept, while the current density drops. Electrical conductivity is attributed to M^+ ion migration along the tunnels existing parallel to the c axis in the crystal.[15] The consumption of the cathode after electrodiffusionwith Au suggests that Au diffuses as a negatively charged ion. Ag, however, diffuses from the anode to cathode as a positively charged ion.

The Ag diffusion pattern in an air atmosphere is substantially different from that in a vacuum atmosphere. Ag diffused in an air atmosphere will initially build up on the cathode and, with further progress of the diffusion process, gradually form layered deposition in the quartz crystal. Fig. 4 shows a cross-section of the layered deposition near the cathode. Layered depositions exhibit metallic luster.

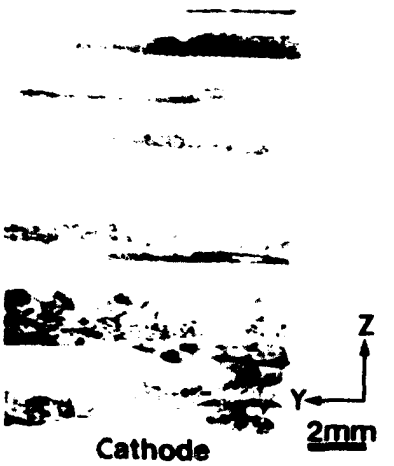


FIG. 4. Layered Deposition Pattern Occurring during Ag Diffusion in an Air Atmosphere

Electrodiffusion of Ag in a vacuum atmosphere does not lead to the formation of a layered deposition. Instead, it completely deposits on the internal surface of the quartz crystal. Fig. 5 shows the Ag deposition pattern during electrodiffusion in a vacuum atmosphere.

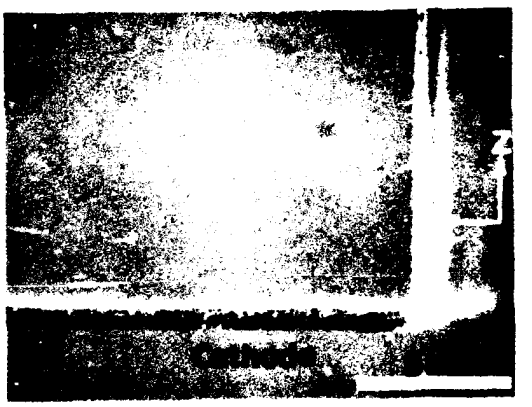


FIG. 5. Cathodic Ag precipitation in Vacuum Electrodiffusion

The analysis results at each stage of the electrodiffusion process are indicative of the state of the diffused substances and the impurity sweeping conditions. These results are most significant because of the extensive information they provide.

Table II represents the analysis data for a specimen subjected to Au diffusion. Na and Li are swept from the quartz crystal as a result of electrodiffusion while the electrode Au is diffused into the crystal.

TABLE II. Changes in Impurity Levels Before and After Au Diffusion

Element	Before Sweeping	After Sweeping
Al	4.20	4.44
Na	0.61	0.003
Li	1.10	0.001
Au	N.D.	0.002

(unit ppm)

Table III shows the analysis results obtained with an electrodiffusion specimen subjected to Ag diffusion. The analyzed specimen was taken from a location in the crystal which had no Ag depositions. Na, Li, are swept from the quartz crystal as a resultt of electrodiffusion while the electrode Ag is diffused into the crystal. The amounts of Al, K, Ca, and Mg do not change. The total of the atomic ratios for Na, Li, and K before electrodiffusion is roughly the same as the atomic ratio for Al. This suggests that all M^+ ions lead to substitutive Al charge compensation.

Figs. 6 and 7 show the analysis results for the impurity concentrations as they change in the course of time during Au and Ag electrodiffusion. At the beginning of electrodiffusion, Ag is present in the quartz crystal in high concentration and the atomic ratio of this concentration is roughly equal to the atomic ratio of Al.

The electrodiffusion process results in the sweeping away of Na and Li so that their concentrations decrease. This is accompanied, in part, with the formation of Al-OH, as is evidence by the appearance of a 78K infrared absorption band. The formation of Al-OH, however, takes place to a lesser extent with Ag diffusion rather than Au diffusion. Fig 8 and 9 show the diffusion time and the proportion of the impurities ions versus the concentration of aluminum ions. The data of Fig. 9 show that the entire Ag is used up on Al for charge compensation to form the association $Al\text{-}Ag^+$. It seems that analyzing the residual Ag is to provide the most reliable testing method for complete aluminum which substitute Si in quartz crystal.

Table III. Change in Impurity Levels Before and After Ag Diffusion

Element	Before Sweeping	After Sweeping
Al	3.36 7.48	3.61 8.04
Na	0.45 1.1?	0.01 0.03
Li	1.00 8.66	<0.00₅ <0.04
K	0.02 0.03	0.02 0.03
Ca	<0.01	0.00₉
Mg	<0.00₁	0.00₁
Ag	N.D.	0.07

{unit upper ppm
lower ppma}

Tables IV and V represent the etch channel formation in the specimen on electrodiffusion of Au and Ag, respectively. Electrodiffusion was performed by changing the applied voltages to suit the different etch channel densities of the samples. Table IV refers to the sample's etch channel formation pattern in the case of Au and Table V, in case of Ag diffusion.

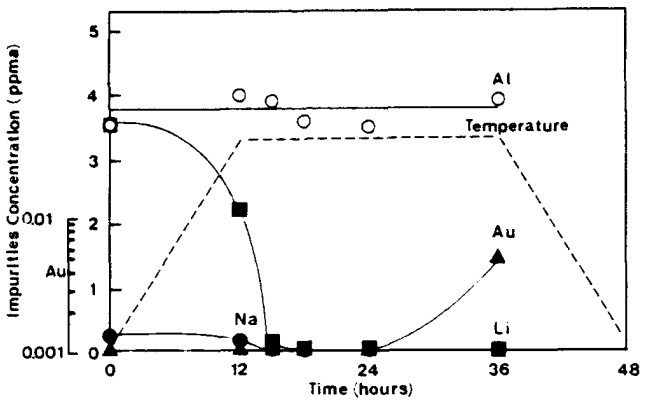


FIG. 6. Changes in Impurity Concentration in a Quartz Crystal Associated with Au Electrodiffusion

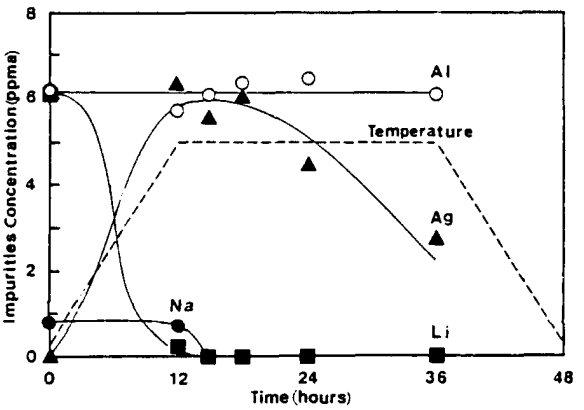


FIG. 7. Changes in Impurity Concentration in a Quartz Crystal Associated with Ag Electrodiffusion

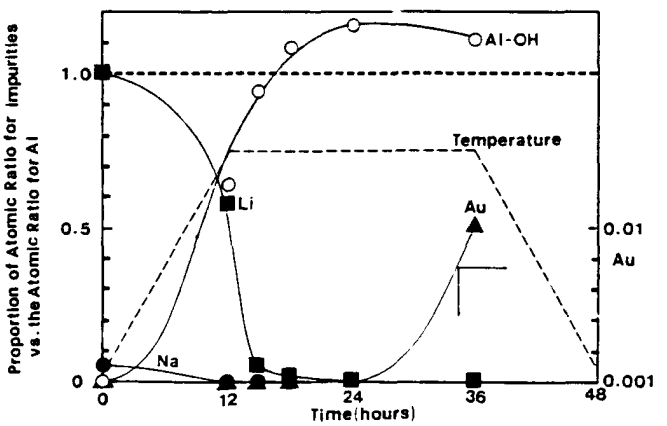


FIG 8. Diffusion Time and the Proportion of the Impurities Ions versus the Concentration of Aluminum Ions in the Au Diffused Quartz Crystals

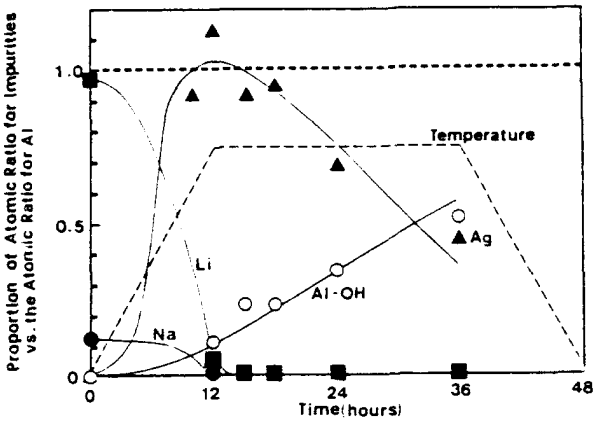


FIG. 9. Diffusion Time and the Proportion of the Impurities Ions versus the Concentration of Aluminum Ions in the Ag Diffused Quartz Crystals

TABLE IV. Au Diffusion Conditions vs. Etch Channel Formation Pattern

(After/Before Sweeping)

Electric Field (V/cm)	Etch Channel Density (cm^{-2})				
	< 20	< 50	< 500	<1000	<2500
2000	–	0/ 32	0/ 231	0/ 721	0/2200
1000	–	0/ 33	0/ 274	0/ 863	0/2314
500	–	0/ 34	0/ 228	1/ 767	–
250	0/ 11	0/ 30	1/ 249	8/ 746	–

TABLE V. Ag Diffusion Conditions vs. Etch Channel Formation Pattern

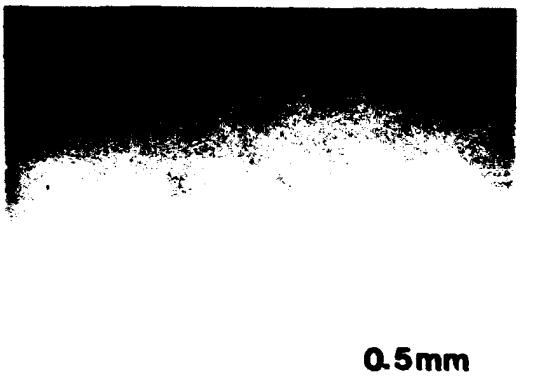
(After/Before Sweeping)

Electric Field (V/cm)	Etch Channel Density (cm^{-2})				
	< 20	< 50	< 500	<1000	<2500
1000	–	–	–	–	0/2139
500	–	–	–	0/ 791	0/2256
250	–	0/ 34	0/ 264	1/ 738	0/2573
125	0/ 10	0/ 33	1/ 259	0/ 764	0/2166

Ag diffusion leads to a reduction in etch channel formation, and so does Au diffusion. Both Au and Ag electrodiffusion had the effect of reducing to zero the etch channel density in a quartz crystal which initially had an etch channel density of approximately 2,500 channels cm^2.

Ag has a higher mobility than Au so that it is capable of diffusing into a quartz crystal at a lower applied voltage to reduce its etch channel density to zero.

Figs.10 and 11 give the etching surfaces before and after the diffusion process for Au and Ag electrodiffusion, respectively.



(b) after Au electrodiffusion

FIG.10. Etching Surface Before and After Au Electrodiffusion

Table VI gives the results of comparing the etch channel, etch pit, and line defect numbers before and after electrodiffusion.

The etch channel and etch pit count were taken after etching an AT plate and counting under a microscope at 50 fold magnification. The line defect count was determined by calculation from an X-ray topography image of a Y plate prepared from the same specimen.

The data show that the line defect count detected by X ray topography shows practically no change between the values before and after diffusion. Nor did the etch pit count present any significant change. The sum total of etch channel and etch pit counts before diffusion is in good agreement with the line defect number detected by X ray topography. Thus the etch channel count is reduced by the diffusion operation, but the line defect density does not change.



(a) before Ag electrodiffusion



(b) after Ag electrodiffusion

FIG.11. Etching Surface Before and After Ag Electrodiffusion



(a) before Au electrodiffusion

TABLE VI. Etch Channel and Etch Pit Counts as well as Line Defect Number (defects/cm^2) Detected by X-ray Topography, Before and After Au Electrodiffusion (etch channel/ etch pit/ line defect)

Electric Field (V/cm)	Before Electrodiffusion	After Electrodiffusion
250	11/78/85	0/67/83
500	11/64/79	0/72/84
1000	10/66/80	0/64/79
2000	10/69/81	0/74/86

Some very interesting results were obtained for the changes in impurity concentration before and after etch channel formation in the sample that had not been treated by electrodiffusion (TABLE VII.).

It can be seen that the sum total of the atomic ratios for Na, Li, and K is in very good agreement with the atomic ratio for Al. This suggests that the M^+ ions are all present in a charge-compensating capacity imparting a charge to Al by way of substitution. Al and the Na, Li, and K ions associated with it do not change after etching as compared with the concentrations before etching. This suggests that the defects due to substitutional manner completely related to the etch channels.

The Ca and Mg concentrations, however, are reduced as a result of the formation of etch channels. This indicates that Ca and Mg segregate out in the proximity of the etch channels. The analysis data for quartz crystals with a high etch channel density include a large amount of Ca impurities and there is a correlation between the two impurities (Fig. 12).

TABLE VII. Changes in Impurity Concentration in the Quartz Crystal Before and After Etch Channel Formation

Element	Before Etching	After Etching
Al	1.19 2.65	1.39 3.10
Na	0.08 0.21	0.08 0.21
Li	0.24 2.08	0.39 3.38
K	0.01 0.02	0.02 0.03
Ca	0.09 0.13	$<0.00_1$ $<0.00_1$
Mg	0.01_5 0.04	0.00_1 0.00_3

(unit upper ppm, lower ppma)

Fig.13 shows the etch channel density and the proportion of the atomic ratio for Na, Li, and K versus the atomic ratio for Al. In a specimen with a low etch channel density the Al atomic ratio and the atomic ratio for the alkali metal ion impurities as a whole are in good agreement. As the etch channel density increases, the total of the atomic ratios of the alkali metal impurities becomes greater than the Al atomic ratio. The total of alkali atoms after etching, however, is in agreement with the atomic ratio for Al.

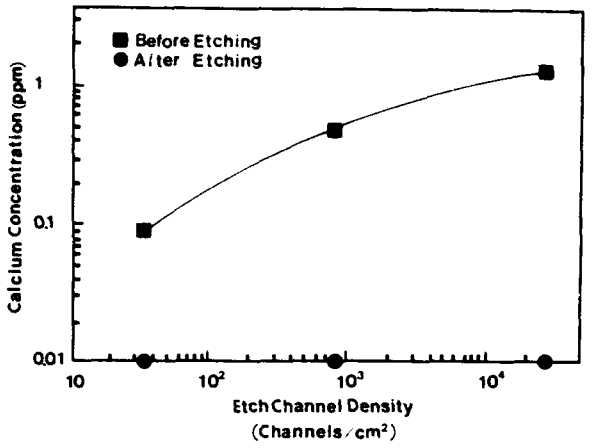


Fig. 12. Amount of Calcium Impurities in the Quartz Crystal vs. Etch Channel Density

In a quartz crystal with a high etch channel density, M^+ ions can be found in the proximity of the etch channel-forming line defects, in addition to Al-M^+ associations which substitute Si in the quartz crystal.

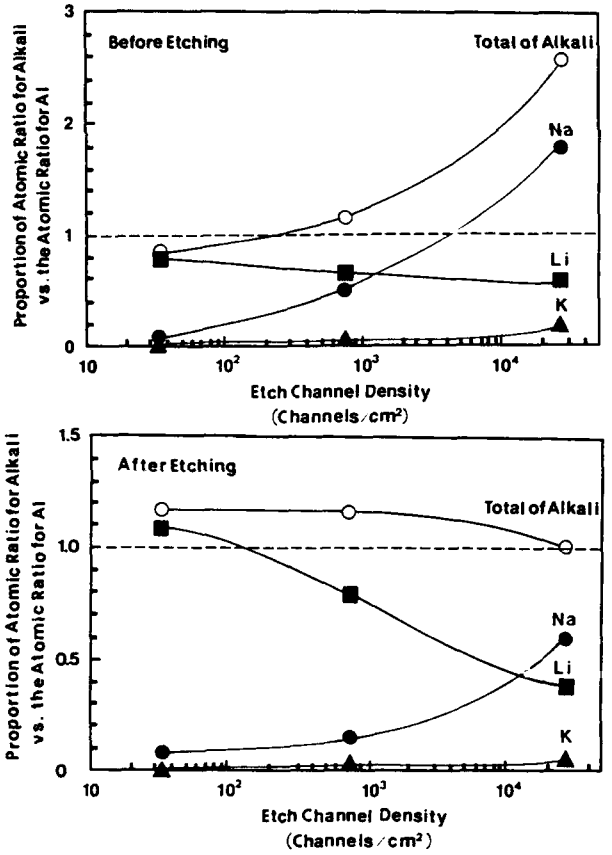


FIG. 13. Comparison of Al and Alkali Impurity Atomic Ratios Before and After Etching

Fig. 14 shows the etch channel counts on the anode and cathode surfaces changing with Au diffusion time. The quartz crystal's pre-diffusion etch channel density was 1l/cm². The etch channels were found to decrease at the cathode first and at the anode.

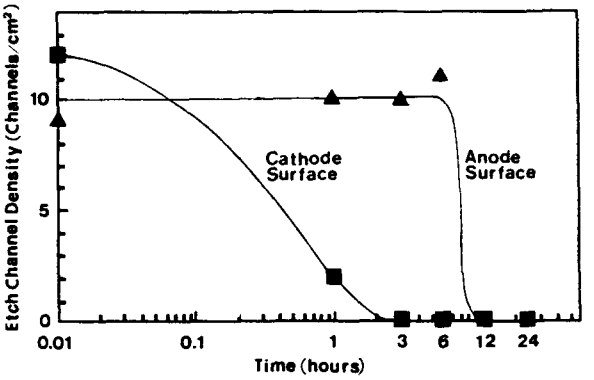


FIG. 14. Changes in Etch Channel Density vs. Diffusion Time (Etch Channel Density of Blank = 1l/cm²)

DISCUSSION

Ca and Mg are present in the proximity of etch channel-forming line defects, and these divalent impurity elements are not swept away by electrodiffusion. This indicates that selective Ca and Mg etching does not lead to the formation of etch channels.

The line defects giving rise to the formation of etch channels form high-strain sites to the extent that divalent ions are present. These have a high selective solubility with respect to the etchant, thus leading to the formation of etch channels.

X-ray topography has shown that the line defects are not eliminated by electrodiffusion. The reduction in etch channel density is thus not due to a decrease in the dislocation density.

When electrodiffusion is performed in an air atmosphere there is a time difference for the reduction in etch channel density at the anode and cathode. This difference shows that metal diffusion is required for reducing the etch channel density. The presence of metals

prevents the ingress of impurities from the atmosphere during electrodiffusion and is thus effective in achieving strain site relaxation.

When Ag is diffused in an air atmosphere, layerd depositions gradually form in the quartz crystal from the cathode surface. Ag disperses in the quartz crystal as a positively charged ion in the form of Ag^{+}. This Ag^{+} is reduced by the negatively charged substance entering from the cathode with the formation of metallic Ag. When electrodiffusion is conducted in a vacuum atmosphere, however, Ag precipitates on the cathode without the formation of layered depositions since there is no ingress of negatively charged ions.

In the absence of intruding negatively charged ions, however, the strain energy of the line defects is sufficiently relieved.

CONCLUSIONS

The following conclusions can be drawn from the present experimental study.

1. The formation of etch channels is due to the occurrence of line defects with such a large strain energy that Ca^{++} and Mg^{++} impurity ions can be present. Electrodiffusion is effective in relieving this strain energy so that a reduction in the number of etch channels is thus possible.

2. When electrodiffusion takes place in an air atmosphere, the strain energy of the line defects is adequately relieved. This may be attributed to the diffusion of the airborne impurities into the quartz crystal. Our experimental studies are to be continued to shed more light on this question.

3. Our electrodiffusion experiments in air and in a vacuum using Au and Ag electrodes have demonstrated that electrodiffusion prevents the diffusion of airborne impurities into the quartz crystal.

ACKNOWLEDGEMENTS

We gratefully acknowledge the assistance Dr. Namikawa and Dr. Kurihara of Nippon Sanso Co., Ltd. in chemical analyzing the quartz crystals used in this study.

REFERENCES

[1] J.R.Vig, J.W.LeBus, and R.L.Filler, "Chemically Polished Quartz", Proc. 31st Ann. Symp. Freq. Control, pp.131-143, 1977.

[2] G.B.Krefft, "Effects of High Temperature Flectrolysis on the Coloration Characteristics and OH-Absorption Bands in Alpha-Quartz", Raiation Effects, Vol. 26, pp.249-259, 1975.

[3] J.Dowsett, D.F.G.Dwyer, F.Stern, R.A.Heinecke, and A.H.Truelove "Etch Processing of Bulk and Surface Wave Devices", Proc. 39th Ann. Symp. Freq. Control, pp.301-308, 1985.

[4] J.G.Gualtieri, "The Influence of Temperature and Electric Field on the Etch-Channel Density in Swept Cultured Quartz", Proc. 39th Ann. Symp. Freq. Control, pp.247-254, 1985.

[5] J.W.Nielsen and F.G.Foster, "Unusual Etch Pits in Quartz Crystals", The Amer. Mineralogist, Vol.45, pp.299-310, 1960.

[6] G.W.Arnold Jr, "Defects in Quartz Crystals", Proc. 11th Ann. Symp. Freq.Control, pp.112-129, 1957.

[7] F.Augustine, "Improving the Quality of Synthetic Quartz", Proc. 11th Ann. Symp. Freq. Control, pp.130-141, 1957.

[8] F.Augustine and D.R.Hale,"Topography and Etch Patterns of Synthetic Quartz ", J. Phys. Chem. Solids, vol. 13, pp.344-346 1960.

[9] T.Hanyu , "Dislocation Etch Tunnels in Quartz Crystals", J. Phys. Soc. Japan , vol.19, p.1489, 1964.

[10] G.Gulzow-Qual."Atzversuche an Synthetishen Quarz", Physica Status Solidi, vol.1, pp.62-67, 1961.

[11] E.V.Tsinzerling and Z.A.Mironova, "Revealing Dislocations in Quartz by Selective Etch Method", Sov. Phys. Crystallogr., vol.8, pp.89-91, 1963.

[12] O.B.Volskaya, "Selective Etchants for Quartz", Sov. Phys. Crystallogr.,vol.13, pp.620-623, 1969.

[13] A.Ser. J.Appl. Crystallography, vol.13, p.50, 1980.

[14] S.Katz, A.Halperin, and M.Schieber,"Growth Tunnels in Quartz Crystals", Proc. 36th Ann. Symp. Freq. Control, pp.193-196, 1982.

[15] A.S.Nowick and H.Jain, "Electrical Conductivity and Dielectric Loss of Quartz Crystals Before and After Irradiation", Proc. 34th Ann. Symp. Freq. Control, pp.9-13, 1980.

Question 1

How were the concentration measurements made over the time of sweeping? Were identical bars used with different sweep times for each - or was the experiment intercrystal and slices removed?

Answer:

The same sample was not used with different electrodiffusion times. I assumed that the samples which were cut from identical lumbered pure-Z bar have the same properties. I prepared eight samples which were made from one lumbered bar. Six samples out of eight were used to measure the change of impurities concentration during the electrodiffusion. Two samples were preserved as reference specimens.

Each sample was maintained with different specified electrodiffusion times and used for chemical analysis. The method that after removing specimen from the bar, identical bar was resumed sweeping again was not adopted.

Before electrodiffusion, I inspected the number of inclusions and etch channel density. Inclusion makes small cracks in the quartz crystal during sweeping process and diffused Au or Ag precipitates at such cracks. These precipitations cause chemical analysis errors. For the inspection of etch channel density, samll specimens which were cut from each sample were etched with 49% HF solusion at 100℃ in a Teflon pressure vessel. If each etch channel densities showed big difference, that samples were rejected.

Question 2

What evidence is there that sweeping relieves strain - since X ray diffraction topographs taken before and after sweeping show little or no difference?

Answer:

I think that the line defect condition must be compared with before and after electrodiffusion using same specimen by X-ray topography. The specimen for X-ray topography is too thin to diffuse Au or Ag. After the comparison of X-ray topography results with defferent diffusion times, I did not conclude whether there was difference or not. Maybe the distorted field was too small to detect any difference by our X-ray topography equipment.

The results of observation on etched surface of AT- and ST-cut wafer show two kind of defects. One defect forms etch pit and etch channel and the other defect forms only etch pit. In my opinion, these two defects can be detected by X-ray topography as line defects.

Chemical quantitative analysis results show that amount of calcium impurities in Z growth region is reduced in the case of etch channel formation. I found good relationship between the amount of calcium impurities and etch channel density. At first, I thought that calcium impurities have high solubility vis-a-vis the etchant and that selective dissolving causes the formation of etch channels.

Chemical analysis results of swept quartz crystals, however, show that calcium impurities are not swept by Au or Ag electrodiffusion. Au or Ag electrodiffusion produce etch channel free elements in spite of seeing no reduction of calcium impurities.

Second, I supposed that the disapperance of line defects is the cause of etch channel reduction. I counted the line defects detected by X-ray topography. There were no observable changes, however, in the number of line defects before and after electrodiffusion.

Finally, I assumed that there are two kinds of line defect. One is just mismatched lattice. The other defect is highly distorted. I presumed that highly distorted defects are surrounded highly distorted field. Divalent impurities such as calcium and magnesium are trapped in this highly distorted field. Trapped divalent impurities in line defects may make highly distorted field. So, it is not possible to decide which is the cause and

which is the effect. It can be considered that highly distorted field has high strain energy. I think that this high strain energy reduces the activation energy required for etchant. Diffused Au or Ag causes relaxation of this highly distorted field.

Question 3

1 Where is the citation for reference [4] in the paper?

2 Since etch-channel density decreases even witnout much electrode migration, e.g. when foil electrodes are used - how do you account for the etch-channel decrease under those conditions?

Answer:

I apologize to you for the omission reference number end of the sentence of "If electrodes of Cr and Au evaporated mettallization are appied, etch channels are not form even when electrodiffusion is carried out in an air atmosphere." in the introduction. Sorry again.

I think that much electrode material migration is not necessary for the reduction of etch channel density.

When silver plate electrodes (thickness = 1mm) were used, the concentration of Ag in the quartz crystal was less than 70 ppb. In case of such a small quantity Ag diffusion, Ag electrodiffusion had the effect of reducing to zero the etch channel density in a quartz crystal which initially had an etch channel density of approximately 2,500 channels/cm^2.

When Au was diffused in the quartz crystal, the concentration of diffused Au was only $1 \sim 3$ ppb. In the case of our Na diffusion in the quartz crystal, etch channel density reduction did not observed in spite of large amount of Na diffusion. So I think that the quantity of migrated material is not important. Au or Ag diffusion is critical even without much electrode material diffusion.

Question 4

Since Au migration increases Al-OH formation and Alkali removal and this leads to near-zero etch-channel density, the suggestion is that the more efficient sweeping plays a major role and Ca and Mg which do not partcipate in the sweeping process plays a minor role - do you agree?

Answer:

The total atomic ratio of the Ag and Al-OH concentration is roughly equal to the atmic ratio of Al during Ag electrodiffusion process in an air atomosphere. This suggests that the entire Ag form the association Al-Ag. Ag moves in the quartz crystal through the Al point defects.

The mechanism of Au diffusion is not clear. There is no evidence that diffused Au form Al-Au. Au diffuses as a negatively charged ion. This shows that it is difficult Au form the association Al-Au.

When Pt plates were used as the electrode in an air atomosphere, the reduction of etch channel density was not observed althouth Na and Li were swept out of crystal. This shows that Na and Li sweeping is less effective for the reduction of etch channel density.

The impurity analysis of no Au or Ag electrodiffused samples showed that Ca and Mg significantly decreased with the formation of etch channel. This shows Ca and Mg are localised to the highly distorted lattice surrounding a line defect and are preferntially removed during etch channel formation.

In compraring the results of the impurity analysis of samples before and after Au or Ag electrodiffusion, the concentration of Ca and Mg were not changed, While Na and Li were easily diffused to the outer surface of the crystal.

Ca and Mg do not partcipate in the electrodiffusion process, but these ions are one of causes of line defect which form the etch channels.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

THERMAL EXPANSION OF ALPHA QUARTZ

J. A. Kosinski, J. G. Gualtieri, and A. Ballato*

US Army Electronics Technology and Devices Laboratory (ETDL),
SLCET-MA-A, Fort Monmouth, NJ 07703-5601

* US Army ETDL, SLCET-DS, Fort Monmouth, NJ 07703-5601

Abstract

Existing data for thermal expansion of alpha quartz (between -50° C and 150° C) have been critically analyzed through a program funded by ETDL. A recommended "best" set of values was received as were third-, fourth-, and fifth-order power series expansions for the coefficients of thermal linear expansion (CTE), referenced to 0° C. In order to fully utilize the results, relationships between the CTEs and the thermal expansion coefficients (α_{ij}) were derived and α_{ij} referenced to 25° C were obtained. Based on the results, an additional analysis to third order in α_{ij} was performed. The new α_{ij} values allow direct comparison with previously published thermal expansion coefficients. The influence of the new α_{ij} on determinations of quartz material temperature coefficients and on the calculation of temperature coefficients of frequency for the case of the AT-cut are discussed.

Introduction

There is an internationally recognized interest in obtaining more reliable alpha-quartz material constants [1,2]. This interest also is recognized, e.g., by ongoing projects sanctioned by the Electronic Industries Association (EIA) [3], which have highlighted the need to examine all of the quartz material constants.

The thermal expansion coefficients $\alpha_{ij}^{(n)}$, also known as the thermoelastic constants, are critical parameters in piezoelectric crystal resonator and filter calculations. For example, the first order temperature coefficient of frequency of the simple thickness modes of a crystal plate can be shown to be of the form [4]:

$$2T_f^{(1)} = T_{\bar{c}}^{(1)} - T_{\rho}^{(1)} - 2T_h^{(1)} \qquad (1)$$

where f= frequency, $\bar{c}$= piezoelectrically stiffened elastic constant, ρ = mass density, and h=thickness. The quantities $T_{\rho}^{(1)}$ and $T_h^{(1)}$ depend directly on thermal expansion coefficients and are given by:

$$T_{\rho}^{(1)} = -(\alpha_{11}^{(1)} + \alpha_{22}^{(1)} + \alpha_{33}^{(1)}) \qquad (2)$$

and

$$T_h^{(1)} = +\alpha_{22}^{(1)} \qquad (3)$$

where the $\alpha_{ij}^{(1)}$ are thermal expansion constants and $\alpha_{22}^{(1)}$ is the thermal expansion constant rotated to the crystallographic direction of the plate thickness, which is taken as the x_2-direction. The $\alpha_{ij}^{(n)}$ are components of tensors of the second rank.

At present there are three sets of thermal expansion coefficients in general use: the Bechmann, Ballato, and Lukaszek (BBL) set [4]; the American Institute of Physics (AIP) set [5] ; and the Mizan and Ballato (MB) set [6]. The MB set is based on a digitization of a graph published by Sosman [7]. Brice [8] and James [9] have also contributed alternative data sets for alpha quartz thermal expansion. In Fig. 1 it is shown that the thermal expansion coefficients in general use do not agree with one another over the temperature range -50° C to + 150° C.

Experimental Data

Funding and time constraints did not allow the gathering of new experimental data. Instead, a literature search and critical analysis of published data were performed. This phase was accomplished by the Purdue University - Center for Information and Numerical Data Analysis and Synthesis (CINDAS). CINDAS operates the High Temperature Materials - Mechanical, Electronic and Thermophysical Properties Information Analysis Center (HTMIAC) for the Department of Defense [10].

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

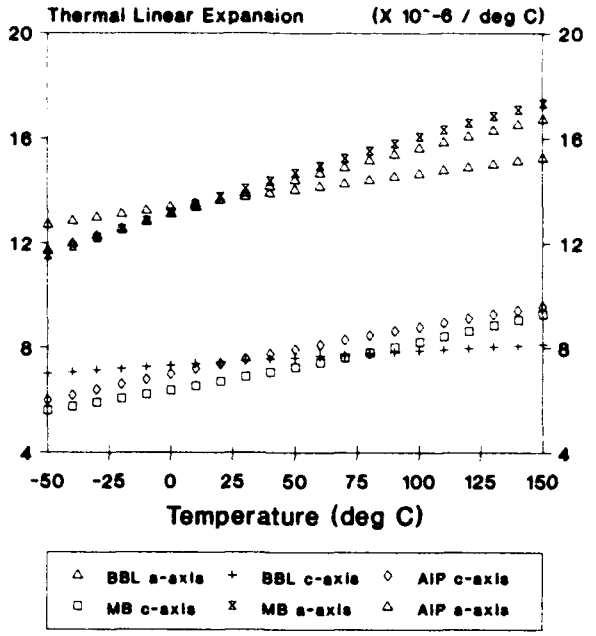


Fig. 1. Thermal expansion calculations using the thermal expansion coefficients of BBL [4], AIP [5], and MB [6]. There is little agreement for the temperature range -50° C to 150° C.

The HTMIAC analysis involved searching, compiling, and critically analyzing existing thermal expansion data. The HTMIAC recommended values for the coefficient of thermal linear expansion (CTE) along the a-axis were obtained by fitting to the data of White [11] and Barron [12], in conjunction with the data of Jay [13], Johnson and Parsons [14], and Lager et al. [15]. Data of Dorsey [16] and Buffington and Latimer [17] also were included. The recommended values for CTE along the c-axis were obtained from Refs. [11-15] and also included the data of Nix and McNair [18].

The recommended coefficient values for CTE were fitted to 3rd-, 4th-, and 5th-order power series of temperature referenced to 0° C and valid for the temperature range -50° C to +150° C. The CTE curve fit coefficients are listed in Table I. See Fig. 2 for a plot of the fit to a 5th-order power series.

As shown in Figs. 3-5, none of the thermal expansion coefficient sets in general use agree well with the HTMIAC recommended data.

ETDL Analysis

Unfortunately, the CTE used by HTMIAC are not equivalent to the thermal expansion coefficients, $\alpha_{ij}^{(n)}$, that are normally used in frequency control R&D. Specifically,

$$CTE = 1/h(\theta_o)\ \partial h/\partial\theta \qquad (4)$$

with reference temperature θ_o=0° C, whereas the more usual thermal expansion coefficients, $\alpha_{ij}^{(n)}$, are obtained from a Taylor series expansion of length or thickness (h) about a reference temperature θ_o=25° C :

$$h(\theta)=h(\theta_o)\sum_{n=o}^{m} 1/n!\ 1/h(\theta_o)\ \partial^n h/\partial\theta^n\ (\theta-\theta_o)^n \qquad (5)$$

whence

$$\alpha_{ij}^{(n)} = T_h^{(n)} = 1/n!\ 1/h(\theta_o)\ [\partial^n h/\partial\theta^n]_{\theta_o}. \qquad (6)$$

In the past, most analyses have been limited to order m = 3.

Mathematical relationships between the CTE and the $\alpha_{ij}^{(n)}$ were derived, accounting also for the different reference temperature. The HTMIAC expansion of (4) can be written as

$$CTE = \sum_{n=o}^{m} a_n \theta^n. \qquad (7)$$

We can then write h(0) in terms of the $\alpha_{ij}^{(n)}$ and $h(\theta_o)$ as

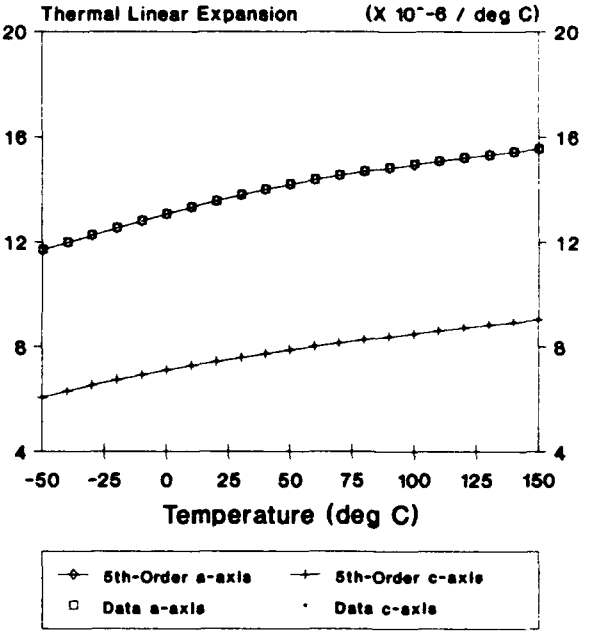


Fig. 2. A 5th-order power series least-squares fit to the recommended HTMIAC data for the coefficient of thermal linear expansion.

$$h_o = h(0) = h(\theta_o)\left[\sum_{n=o}^{m} \alpha_{ij}^{(n)} (-\theta_o)^n\right] \quad (8)$$

where $\theta_o = 25^\circ$ C.

Table I

HTMIAC CTE POWER SERIES COEFFICIENTS

a-axis

Coeff.	Order of Fit 5	4	3	Units
a_0	13.1	13.1	13.1	X 10^{-6}/K
a_1	26.1	25.8	25.0	X 10^{-9}/K^2
a_2	-60.6	-54.3	-64.7	X 10^{-12}/K^3
a_3	-466.2	-320.3	45.1	X 10^{-15}/K^4
a_4	4645	1827		X 10^{-18}/K^5
a_5	-11270			X 10^{-21}/K^6

c-axis

a_0	7.1	7.1	7.1	X 10^{-6}/K
a_1	17.2	17.5	17.7	X 10^{-9}/K^2
a_2	-45.5	-50.9	-48.4	X 10^{-12}/K^3
a_3	323.0	-200.0	111.7	X 10^{-15}/K^4
a_4	-2817	-441.6		X 10^{-18}/K^5
a_5	9503			X 10^{-21}/K^6

From (5) and (6) we find that

$$\partial h(\theta)/\partial\theta = h(\theta_o)\left[\sum_{n=o}^{m} n\alpha_{ij}^{(n)} (\theta-\theta_o)^{n-1}\right] \quad (9)$$

thus

$$CTE = 1/h_o\ \partial h(\theta)/\partial\theta \approx \sum_{n=o}^{m} n\alpha_{ij}^{(n)} (\theta-\theta_o)^{n-1}, \quad (10)$$

since the summation in (8) is $\approx$ 1. Therefore, we find

$$CTE = \sum_{n=o}^{m} a_n\theta^n = \sum_{n=o}^{m} n\alpha_{ij}^{(n)} (\theta-\theta_o)^{n-1}. \quad (11)$$

Expanding both summations and equating coefficients of equal powers of θ, one obtains (after some manipulation) the following thermal expansion coefficients up to order six in terms of the CTE power series coefficients:

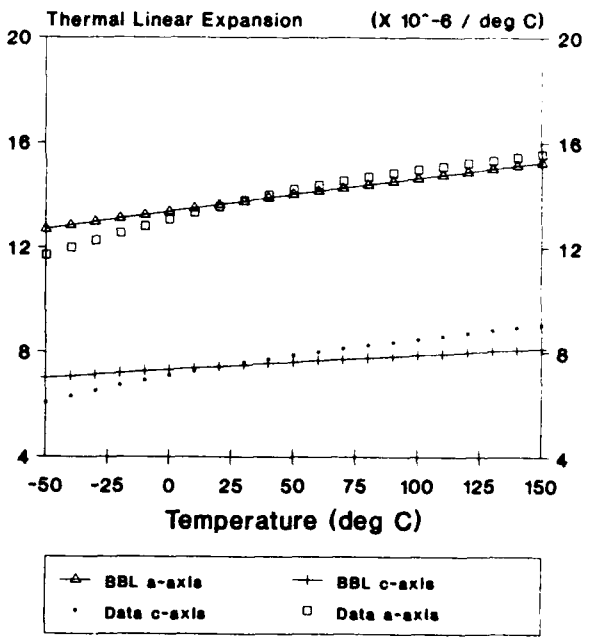


Fig. 3. Thermal expansion calculated using the thermal expansion coefficients of BBL [4] vs the HTMIAC data.

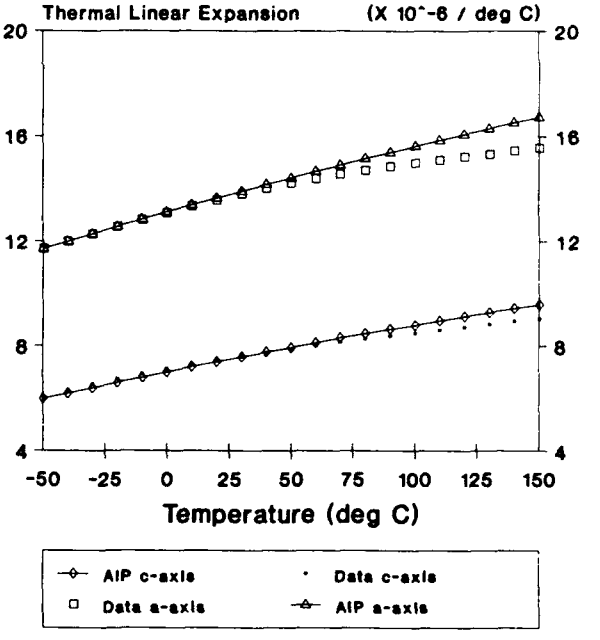


Fig. 4. Thermal expansion calculated using the thermal expansion coefficients of AIP [5] vs the HTMIAC data.

$$\alpha_{ij}^{(1)} = a_o + a_1\theta_o + a_2\theta_o^2 + a_3\theta_o^3 + a_4\theta_o^4 + a_5\theta_o^5$$
$$\alpha_{ij}^{(2)} = 1/2[a_1 + 2a_2\theta_o + 3a_3\theta_o^2 + 4a_4\theta_o^3 + 5a_5\theta_o^4]$$
$$\alpha_{ij}^{(3)} = 1/3[a_2 + 3a_3\theta_o + 6a_4\theta_o^2 + 10a_5\theta_o^3] \qquad (12)$$
$$\alpha_{ij}^{(4)} = 1/4[a_3 + 4a_4\theta_o + 10a_5\theta_o^2]$$
$$\alpha_{ij}^{(5)} = 1/5[a_4 + 5a_5\theta_o]$$
$$\alpha_{ij}^{(6)} = 1/6[a_5].$$

Note that an analysis of CTE $^{(n)}$ to order m corresponds to an analysis of $\alpha_{ij}^{(n)}$ to order m+1.

The HTMIAC results recast in thermal expansion form are listed in Table II. Most calculations of quartz resonator temperature behavior are carried out only to third order. If any one of the HTMIAC analyses is truncated to third order, the calculated results do not reproduce the HTMIAC data (see, e.g., Fig. 6). Because of this, a least-squares fit to the HTMIAC data to third order in α_{ij} was developed. This fit is called KGB and is plotted in Fig. 7. The values up to third order are listed together with BBL, AIP, and MB in Table III.

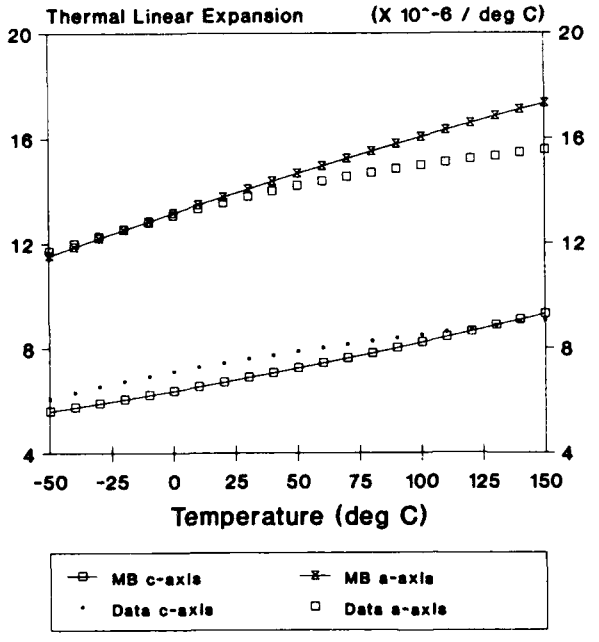


Fig. 5. Thermal expansion calculated using the thermal expansion coefficients of MB [6] vs the HTMIAC data.

Discussion

To understand the impact of using the KGB $\alpha_{ij}^{(n)}$ instead of, for example, using the BBL $\alpha_{ij}^{(n)}$, we have investigated the determination of material-constant

Table II

HTMIAC RESULTS IN THERMAL EXPANSION FORMAT

$\alpha_{11}^{(n)}$

(n)	Order of Fit 6	5	4	Units
1	13.673	13.668	13.659	X $10^{-6}/K$
2	11.231	11.299	10.926	X $10^{-9}/K^2$
3	-26.646	-23.834	-20.422	X $10^{-12}/K^3$
4	-18.034	-34.400	11.278	X $10^{-15}/K^4$
5	647.250	365.400		X $10^{-18}/K^5$
6	-1878.333			X $10^{-21}/K^6$

$\alpha_{33}^{(n)}$

1	7.508	7.513	7.515	X $10^{-6}/K$
2	7.686	7.653	7.746	X $10^{-9}/K^2$
3	-10.131	-12.502	-13.328	X $10^{-12}/K^3$
4	25.173	38.960	27.925	X $10^{-15}/K^4$
5	-325.825	-88.320		X $10^{-18}/K^5$
6	1583.833			X $10^{-21}/K^6$

Table III

THIRD ORDER THERMAL EXPANSION

$\alpha_{11}^{(n)}$

(n)	BBL	AIP	MB	KGB	Units
1	13.71	13.77	13.92	13.65	X $10^{-6}/K$
2	6.50	13.03	15.09	11.02	X $10^{-9}/K^2$
3	-1.90	-6.33	-7.86	-19.32	X $10^{-12}/K^3$

$\alpha_{33}^{(n)}$

1	7.48	7.48	6.79	7.50	X $10^{-6}/K$
2	2.90	9.41	8.69	8.00	X $10^{-9}/K^2$
3	-1.50	-5.44	6.88	-10.44	X $10^{-12}/K^3$

temperature coefficients (obtained by the resonator method) as well as the calculation of frequency-temperature behavior from the material constants.

The use of KGB thermal expansion coefficients instead of BBL coefficients will result in nontrivial differences when extracting temperature coefficients of the piezoelectric and dielectric constants from measured resonator frequency-temperature behavior. Values of the temperature coefficients of selected elastic constants for rotated Y-cuts extracted from measurements on rotated Y-cut resonators are given in Table IV. The previously reported [3] values using BBL thermal expansion coefficients may be compared to the values obtained using KGB thermal expansion coefficients.

The influence of the $\alpha_{ij}^{(n)}$ on the calculation of frequency-temperature behavior has been examined by using (1) in conjunction with the BBL set of material constants (with the exception of the BBL $\alpha_{ij}^{(n)}$). It is best to write (1) in the form [19]:

$$2\hat{T}_f^{(n)} = \hat{T}_c^{(n)} - \hat{T}_\rho^{(n)} - 2\hat{T}_h^{(n)} \qquad (13)$$

where

$$\hat{T}^{(1)} = T^{(1)}$$
$$\hat{T}^{(2)} = T^{(2)} - 1/2(T^{(1)})^2 \qquad (14)$$
$$\hat{T}^{(3)} = T^{(3)} - T^{(2)}\,T^{(1)} + 1/3(T^{(1)})^3.$$

If we then assume $T_{cB}^{(n)} = T_{cK}^{(n)}$, where B=BBL and K=KGB, then

$$[\hat{T}_{fB}^{(n)} - \hat{T}_{fK}^{(n)}] = 1/2\,([\hat{T}_{\rho K}^{(n)} - \hat{T}_{\rho B}^{(n)}] + 2[\hat{T}_{hK}^{(n)} - \hat{T}_{hB}^{(n)}]). \qquad (15)$$

The $\alpha_{ij}^{(n)}$ values found in Table III together with

$$T_\rho^{(n)} = -(2\alpha_{11}^{(n)} + \alpha_{33}^{(n)}) \qquad (16)$$

allow the calculation of temperature coefficients of density. These coefficients are given in Table V.

Similarly, the temperature coefficients of thickness for rotated Y-cut resonators may be calculated using the equation

$$T_h^{(n)} = \alpha_{11}^{(n)} \cos^2\theta + \alpha_{33}^{(n)} \sin^2\theta \qquad (17)$$

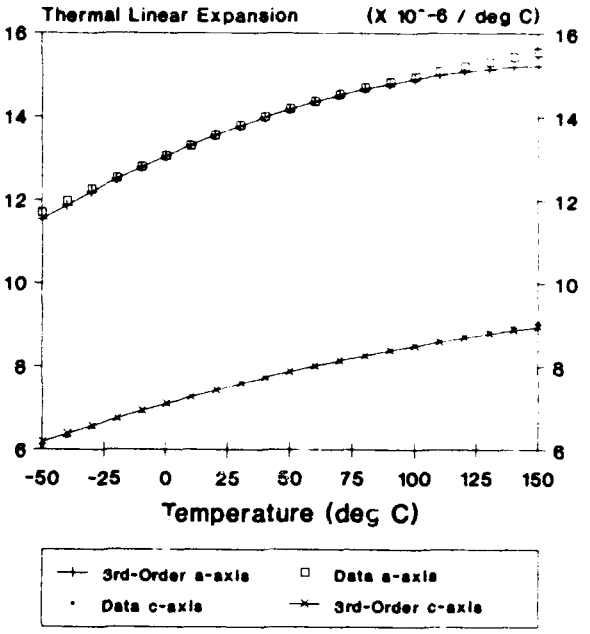


Fig. 6. Calculation of thermal expansion using 6th-order thermal expansion coefficients (truncated to 3rd-order) vs the HTMIAC data.

Table IV

TEMPERATURE COEFFICIENTS OF ELASTIC CONSTANTS ($T_{c_{ij}}^{(n)}$)

(n)	ij -> 66	44	14	Units
		BBL		
1	176.9	-170.4	103.9	X $10^{-6}/K$
2	163.8	-252.5	-19.4	X $10^{-9}/K^2$
3	43.9	-201.1	-509.1	X $10^{-12}/K^3$
		KGB		
1	176.9	-170.3	103.9	X $10^{-6}/K$
2	158.7	-256.6	-23.9	X $10^{-9}/K^2$
3	51.6	-177.3	-492.8	X $10^{-12}/K^3$

The temperature coefficients of thickness for AT-cut and BT-cut resonators are given in Table VI:

Table V

TEMPERATURE COEFFICIENTS OF DENSITY

(n)	KGB $T_\rho^{(n)}$	BBL $T_\rho^{(n)}$	KGB $\hat{T}_\rho^{(n)}$	BBL $\hat{T}_\rho^{(n)}$	Units
1	-34.80	-34.90	-34.80	-34.90	X $10^{-6}/K$
2	-30.04	-15.90	-30.65	-16.51	X $10^{-9}/K^2$
3	49.08	5.30	48.02	4.73	X $10^{-12}/K^3$

Table VI

TEMPERATURE COEFFICIENTS OF THICKNESS

AT-Cut

(n)	KGB $T_h^{(n)}$	BBL $T_h^{(n)}$	KGB $\hat{T}_h^{(n)}$	BBL $\hat{T}_h^{(n)}$	Units
1	11.60	11.63	11.60	11.63	X $10^{-6}/K$
2	10.01	5.30	9.94	5.23	X $10^{-9}/K^2$
3	-16.36	-1.77	-16.48	-1.83	X $10^{-12}/K^3$

BT-Cut

1	10.13	10.14	10.13	10.14	X $10^{-6}/K$
2	9.29	4.44	9.24	4.39	X $10^{-9}/K^2$
3	-14.23	-1.67	-14.32	-1.71	X $10^{-12}/K^3$

The differences in the temperature coefficients of frequency, density, and thickness obtained from (15) are listed in Table VII. A number of observations may be made. First, the differences in predicted first order temperature coefficient of frequency are very small. For example, from Bechmann [20] we know that:

AT-cut $\partial T_{fR}^{(1)}/\partial\Theta$ = -5.15 X $10^{-6}/K,°\Theta$

and for the

BT-cut $\partial T_{fR}^{(1)}/\partial\Theta$ = 2.14 X $10^{-6}/K,°\Theta$

Thus, the difference between the predicted AT-cut frequency-temperature behaviors is equivalent to an angular error of -14" of arc and the BT-cut difference is equivalent to an angular error of 1'07" of arc. Second, the differences in higher order temperature coefficients of frequency are

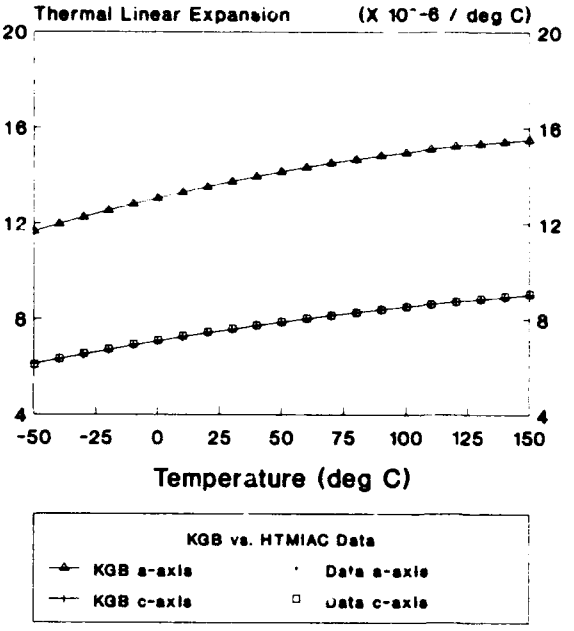


Fig. 7. Third-order thermal expansion coefficient least-squares fit to the HTMIAC data.

Table VII

DIFFERENCES IN TEMPERATURE COEFFICIENTS OF FREQUENCY, DENSITY, AND THICKNESS

(n)	$[\hat{T}_{fB}^{(n)} - \hat{T}_{fK}^{(n)}]$ AT	$[\hat{T}_{fB}^{(n)} - \hat{T}_{fK}^{(n)}]$ BT	$[\hat{T}_{\rho K}^{(n)} - \hat{T}_{\rho B}^{(n)}]$	$[\hat{T}_{hK}^{(n)} - \hat{T}_{hB}^{(n)}]$ AT	$[\hat{T}_{hK}^{(n)} - \hat{T}_{hB}^{(n)}]$ BT
1	0.02	0.04	0.10	-0.03	-0.01
2	-2.36	-2.22	-14.14	4.71	4.85
3	7.00	9.04	43.29	-14.65	12.61

Note: The usual units apply.

values of the differences in temperature coefficients of frequency for the AT-cut with the actual magnitudes of these coefficients as published by Ballato [21]:

AT-cut $T_{fA}^{(2)}$ = -0.45 X 10^{-9} $T_{fA}^{(3)}$ = 108.6 X 10^{-12}

Conclusions

An analysis of existing experimental thermal expansion data for quartz has been performed, yielding an improved average, set of higher order thermal expansion coefficients α_{11} and α_{33}. Comparisons with

previously used sets are given. The new values are required for redeterminations of the temperature coefficients of the elastic, piezoelectric, and dielectric constants of this important crystalline material.

References

[1] B. J. James, "A new measurement of the basic elastic and dielectric constants of quartz," Proc. of the 42nd Annual Frequency Control Symposium (AFCS), pp. 146-154, 1988.

[2] Jin-Feng Wang, Chen, Lei Zhang, and Zi-Kai Qin, "Determination of the dielectric, piezoelectric, and elastic constants of crystals in class 32," Phys. Rev. B, Vol. 39, No. 17, pp. 12888-12890, 1989.

[3] J. Kosinski, A. Ballato, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Temperature derivatives of the dynamic permittivity and permeability of the simple thickness modes of quartz plates," Proc. 42nd AFCS, pp. 53-64, 1988.

[4] R. Bechmann, A. Ballato, and T. Lukaszek, "Higher-order temperature coefficients of the elastic stiffnesses and compliances of alpha quartz," Proc. IRE, Vol. 50, No. 8, pp. 1812-1822, Aug 1962.

[5] R. A. Kirby, T. A. Hahn, and B. D. Rothrock, American Institute of Physics Handbook, ed. D. E. Gray, McGraw Hill, NY, 1972.

[6] M. Mizan and A. Ballato, "Stress-frequency coefficients of doubly rotated, low temperature coefficient, quartz resonators," USAERADCOM Technical Report DELET-TR-84-10, US Army Electronics Technology and Devices Laboratory, Fort Monmouth, NJ, p. 76 , Sep 1984.

[7] R. Sosman, The Properties of Silica, New York: Chemical Catalog Co., 1927.

[8] J. C. Brice, "Crystals for quartz resonators," Rev. of Mod. Phys., Vol. 57, No. 1, pp. 105-146, 1985.

[9] B. J. James, "Determination of elastic and dielectric properties of quartz," Ph.D. Thesis, p. 57, Royal Holloway and Bedford New College, Univ. of London, Spring 1987.

[10] P. D. Desai, "Critical analysis of alpha-quartz thermal expansion data," HTMIAC SPECIAL REPORT 40, Defense Electronics Supply Center Contract No. DLA900-86-C-0751, July 1990. HTMIAC is operated by CINDAS, Purdue University, 2595 Yeager Rd. West Lafayette, IN 47906.

[11] G. K. White, Cryogenics, Vol. 4, pp. 2-7, 1964.

[12] T. H. K. Barron, J. F. Collins, T. W. Smith, and G. K. White, " Thermal expansion, Grüneisen functions and static lattice properties of quartz," J. Phys. C: Solid State Phys., Vol. 15, pp. 4311-4326, 1982.

[13] A. H. Jay, Proc. Roy. Soc. (London), Vol. 142A, pp. 237-247, 1933.

[14] W. H. Johnson and W. H. Parsons, J. Res. Natl. Bur. Stds., Vol. 32, pp. 101-126, 1944.

[15] G. A. Lager, J. D. Jorgenson, and F. J. Rotella, API Conf., Proc., Vol. 89, pp. 75-77, 1982. (ISSN:0094-243X/82/890075-03$3.00 Copyright 1982 American Institute of Physics)

[16] H. G. Dorsey, Phys. Rev., Vol. 27(1), pp. 1-10, 1908.

[17] R. M. Buffington and W. M. Latimer, J. Am. Chem. Soc., Vol. 48, pp. 2305-2319, 1926.

[18] F. C Nix and D. MacNair, Rev. Sci. Instr., Vol. 12, pp. 66-70, 1941.

[19] A. Ballato and T. Lukaszek, "Higher-order temperature coefficients of frequency of mass-loaded piezoelectric crystal plates," Proc. 29th AFCS, pp. 10-25, 1975.

[20] R. Bechmann, Frequency-temperature-angle characteristics of AT- and BT-type quartz oscillators in extended temperature range," Proc. IRE, Vol. 48, No. 8, p. 1494, Aug 1960.

[21] A. Ballato, "Doubly rotated thickness mode plate vibrators,"Physical Acoustics, Vol. 13, Chapter 5, pp. 115-181, 1977. Academic Press, New York.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INFLUENCE OF CONVECTIVE FLOWS ON THE GROWTH OF SYNTHETIC QUARTZ CRYSTALS

V.A.Klipov, N.N.Shmakov

Research Department of FONON in Yuzhnouralsk,
Cheliabinsk Region, 457040, USSR

The influence of concentrative and convective flows of the crystal growing medium on the shape, morphology of the surfaces, enternal structure, growth rate, macroscopic defects and inclusion density was investigated. Dependance of above mentioned characteristics on the convective fiow direction and crystal's place within the flow was determined.

Experimental

Quartz crystals were grown in the autoclaves with inside volume 1.5-7 cubic meters. The (0001) cut seeds (Z-plates and Y-bars) as well as the $(01\bar{1}1)$ and $(11\bar{2}0)$ cut seeds (r-slices and x-slices accordingly) were used. The growth process has utilized alkaline water solution confined within autoclave at the temperature under 400°C and pressure from 70 to 120 MPa. The method of temperature gradient was used.

Morphological peculiarities were studied visualy and, also, by taking some of the specific measurements either of the every of the crystals grown in the crystallization chamber or of the crystals of some chosen cross sections.

The internal structure of the crystal was investigated by the shadow projection method [1], Lang's X-ray projection topography method [2] and by the gamma irradiation of the crystals. To determine the influence of concentrative flow only, on the shape and morphology of the growing crystal faces a special metalic box with double perforated sides was used. Smaller box had the 0.8 mm holes in quantity of 25 per square cm and was placed with some clearance inside the bigger one having 1 mm holes in quantity of 5.3 per square cm. This double box, with the seed suspended inside it, was placed in the growth zone of the autoclave. Therefore, the main convective current in the autoclave as well as various solid particles carried by it, could not influence the crystal growth. One of the crystals grown in such a box is shown in Fig.1.

Wedge-shapedness of the crystal

It was established that the ascending concentrative flow is generating the wedge-shapedness of the crystal. Lower part of the crystal is growing faster (in both X and Z directions) than the upper end. Besides, a directed accessory relief is emerging on the basic surface, and on the edges between the basic and rhombohedral faces a border projections are revealed. These border projections are serrated often; sometimes, with cracks in their growth area.

Futher, many crystals grown outside of mentioned perforated box were investigated. It was revealed that almost all of them became wedge-shaped. Some of them grew narrower in their upper end direction, the others, on the contrary, in the lower end direction (Fig.2).

There is no doubt that both concentrative and convective flows influence in the same manner on the shape of the crystals of various compounds [3]. Now, we can affirm that synthetic quartz crystals grow narrower in the upper part of them in ascending convective flow, and vice versa the upper part of the crystal grows faster in discending flow.

To estimate the crystal's vertical wedge-shapedness a following equations were used

$$Ko = \frac{H_1 - H_2}{L - (l_1 + l_2)}; \qquad K = \frac{H_1 - H_2}{L}$$

were Ko and K are the exact and approximate (accordingly) linearly interdependent wedge-shapedness coefficients, and H_1 ,H_2 ,L, l_1, l_2 are special dimensions as shown in the Fig.2.

CH2965-2/91/0000-029 $1.00 © 1991 IEEE

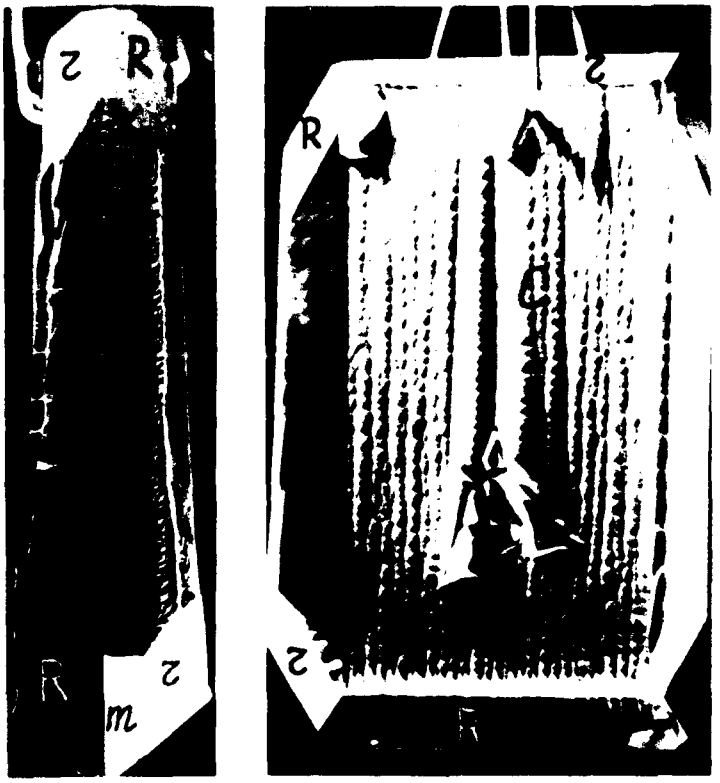
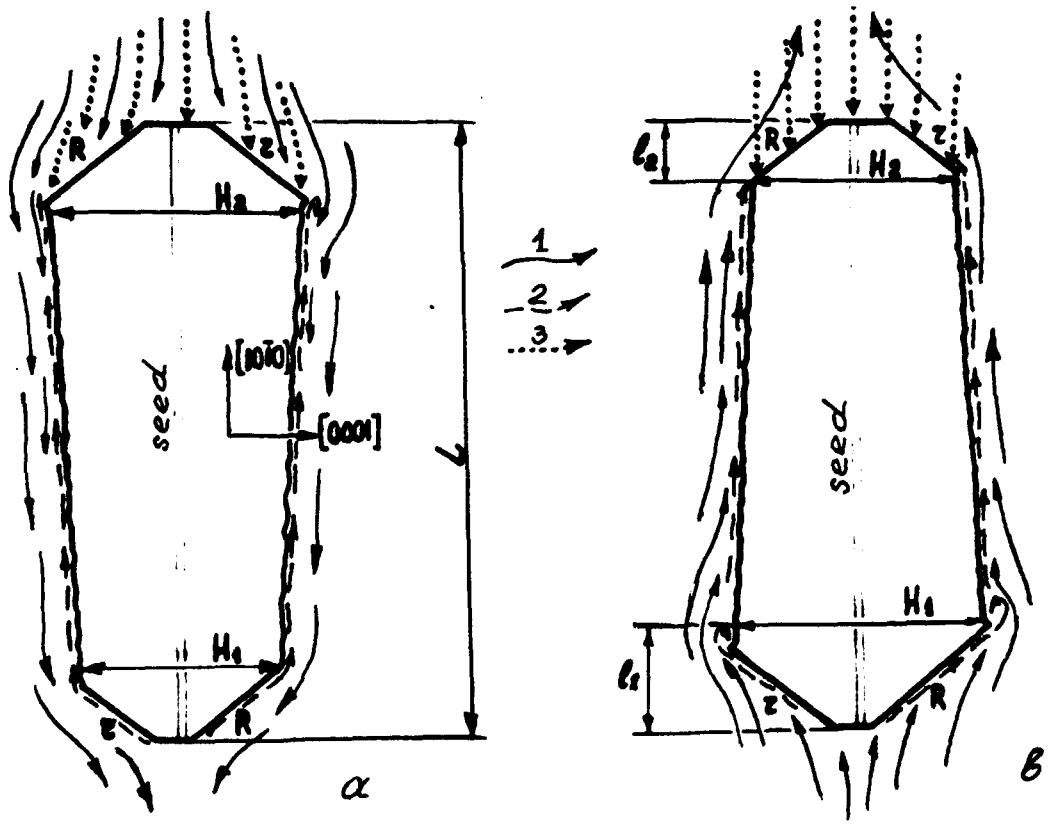


Fig.1. Quartz crystal grown inside the double box. Sideview and view from the side of basic surface.

Fig.2. Scheme of the discending (a) and ascending (b) convective flows 1 passing along the crystal. 2 - concentrative flows. 3 - solid particles settling upon crystal surfaces.



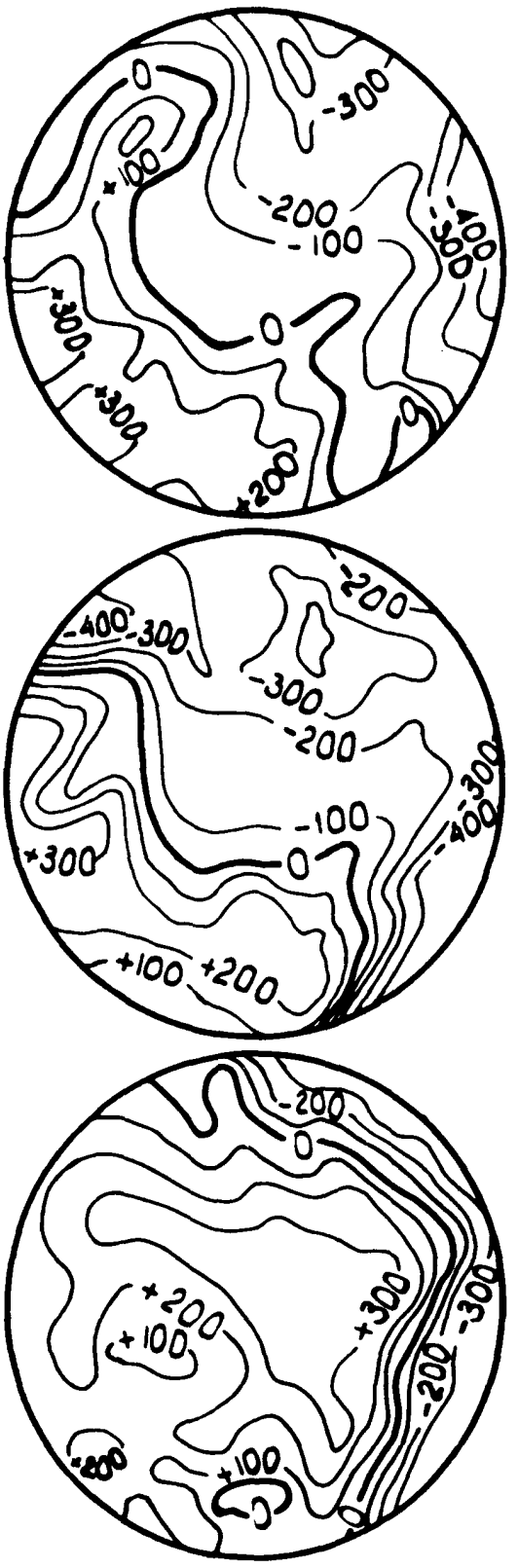


Fig.3. Distribution of the coefficient of the crystal wedge-shapedness, $Kx10^4$, in the cross-sections of cristallisation chamber. Positive isonephs show the field of ascending convective flow, negative - discending flow.

Fig.3 shows, by means of isonephs, the distribution of K in the upper, middle and lower cross sections of the autoclave's cristallisation chamber in the case of synthetic quartz crystals grown on the Z-cut seeds. A zero isoneph cor-responds with the crystals showing no wedge-shapedness and marks a border between ascen-ding and discending flows.

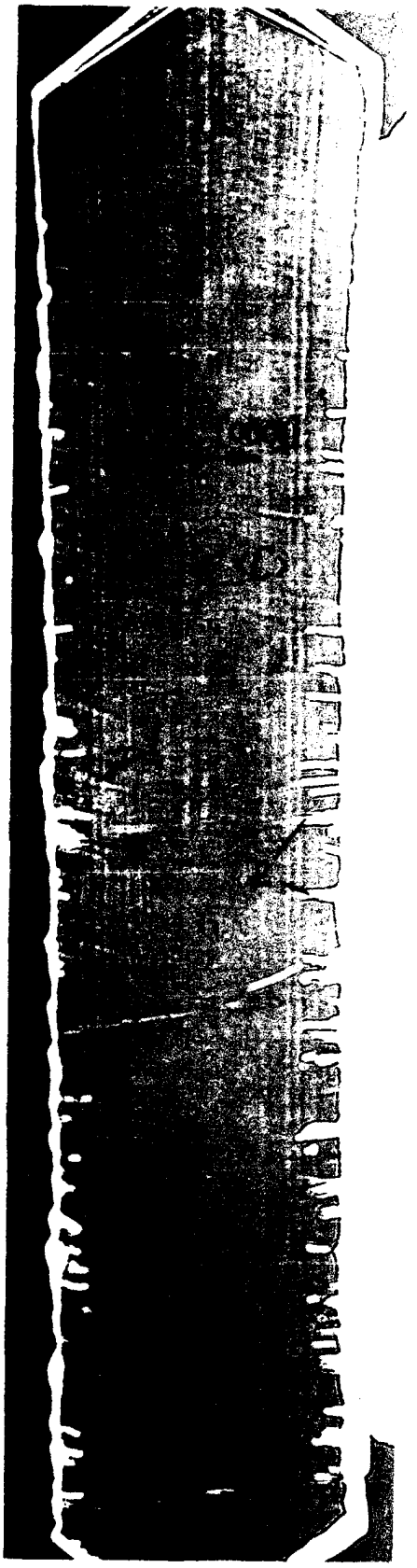
Stability of the convective flow

To answer the question of convective flow stability during the growth time a special investigation was accomplished. Some of the (1120) slices were cut out along the crystal's length from the crystals grown in various parts of an autoclave. Futher, the shadow projections of them were investigated.

It was revealed no slice showing the wedge-shapedness direction change during the crystal growing time, i.e. every of the growth layer throughout the all Z-zone (in figures shown as <C>) is narrowing in the same direction.

Inside structure and defects in the crystal

The inside structure of the Z growth zone and morphology of the (0001) surface was investigated. In the direction of the convective flow motion was revealed the formation of "parasite" growth zones (Fig.4) and connected with them cracks. On the basic surfaces gradualy change of accessory types or increasing of the accessory's dimensions is taking place (Fig.5). On the crystal's end facing the current were grew the border projections on the edges. The internal structure of border projections was investigated by the Lang's method and, also, by the gamma-irradiation of slices of various cuts (Fig.6). It was found that serrated border projections intrude into Z growth zone and have complicated structure because of having their own growth zones and defective tense surfaces. The growth of the serrated border projections is accompanied



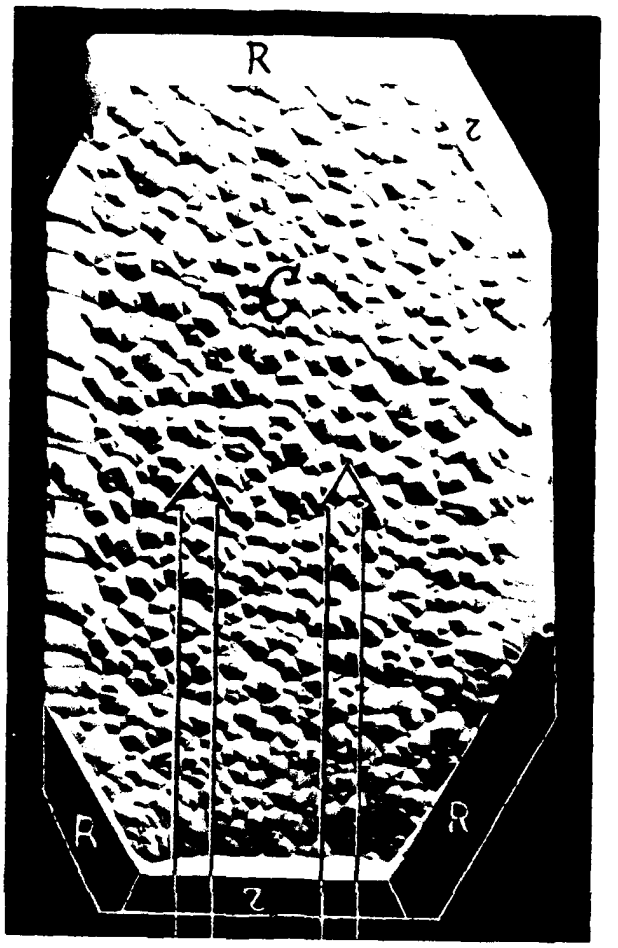


Fig.5. Increasing of the basic surface accessory's dimensions in the direction of the convective flow motion and formation of the border projections on the edges r/C and R/C of lower end of the crystal

Fig.4. Inside structure of the crystal grown inside the discending convective flow and detected by [1] method. 1-"parasite" growth pyramids; 2-cracks.

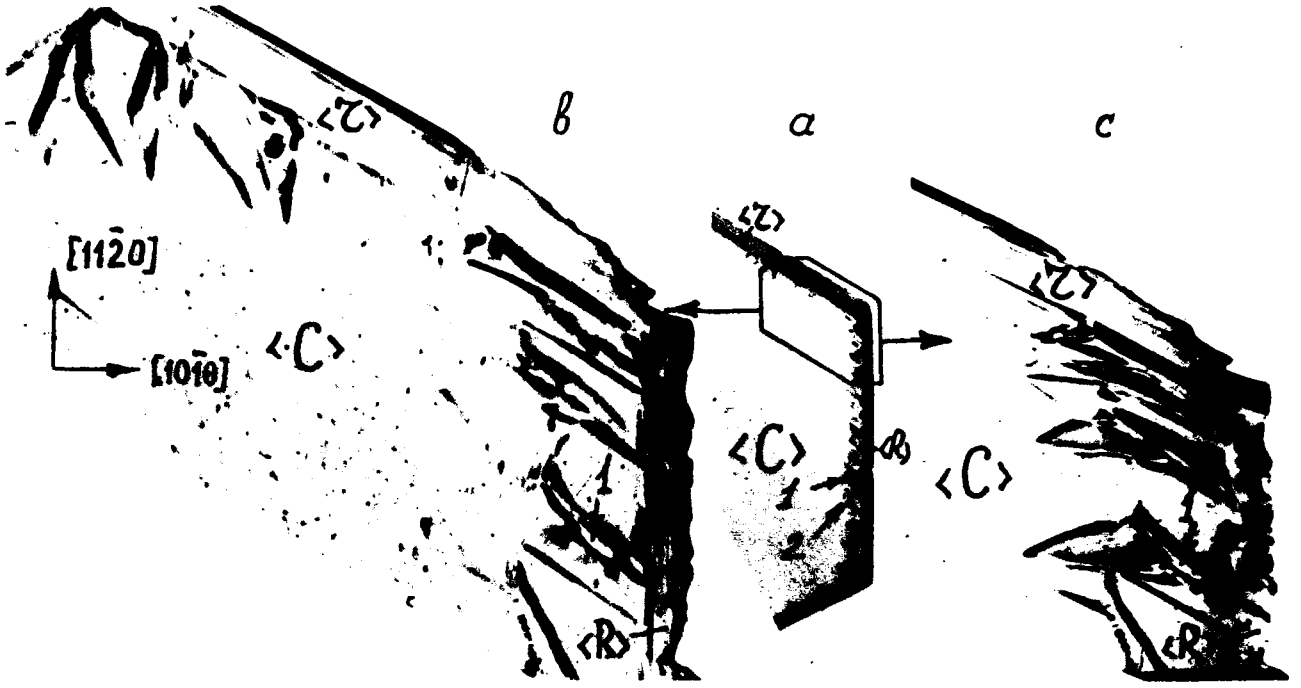


Fig.6. Morphology of the serrated border projections. a- gamma-irradiated Z-cut slice; b- X-ray photograph of the part of (0001) slice, reflex from (10$\bar{1}$0) under the R; c- reflex from (10$\bar{1}$0) under r, Mo K_α ,8^x; 1-2 - growth cracks connected with bubbles.

by the formation of the two types of the different orientated flat morphological cracks opened to the surface or represented only as a series of separated bubbles (Fig.7).

Crystal growth rate

It was investigated the dependence of growth rate on the position of the growing crystal within the cross section of the convective flow. Fig.8 shows the distribution of the mean Z dimensions (calculated as $(H_1+H_2)/2$) in some of the cross-sections of the crystallization chamber. Zero isoneph shows the border between opposite flows (cross-sections are the same as in Fig.3).
It was found that most uniform crystals grow inside the ascending convective flow. On the contrary, the greatest variety in the growth rate



Fig.7. Morphology of the cracks and bubbles of first (1) and second (2) types.



Fig.8. Distribution of the crystal mean Z-thickness in some cross-sections of the cristallization chamber.

revealed itself inside the discending flow's cross section. It is more noticeable in the lower part of the crystallization chamber. The fields of minimum growth rate are at the border between ascending and discending flows, especialy at the areas of its harsh twisting. Exactly these crystals reveal serrated border projections on the edges between the rhombohedral and basic surfaces and, also, the cracks connected with them and located inside their growth zones and Z growth zones and, also, between the <0001> and <11$\bar{2}$0> growth zones.

Dencity of twin accessories on (10$\bar{1}$1) and inclusion density

It was established that the densities of twin accessories upon great rhombohedron surfaces of the crystal grown inside the ascending and discending flow are different. The greatest density of twin accessories (as much as 41.6 cm^{-2} in one of the experiments) was found upon upon faces of the crystals grown inside of ascending flow, the lowest values (10.5 cm^{-2} in the same experiment) were corresponding with discending flow mainly near of autuclave sides. It turned out that the dencity of twin accessories changes gradualy from ascending to discending flow if observing within one of the horizontal cross sections of the crystallization chamber.

The dencity of inclusiones was counted in accordance with the last IEC Publication. It was found inclusion dencity increasing from ascending to discending flow (Fig.9).

Results and discussion

Theoretical principles of interconnection between the growing crystal and surroundings were laid by P.Curie [4], in 1894, in his Universal Principle of Symmetry. In conformity with our investigation it means that the growing crystal preserves in its appearance those symmetry elements only, wich concur with the symmetry elements of nutrient medium. So, if the vertical current of soluted nutrient having

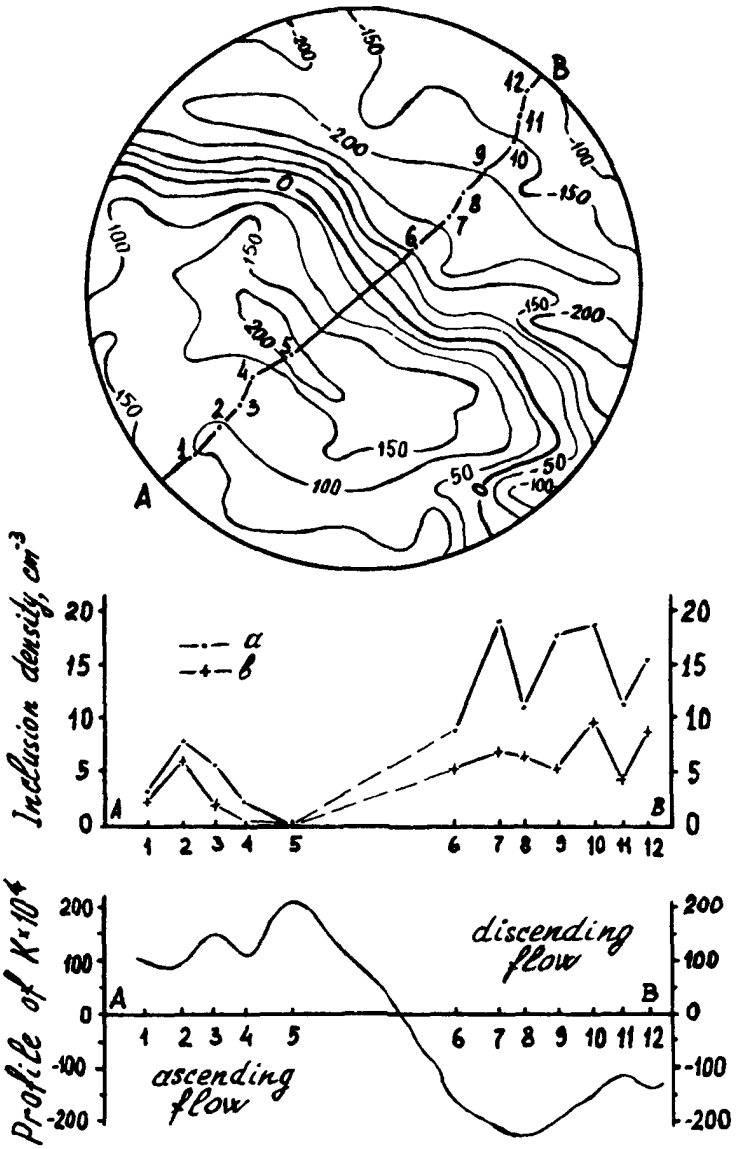


Fig.9. Inclusion dencities in crystal grown inside the ascending and discending flows. Isonephs show the distribution of Kx10^4 wedge-shapedness coeficient. 1-12 - are the positions of the samples in the A-B profile; a- general inclusion density (sizes from 10 to 125 microns); b- dencity of the inclusions wich are lesser than 30 microns.

(in gravitational field) a symmetry of cone ($L_{\infty}\infty P$) is interacting with synthetic quartz crystal having $L_2 2P$ external symmetry the latter has to preserve the symmetry elements common with those of the flow. That common element, in this case, is a vertical flatness of the external symmetry of the crystal coincided with the seed flatness and vertical flatness of the flow. Horizontal flatness of external symmetry and the L_2 disappear. This phenomena reveals itself in the appearance of the wedge-shapedness in the shape of the crystal grown on the ZYb/0° cut seed.

From a position of crystal growing in the directed current of nutrient solution, it ought to be marked that in more advantageous conditions are the surfaces and edges of a crystal orientated to meet the flow. As the nutrient solution is passing along the crystal surfaces its oversaturation is decreasing gradualy, and as a consequence, decreases the growth rate of the surfaces, and parasite pyramids and connected with them cracks and changing in the relief of surfaces take place. The generation of the serrated border projections is determined, probably, as in the case with KDP crystals [5], by the existence of some critical velocity of the current passing along the crystal or formation of the closed vortexes on the border between the opposite convective flows wich is proved by existence of minimum growth rate in this area.

Natural changing of the twin accessory density upon the upper great rhombohedron faces is connected, probably, with character of solution motion bypassing it. The discending flow (see Fig.2) does not assist in holding of microcrystals on the crystal surfaces , but quite the reverse, it is washing them away. In the case of ascending flow, the upper surfaces are in the "shadow" of the passing flow, and the microcrystals are deposited on them easily.

The cause of the inclusion dencity increasing in the crystal grown in the discending convective flow is, probably, in the turbulent or even vortical manner of solution motion near the crystal surfaces where the counteraction between concentrative and convective flows is taking place. The time of the particles being near the crystal surface is longer (and consequently increases the probability of their holding by the crystal) than that in the case of ascending flow when concentrative and convective flows are passing in the same direction and their resulting speed increases.

Conclusions

The shape of the real quartz crystal as well as relief of it surfaces and faces, its homogenity and structure are the result of the interaction between the structure of the crystal and that of the crystal growing envirement. Under the influence of the flows of nutrient medium the crystals (grown on the seed of every cut) become wedge-shaped. Greater growth rate have surfaces facing towards the current movement.

With the direction of the wedge-shapedness correlate particularities of the surface relief and homogenity of the growth layers as well as border projection formation.

On the basis of account of the morphological particularities and enternal structure of the crystals a method was eleborated intended to determine the current motion direction and structure of the convective flow.

A stable convective circulation of the solution in the autoclave determines the crystal growth rate, defects formation as well as dencity of of the inclusions and accessories on the upper surfaces of the major rhombohedron.

The mechanism of interaction between the crystal and crystal growing medium is highly complicated, and there is much enough to investigate yet.

We believe that investigation of the factors influencing the formation of convective flows in autoclave as well as their control is also worth of discussion, but it is in our following paper.

References

[1] Доладугина В.С. Приборы и методы исследования однородности кварца.- В кн.:Рост кристаллов,т.3,М., изд. АН СССР,1961, с. 481-494.

[2] Lang A.R. the projection topography, a new method in X-ray microradio-graphy. - Acta Cryst., 1959, 12.

[3] Мокиевский В.А. К вопросу об искажении облика природных кристаллов.- Зап. ВМО,1953, ч.82, вып.3, с.173-184.

[4] Кюри П. Избранные труды.- М. - Л., "Наука", 1966, 399 с.

[5] Janssen-van Rosmalen R. e.a. The Influence of the Hydrodinamic Environment on the Growth and the Formation of Liquid Inclusions in the KDP Crystals. - Krist.Technik, 1978, 13, 1, pp. 17-28.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SWEEPING AND IRRADIATION EFFECTS ON HYDROXYL DEFECTS IN CRYSTALLINE QUARTZ

Harish Bahadur,
National Physical Laboratory,
Hillside Road, New Delhi–110 012 (India)

ABSTRACT

A study of electrodiffusion (sweeping) and irradiation effects on hydroxyl defects has been carried out on good quality optically clear natural quartz crystals. The crystals were of Arkansas and Brazilian origin which are generally used as starting material in hydrothermal synthesis of high quality cultured quartz crystals. In particular, various OH^- related point defects have been monitored using infrared absorption measurements in the 3100-3700 cm^{-1} range. Our estimates from the irradiation results show that the Brazilian crystal had relatively much higher concentration of hydrogen than the Arkansas crystal. We have therefore designated the Brazilian sample as 'high-H' and Arkansas as 'low-H' quartz.

High- and low-H samples of natural quartz in three different conditions, unswept, Na-swept and H-swept, were subjected to a sequence of low- and room -temperature irradiations that separate proton and alkali motion. The results are discussed in terms of proton and alkali motion to shallow and deep traps and compared with cultured quartz.

INTRODUCTION

Quartz oscillator crystals play a significant role as precision oscillators for a variety of frequency control and timing devices. In fact, in the present day technology for various kinds of precision equipment the name 'quartz' has become the symbol of quality. In recent years, studies on radiation effects in crystalline quartz have received considerable attention due to their use in flying clocks for space satellites. It is now believed that a variety of impurities present in quartz crystals, both natural and synthetic (also called 'cultured quartz'), incorporated during their growth are responsible for the performance of oscillator crystals in a radiation environment [1-3]. At present, most of the industrial demands on quartz crystals are met by the different grades of commercially available cultured quartz except for those where a radiation-resistant quartz is required for various aerospace applications. Therefore, radiation effects in quartz is a subject of considerable interest both from fundamental as well as applied aspects. Also, it will be of interest to ivestigate the role of different impurities present in natural quartz which in most cases is the starting material for growth of high purity cultured quartz.

Various interesting point defects are found to occur in quartz crystals [2,4,5-11]. Aluminum is the most pervasive impurity in quartz and is present as Al^{3+} at substitutional sites of Si^{4+}. The charge compensation at Al^{3+} is provided by one of the monovalent impurities such as Li^+, Na^+, K^+, or H^+ or an electron hole. These centers are respectively designated as Al-M^+ where M^+ is either Na^+, Li^+ or K^+ and Al-OH^- in case of an interstitial proton compensating the negative charge of aluminum; the protons forms an OH^- molecule with the adjacent oxygen. The Al-hole centers are designated as Al-h^+ or $[Al_{e+}]^o$. A variety of techniques are used to characterize the point defects in quartz. For example, Al-Na^+ centers have a strong acoustic loss peak at 53K and a weaker peak at 135K. Al-hole and oxygen vacancies and some of the Ge-related centers are detected by using electron spin resonance. Martin [8] has shown that the acoustic loss peaks at 23K, 100K and 135K are related to the production of Al-hole center.

In addition to the Al-associated centers there are other types of point defects also present in quartz [12]. They include, for example, oxygen vacancies, Ge-associated centers and numerous OH^- related defects. Most of the OH^- related defects are the stretching vibrations of OH^- molecules trapped at a variety of unknown sites and are detected by infrared absorption measurements. The OH^- intensities are very sensitive to the temperature at which they are monitored. In fact, at room temperature most of the bands due to temperature broadening are beyond recognition. It is therefore imperative that measurements be made at 77K or lower if defect bands are to be monitored. A considerable amount of work has already been done to date but there is no clear understanding of the nature and origin of different OH^- related bands. To date, only the bands at 3367 and 3306 cm^{-1} have been attributed in literature to be due to protons trapped at oxygen sites around an Al^{3+} substituted Si^{4+} in the quartz lattice.

Studies on near infrared absorption prior to Kats [13] were inconsistent and speculative. Kats [13] in his classic study of infrared absorption in

CH2965-2/91/0000-037 $1.00 © 1991 IEEE

quartz has reported numerous OH^- related absorption bands in a variety of natural and synthetic crystals and attributed some of the bands due to specific alkalis associated with them. Dodd and Fraser [14] corroborated Kats' work and found a direct relationship between the 100°C anelastic relaxation peak in 5 MHz resonators and the hydrogen bonded to broad OH^- absorption in as grown cultured quartz. While the number of absorption bands in natural quartz can be quite high and their relative strengths are uncorrelated [13,14], the cultured or synthetic quartz has always a set of four principal bands associated with as-grown OH^- impurities with peaks at 3348, 3437, 3400 and 3581 cm^{-1}. These intrinsic bands in cultured quartz are called growth-defect bands and their strength has been found directly proportional to the growth rate of the crystals [15]. Often, these bands are designated as s_1, s_2, s_3, and s_4 respectively in order of wavenumbers [16-22]. The Al-OH^- bands which are produced by irradiation or electrodiffusion in hydrogen or air, positioned at 3306 and 3367 cm^{-1} are designated as e_1 and e_2. Brown and Kahan [23] studied the effect of irradiation and electrodiffusion on various specimens of natural and cutured quartz and confirmed the proposal of Kats [13] that the bands in crystalline quartz located at 3306 and 3367 cm^{-1} arise due to the location of hydrogen around substituted Al^{3+} for Si^{4+} in quartz lattice. The paper of Brown and Kahan [23] also gives a comprehensive background of the previous work on infrared absorption in quartz crystals. Krefft [24] carried out a correlation study of OH^- absorption bands and the visible-ultraviolet coloration bands before and after electrolyzing the crystals at high temperature under vacuum; thus apparently after sweeping the holes into the crystals.

Sibley et.al.[25] extended the investigation on 3367 and 3306 cm^{-1} bands and included the effects of irradiation and annealing of high quality cultured quartz. These authors [25] reported that following an intense irradiation at 77K the Al-OH^- bands are nearly equally populated; and retain their relative strength up to 230K during annealing. Above 230K a reorientation takes place in the temperature region 230-270K. They found that while the band at 3367 cm^{-1} continues to grow uptil 300K and saturates the 3306 cm^{-1} band decreases up to 300K and saturates. Sibley et.al. [25] interpreted this to mean that these bands arise from OH^- molecules in two different but closely related sites relative to aluminum ions and that a transfer from site to the other takes place in this temperature region. Halliburton et.al. [5,26] have recently obtained additional evidence supporting the Al-OH^- model for the 3306 and 3367 cm^{-1} bands from a parallel study using electron spin resonance (ESR) and infrared absorption (IR) techniques.Lipson et.al.[16-20] studied the production behavior and polarization characteristics of the four principal (s_1-s_4) bands in a variety of cultured quartz samples both unswept and swept versions of different grades of commercially available stones. Later, these workers [21,22,27] have extended their investigations in evaluating the inhomogeniety into the crystals by nonuniform growth rates. The peaks at 3581 cm^{-1} and 3367 cm^{-1} were used to monitor respectively the as-grown OH^- and Al-OH^- point defect centers [19]. Lipson and Kahan [22] observed large variations in substitutional impurity concentrations along the crystal growth axis. Subramaniam, Halliburton and Martin [28] investigated the effects of irradiation at 80K, 196K and 300K in a variety of synthetic quartz. In addition to the well known as-grown s-bands and radiation induced e-bands, these authors [28] found a series of five bands at 3493, 3503, 3557, 3547 and 3603 wavenumbers produced after an intense irradiation at 80K. The new set of bands was labled as r_1, r_2, r_3, r_4, and r_5 bands respectively. Irradiation at 196K produced only r_1, r_4 and r_5 but not $r_2$and $r_3$bands. Room temperature irradiation did not produce any of the r-bands but instead produced the well known e_1 and e_2 (Al-OH^-) bands at 3306 and 3367 cm^{-1}.These authors [28] in their studies included the effect of irradiation on s-, r and e-bands and investigated their thermal annealing behavior. They found that all the four s-bands decrease in their intensity upon irradiation at low temperature and recover to their original strength upon annealing to nearly 760K. The r-bands showed their initial growth with increasing radiation dose and saturation on extended irradiation except the r_4 band at 3547 cm^{-1} which continued to grow uptil the maximum dose of 1.8 x 10^{16} MeV/cm^3. Thermal anneals of r-bands showed that the r_2 band anneals out completely by 400K; the other r-bands show a marked decrease in the 600-650K region. The r_4 and r_5 anneal out almost completely and the r_1 and r_3 do not completely vanish.

In a recent study, Bahadur [29] has shown that in a variety of natural quartz of different origin a set of major bands always present is at 3476 cm^{-1}, Al-OH^- at 3306 and 3367 cm^{-1}. A number of other OH^- related bands may or may not be present in conjunction with the major bands. While the Al-OH^- bands are present in the natural quartz in their as-received state in significant strength, these bands in cultured samples are produced only after either electrodiffusion (sweeping) in air or hydrogen or irradiation by ionizing radiations [23-25,30]. On the other hand, electrodiffusion or irradiation in natural quartz increases the strength of Al-OH^- centers in the limit to saturate all substitutional aluminum compensated with hydrogen [29]. Except for the Al-OH^- bands, defects responsible for the other infrared bands have not been identified in terms of any specific atomic model. In cultured samples, recently, Martin et.al.[31] have reported that these defects can also trap alkalis. While the hydrogen trapped at these "growth-defect" sites does not cause any anelastic

loss, lithium or sodium trapped at the site produces loss peaks at 305K or 335K respectively. These loss peaks grow rapidly with irradiation and then show some decay at higher radiation doses. For samples with small OH^- -growth defect to aluminum ratios the decay of these loss peaks is nearly complete while for samples with relatively large OH^- -growth defect to aluminum ratios the peaks show only a small decay. The OH^- related growth defects are the source of hydrogen that is transferred to aluminum during the irradiation process. The "growth-defects' also seem to act as one of the traps for the alkalis that are released from the aluminum.

Radiation-induced mobilities of alkali ions and protons in quartz are the most direct effects that are apparently responsible for the performance of a quartz oscillator in a radiation field [5,7, 32-45].While the protons under radiation field are mobile at all temperatures, the movement of alkalis starts only about and above 200K [5,7,25,46]. If quartz crystals are irradiated above 200K, the $Al\text{-}M^+$ centers are converted into a mixture of $Al\text{-}OH^-$ and Al-hole centers. The relative concentration of production of Al-hole centers depends upon the number and nature of available electron traps. On irradiating quartz at low temperatures (77K), hydrogen atoms are formed due to dissociation of OH^- molecules followed by a displacement of hydrogen to an isolated interstitial site where it becomes stably trapped [47]. Also, in case of natural quartz where $Al\text{-}OH^-$ centers are present in the as-received unswept state or H-swept or irradiated cultured quartz, the low temperature irradiation as in cultured crystals [4,5,7] would convert the $Al\text{-}OH^-$ centers into Al-hole centers.

Previous studies on infrared characteristics in natural quartz [13,14,23] were confined only to the irradiation at room temperature and electrodiffusion. Recently, only limited information on irradiation effects in an unswept Brazilian quartz were published [29] as a function of irradiation at room temperature and low temperature. In this paper we present results of our investigations on the radiation induced ion movements in natural quartz as a function of H-concentration (high and low), the sample condition (unswept or Li-swept, Na-swept and H-swept) and irradiating temperature; 77K, room temperature and final irradiation at 77K.

EXPERIMENTAL DETAILS

Two natural stones, one from Arkansas and the other from Brazil used in the present investigation. Samples of matching thickness were prepared such that the c-axis was always perpendicular to the incident beam of light. In this case the electric vector $\vec{E}$ of the incident light was always parallel to the c-axis. All spectra were recorded for $E \perp c$. After the crystals were cut and polished, they were mounted in an optical Dewar equipped with CaF_2 windows and capable of reaching $\pm$ 2K. Near-infrared absorption was measured with a Beckman 4240 double beam spectrophotometer in which ratio recording of sample and reference signals was done electronically. Crystals were irradiated first at 77K, warmed to room temperature, and finally irradiated at 77K. The irradiation was carried out by a Van de Graaff electrostatic accelerator. The acceleration voltage used was 1.75 MeV and the current density on the sample was 10 uA/cm^2.The distance of the sample in cryostat and the window of the accelerator was 5 in.; this setting was used at each step of irradiation. Time of irradiation was 2 min every time. The irradiation time was controlled by a Gralab model 900 programmable timer. This irradiation deposited a dose typically around 2 Mrad. Samples were mounted in the optical Dewar and irradiated in the Dewar itself without demounting. The radiation port for the cryostat consisted of a 5-mil aluminum window.

Samples were swept by Na, Li and hydrogen. Prior to sweeping, the samples were given an optical finish. Our sweeping process began by vapor depositing gold electrodes on the two samples faces. If Li or Na was to be swept the appropriate salt was first evaporated on one side, then a gold electrode was deposited over the salt. The electrodiffusion was carried out at a temperature of 490° C. Alkali sweeps i.e., taking alkalis into the crystal, were done in a vacuum while an H_2 atmosphere was used for sweeping protons into the crystal. The applied electric field was approximately 30V/cm for alkali sweeps and 1500 V/cm for hydrogen. A digital multimeter and a laboratory computer were used to monitor the sweeping current as a function of time. Near the completion of sweeping, we believe only one species of ion was being transported and the sweeping current reached a steady state; this usually took about 20-24 hours for a hydrogen sweep and about two hours for alkali sweep. Once the sweeping was complete the furnace temperature was slowly programmed down. During the cool down process the electric field remained still on to prevent back diffusion of alkali ions or protons.

RESULTS AND DISCUSSION

Figure 1 shows a representative case of near infrared absorption spectra of unswept, Na-swept and H-swept condition of a natural quartz. Out of a number of absorption bands shown in Figure 1 only the bands for which a model has been suggested are at 3306 and 3367 cm^{-1}. Kats [13] first suggested that these bands are due to an OH^- bond adjacent to substitutional aluminum and the subsequent research on natural as well as cultured quartz [5-11, 14-25,29,30] have corroborated Kats' model [13]. The bands at 3306 wavenumber represents the proton trapped at the short-bond oxygen and the proton trapped the long-bond oxygen shows absorption at 3367 wavenumber. In the text to follow the 3306 cm^{-1} band will be represented as Al_s and the 3367 cm^{-1} as Al_l; the suffixes s and l respectively denote as short and long bond $Al\text{-}OH^-$ center. The other spectral bands may be attributed to impurity elements which have (i) alkalis and hydrogen together, (ii) only hydrogen and (iii) isolated minor lines e.g. the one at 3595 wavenumbers present in all swept conditions. In the alkali

swept conditions four separate groups of lines are observable and in each group there is one major and in each group there is one major and several minor lines while in the hydrogen-swept condition the situation degenerates to two groups having one major line and several minor lines in each group. These group numbers are arbitrarily marked in the Figure in increasing order of wavenumbers.

In natural quartz of different origins, it has has been found that the bands always present in significant strength are at 3476, 3367 and 3306 cm^{-1}. While the latter two are the known $Al\text{-}OH^{-}$ defect centers, the 3476 cm^{-1} has been found [29] to be associated with the presence of Li and H ions together. Na-sweeping results in a prominent band at 3451 cm^{-1} along with $Al\text{-}OH^{-}$ centers. In the H-swept condition the major band positions occur at 3468 cm^{-1} and the $Al\text{-}OH^{-}$ bands. Henceforth, these major bands dependent on alkali and H or H-sweeping will be given a generic designation e.g. 3476 cm^{-1}; Li and H dependent will be designated as Li-n3a, the 3451 cm^{-1}: Na and H dependent as Na-n2a and the 3468 cm^{-1} induced by H-sweeping as H-n2a where n stands for natural quartz, group number 1 to 4 (or 1 to 2 for H) by increasing wavenumber letter for the lines within each group by decreasing intensity and a for the strongest line in the group. Each line group may possibly be attributable to defects associated with one of the four inequivalent oxygen positions. With H-sweeping induced defects the situation is degenerate resulting in two groups as one finds it also with the $Al\text{-}OH^{-}$ defects. The smaller lines within each group signify different kinds of local displacements essential to particular point defect species.

Figure 2 shows the near infrared spectra in unswept or as-received condition of two crystals for which the irradiation results will be presented later. The crystals as mentioned earlier are of Brazilian and Arkansas origin. It may be seen that the two crystals have primarily the same spectral features except that the intensity of absorption bands in Brazilian quartz is much higher than in the Arkansas quartz. Some of the bands present in the significant strength in the Brazilian quartz are not observable in the Arkansas quartz. Using the H-swept spectra of the two crystals we find that the Al-ratio calculated using the formula $[Al_l + Al_s$ (Brazilian)$/Al_l + Al_s$ (Arkansas)] is nearly 1.2. By comparing the H-swept spectra of the two crystals with the spectra of their respective unswept state we find that in the particular Brazilian sample of Figure 1 or 2 in its unswept state nearly 40% of the substitutional aluminum is charge compensated with hydrogen ions and remaining 60% with alkalis mostly Li. Such a comparison in the Arkansas crystal shows that almost 30% of the substitutional aluminum is charge compensated by alkalis, evidently Li.

Halliburton et.al [5,46] have shown that during irradiation the onset of movement of alkali ions from the aluminum centers in quartz starts only

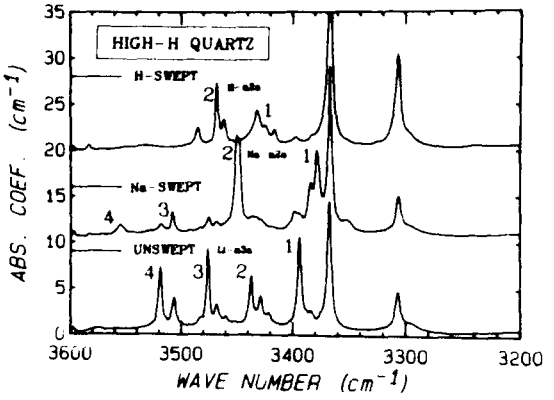


Figure 1: Infrared absorption spectra of a natural quartz crystal depicting all the major bands in its unswept, Na-swept and H-swept conditions.

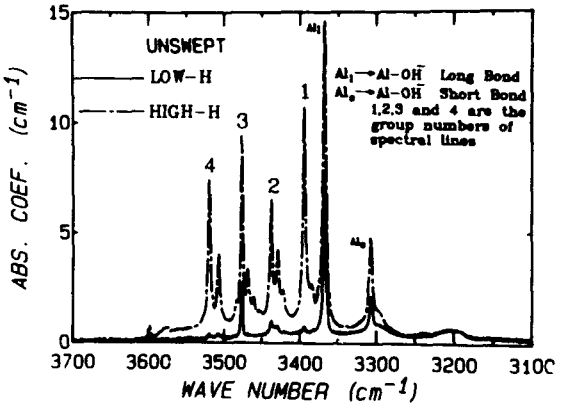


Figure 2: Infrared absorption spectra of low-H and high-H natural quartz crystals in their as-received condition.

when the sample temperature is about or above 200K. Sibley et.al.[25] on the other hand have shown that the protons in quartz crystals are mobile under irradiation at all temperatures down to 10K. We believe that the situation with regard to detrapping of alkalis from defects even other than those involving aluminum may be taken as analogous as far as the irradiationg temperature of the sample is concerned. Thus, from the spectra of the irradiated crystals and the fact that at 77K irradiation only protons move and upon irradiation at room temperature both alkalis and protons are mobile, our estimates show that the Li concentration in the Brazilian sample is approximately 5.2 times higher compared to that in the Arkansas sample. Similarly, Al ratio is 3.25 and H ratio is 3.2. Therefore, it is reasonable to believe that due to relatively much high concentration of both Li and H in the Brazilian sample in its unswept condition the intensities of all Li-H related infrared bands are much higher compared to the intensities of these bands in the Arkansas sample. From the above we see that in the Brazilian sample the H ratio is substantially higher than the Al-ratio in comparison with the Arkansas quartz. We shall therefore designate the Brazilian sample of the present study as high-H and the Arkansas sample as low-H quartz. With this H-limitation, only the most preferred defect sites are occupied, i.e. the minor lines are all very weak and even the dominent lines of the non-preferred groups are quite weak. After Na - sweeping the spectra show dominance of one line, even without H-limitation, presumably caused by the larger ion size preventing the occupency of some of the defect sites. After H-sweeping, degeneracy reduces the number of groups to two. The high-H and low-H spectra in the H-swept condition were found to be somewhat less different than in the unswept condition, presumably the great H-influx during H-sweeping has somewhat relieved the H-limitation.

In the present work irradiation is used as a tool for characterization of infrared active defect centers in quartz. At 77K irradiation, only protons move; they come to rest, most in shallow and some in deep traps. Warm up to room temperature will release most of them back to their original defect species; those not returning must have been captured by deep traps and need higher temperature for release. With shallow and deep defined in this way, irradiation at 77K and subsequent warmup can be used to get an estimate for the concentration of shallow and deep proton traps. Irradiation at room temperature allows movement of both, protons and alkali ions. Spectra taken some time afterward would then yield the combined concentration of deep trapping sites for H and alkali. Subsequent effects of second irradiation at 77K should be compared with the first low-temperature irradiation.

Comparative studies of Irradiation Characteristics of Unswept High-H and Low-H Quartz:

Figures 3 and 4 respectively show the spectra upon irradiation sequence of unswept high-H and low-H quartz. At the first 77K irradiation the general feature was that all bands decreased in their strength significantly except the Al_s at 3306 cm^{-1} which has been found to grow upon low temperature irradiation in both natural [29] and cultured quartz [25]. It was further noticed that at this stage some new low strength IR bands were produced in high-H sample only. These bands were positioned at 3625, 3618, 3455 and 3320 cm^{-1}. Upon warming to room temperature the bands at 3625 and 3618 cm^{-1} stayed but their intensity was much reduced. Other new bands introduced were lost in the background at this stage. The 3595 cm^{-1} which was removed at the first 77K-irradiation recovered to its as-received condition. All other bands showed a recovery between 85-95% of their as-received strength. Since, most of the bands recover significantly upon warming to room temperature following a 77K-irradiation, it is suggested that the protons which are released from the specific defect centers get trapped mostly in shallow traps; those not returning get into deep trapping centers in the quartz lattice. Table (I) summarizes the resultant percentage change in the band strength in both high-H and low-H unswept quartz as a function of irradiation sequence. It may be seen that some bands such as at 3595, 3575, 3520, 3507, 3437, 3430, 3395 and 3383 wavenumbers decreased in identical proportion in both samples at the first 77K-irradiation as well as the final 77K-irradiation. Some bands which were knocked down at room temperature irradiation such as at 3575, 3520, 3507, 3468, 3460 3437 and 3383 cm^{-1} were decreased at the subsequent 77K-irradiation in the same proportion as they were decreased under the first 77K-irradiation from their as-received strength. These bands therefore represent some specific point defect configuration which is typical of natural quartz crystal in its unswept state. Interestingly, the very small band at 3451 cm^{-1} which was lost to background at the first 77K irradiation, started appearing at room temperature irradiation in the low-H sample and appeared in nearly four times of its as-received strength in the high-H sample at the stage of final 77K irradiation. Only in the case of high-H sample two new small bands one at 3535 cm^{-1} and the other at 3320 cm^{-1} were observed at the first 77K irradiation. Another interesting feature can be observed that during the entire sequence of irradiation the Li-n3a band and its shoulder band at 3480 cm^{-1} were found to decrease at each step. However, for the low-H sample, the decrease of Li-n3a band was faster than for high-H sample at all stages of irradiation.

Some interesting features emerged at room-temperature irradiation. A very small new band at 3560 cm^{-1} appeared only in high-H sample. The very small 3535 cm^{-1} bandthat appeared only in high-H quartz after 77K irradiation started appearing in the low-H sample at very low strength at this stage. In the case of high-H sample the 3535 cm^{-1} band at room-temperature-irradiation attained an increased strength of about three times of its 77K-irradiated value. A new band at 3410 cm^{-1} appeared in both samples but was much more prominent in the high-H sample. Brown and Kahan [23] have observed that in natural quartz a line at 3434 cm^{-1} appeared upon

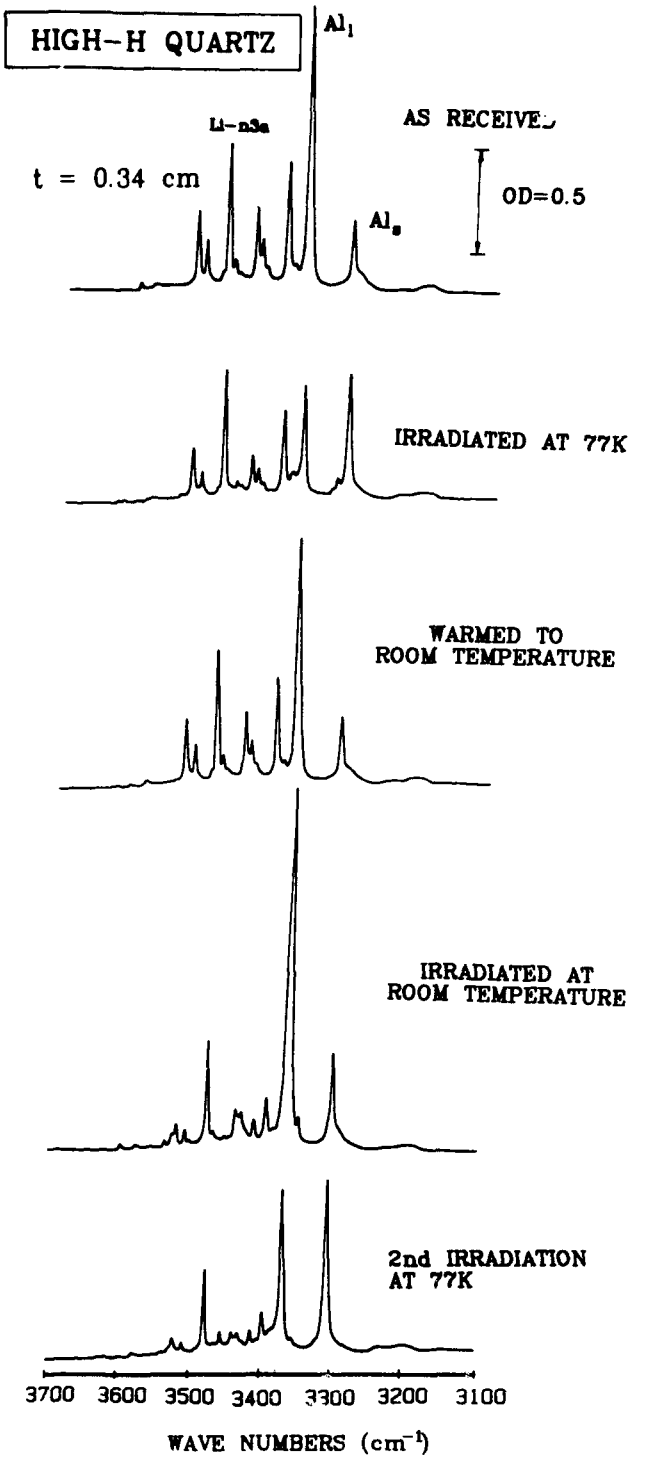


Figure 3: Infrared absorption spectra showing the effect of irradiation sequence on unswept high-H (Brazilian origin) quartz crystal.

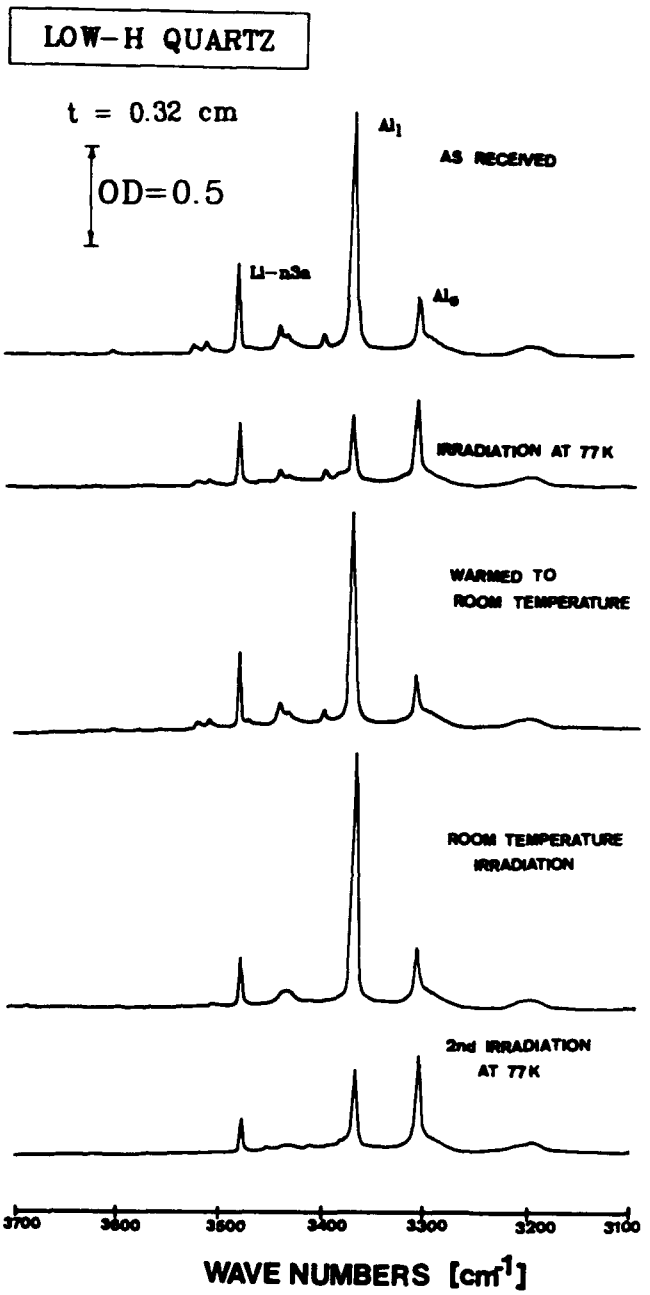


Figure 4: Infrared absorption spectra showing the effect of irradiation sequence on unswept low-H (Arkansas origin) quartz crystal.

room temperature irradiation and the intensity of lines at 3367 and 3306 wavenumbers was increased. These authors [23] found that the 3434 cm^{-1} band appeared at relatively high radiation dose =100 M rad. We have found that this line appears in low-H sample at a relatively much less radiation dose of 2 M rad at room temperature irradiation. On the other hand, in the high-H sample, the bands in this region decreased at all steps of irradiation. In the present work we have also found that the room temperature irradiation increases the intensity of $Al\text{-}OH^-$ (3306 and 3367 cm^{-1}) lines. The radiation effects on $Al\text{-}OH^-$ in differently swept natural quartz is described in a separate section. Another interesting feature in the case of high-H sample only was that the 3320 cm^{-1} band produced at 77K-irradiation was removed and another band at 3355 cm^{-1} came up. At the stage of final irradiation at 77K the 3595 cm^{-1} band was removed in both the samples and the 3560 cm^{-1} band introduced at room temperature irradiation in high-H sample was also removed. The band at 3240 cm^{-1} which started appearing at room temperature irradiation became more prominent and readily seen in both the samples.

Comparative Studies of Irradiation Characteristics of Na-swept High-H and Low-H Quartz:

Figure 5 and 6 respectively show the spectra of as Na-swept and after each step of irradiation for both high-H and low-H quartz. It may be seen from the spectra that while some bands on the high frequency side of Al_1 in Na-swept condition are distinctly resolved in high-H samples, they may be minutely seen in the low-H sample. After the first 77K irradiation the general feature was that all the bands except those at 3476 and 3468 cm^{-1} both present in small strength and the Al_s at 3306 cm^{-1} showed a reduced strength. On the other hand the 3476 cm^{-1}, 3468 cm^{-1} and the Al_s bands showed an increased strength. The increase in the Al_s band at low temperature irradiation is the usual behavior [25,29]. However, the behavior of 3476 cm^{-1} and 3468 cm^{-1} bands in the Na-swept condition under the first low temperature irradiation is opposite of the unswept condition. While in the unswept condition these bands showed a reduction in their intensity at all stages of irradiation in the sequence; in the Na-swept condition upon the first irradiation at 77K these bands showed asignificant increase in their intensity. Both the bands increased more in the low-H sample than in the high-H one. Table (II) summarizes the resultant percentage change in the band strength in both high-H and low-H Na-swept quartz as a function of irradiation sequence. An interesting point may be noted here. The 3476 cm^{-1} band in both the Na-swept samples behaved almost identically uptil the stage of room temperature irradiation. At the subsequent final irradiation at 77K the low-H sample showed an increase and the high-H quartz exhibited a decrease in the intensity of this band.

The 3468 cm^{-1} band in high-H sample showed about 150% increase in the intensity at the first 77K irradiation while in the low-H sample this band remained almost unaffected. Upon warming to room temperature this band appeared in about triple of its as-Na-swept strength in the low-H sample while it remained at its 77K-irradiated value. At 300K irradiation in both the samples this band attained a value of about 200% of the as-swept strength. Interestsingly, at the subsequent final irradiation at 77K this band behaved oppositely in low and high-H samples. In the low-H sample it was lost to background while in the high-H one it continued to increase and attained a value of about 250% of its as Na-swet strength.

It may be noted from Figures 1 and 3-6 that in the as-Na-swept condition the band at 3476 cm^{-1} is more intense than the 3468 cm^{-1} band; a situation opposite of the unswept state of natural quartz. These bands increase under the first 77K irradiation. Upon warming to room temperature these bands showed a reversal in their relative strength; the 3468 cm^{-1} band increased much while the 3476 cm^{-1} band decreased considerably. Also , the 3437 cm^{-1} band in the Na-swept condition behaved opposite to the case of unswept crystal. At the low temperature irradiation this band was knocked down in identical proportion in both samples and recovered to as-Na-swept strength in both samples. Irradiation at 300K increased its strength by about 25% in both samples and the subsequent final low temperature irradiation reduced this band (65% of as-Na-swept value) in high-H sample. On the other hand in the unswept condition this band showed a decrease at all stages of irradiation. However, the similarity was that at the stage of final low temperature irradiation this band was knocked down more in low-H quartz.

The Na-n2a (3451 cm^{-1}) band in both the samples exhibited a similar character of diminishing in strength at all stages of irradiation as also shown by the Li-n3a band in unswept condition. A comparison of Li-n3a band (in unswept condition) with Na-n2a band (in Na-swept condition) shows that during the irradiation carried out the Na-n2a band decreased much faster than the Li-n3a band at all stages of irradiation in both the samples. This indicates that both Li and H are relatively strongly bonded to the defect center representing the Li-n3a band as compared to the Na and H at the Na-n2a band.

An interesting difference was observed with regard to production of new small bands at 3625 and 3618 wavenumbers in the unswept and Na-swept conditions. In the unswept case these bands appeared only in high-H sample upon first irradiation at 77K. After Na-sweeping, the 3625 cm^{-1} band was produced only in low-H sample. However, the 3618 cm^{-1} small band in conjunction with two others at 3570 and 3545 cm^{-1} appeared in both the samples. The 3520 cm^{-1} band which was not initially present in the

TABLE - I

Summary of resultant percentage change in the infrared band strength of unswept natural quartz crystals of high-H and low-H concentration as a function of irradiation at low-, room- and again low-temperature irradiation (radiation dose at every step = 2 Mrad).

Treatment given	Band positions (cm^{-1}) AS-RECEIVED	Resultant change in relation to the as-received band strength of: High-H Quartz	Low-H Quartz	Comments
First irradiation at 77K	3595	Removed	Removed	(i) The band at 3575 cm^{-1} was very small in high-H sample and was not observable in the low-H sample in its as-received condition.
	3575	Unaffected	———	
	3520	Reduced to 63%	Reduced to 60%	(ii) Two new small bands at 3625 and 3618 wavenumbers appeared along with a shoulder at 3322 wavenumber in high-H sample only. No such band appeared in the low-H sample.
	3507	Reduced to 60%	Reduced to 60%	
	3480	Reduced to 75%	———	(iii) The band at 3480 cm^{-1} was not discernible in low-H as-received quartz.
	3476 (Li-n3a)	Reduced to 86%	Reduced to 70%	
	3468	Lost to Background	———	(iv) There was a tendency for appearance of a very small band at 3535 wavenumber in high-H sample only.
	3460	Lost to Background	———	
	3451	Lost to Background	Lost to Background	(v) The 3468 and 3460 cm^{-1} bands were not discernible in its as-received condition in low-H sample.
	3437	Reduced to 50%	Reduced to 50%	
	3430	Reduced to 50%	Reduced to 50%	(vi) A new and small band at 3455 cm^{-1} appeared in high-H sample with a tendency to appear in the low-H sample also.
	3420	Reduced to 85%	———	
	3395	Reduced to 75%	Reduced to 75%	(vii) The 3420 cm^{-1} band was hardly discernible in low-H sample .
	3383	Almost unaffected	Almost unaffected	
	3367 Al_1 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE - V.		(viii) A new small band appeared at 3320 wavenumbers in high-H quartz only.

Warmed to room temperature	Comments:
	(i) Small bands at 3625 and 3618 cm^{-1} introduced in high-H sample at 77K irradiation stayed upon warming to room temperature. However, their intensities were much reduced.
	(ii) The 3595 cm^{-1} band recovered to its half intensity in both samples.
	(iii) All other bands showed a recovery between 85-95% of their as-received strength.

Treatment given	Band positions (cm^{-1})	Resultant change in relation to the as-received band strength of: High-H Quartz	Low-H Quartz	Comments
Irradiation at room temperature	3595	Reappeared to 75%	Reappeared to 75%	(i) A very small new band at 3560 cm^{-1} was seen only in high-H sample.
	3575	Reduced to 50%	———	
	3520	Reduced to 28%	Removed	(ii) The very small 3535 cm^{-1} band that appeared in high-H quartz after 77K-irradiation started appearing in low-H sample after room temperature irradiation. However, the intensity was very small. In the case of high-H sample the band attained an increased strength of about 3 times of its 77K-irradiated value.
	3507	Reduced to 30%	Reduced to 30%	
	3480	Lost to Background	Lost to Background	
	3476 (Li-n3a)	Reduced to 73%	Reduced to 52%	
	3468	Reduced to 40%	———	(iii) A new band at 3410 cm^{-1} appeared in both samples; prominent in high-H.
	3460	Removed	———	
	3451	Lost to Background	———	
	3437	Reduced to 35%	Reduced to 50%	(iv) The 3320 cm^{-1} band produced at 77K irradiation was removed and another band at 3355 wavenumber started appearing in high-H sample only.
	3430	Reduced to 65%	Lost to background	
	3420	Removed	Removed	(v) Bands at 3625 and 3618 cm^{-1} introduced at 77K-irradiation and stayed upon warming to room temperature vanish at this stage.
	3395	Reduced to 35%	Lost to Background	
	3383	Lost to Background	Lost to Background	
	3367 Al_1 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE -V.		(vi) A small band 3240 cm^{-1} started appearing in both the samples.

Treatment given	Band positions (cm^{-1})	Resultant change in relation to the as-received band strength of		Comments
		High-H Quartz	Low-H Quartz	
Final irradiation at 77K	3595	Removed	Removed	(i) The 3595 cm^{-1} band got removed in both the samples.
	3575	Remained to 50%	———	
	3520	Reduced to 17%	Lost to Background	(ii) The 3560 cm^{-1} band introduced at room temperature irradiation in high-H sample got removed.
	3507	Reduced to 22%	Lost to Background	
	3480	———	———	(iii) In the low-H sample a small envelope was observed consisting of small bands in the 3437-3420 wavenumber region. On the other hand, in the high-H sample the band at 3520 wavenumber started appearing.
	3476 (Li-n3a)	Reduced to 60%	Reduced to 65%	
	3468	Lost to Background	Lost to Background	
	3460	Lost to Background	Lost to Background	
	3451	Increased to 4 times	Started appearing	
	3437	Reduced to 18%	Reduced to 15%	(iv) The band at 3240 cm^{-1} which started appearing at room temperature irradiation became more prominent and readily observable in both the samples.
	3430	Reduced to 18%	Reduced to 15%	
	3420	Started appearing	See comments	
	3395	Reduced to 24%	———	
	3383	Lost to Background	Lost to Background	
	3367 Al_1 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE -V.		

TABLE - II

Summary of resultant percentage in the infrared band strength of Na-swept natural quartz crystals of high-H and low-H concentration as a function of irradiation at low-, room- and again low-temperature irradiation (radiation dose at every step = 2 Mrad).

Treatment given	Band positions (cm^{-1}) AS-Na-SWEPT	Resultant change in relation to the as-Na-swept band strength of		Comments
		High-H Quartz	Low-H Quartz	
First irradiation at 77K	3595	Reduced to 40%	Lost to Background	(i) Three new small bands at 3618,3570 and 3545 wavenumbers are produced in both samples. A band at 3625 wavenumber (in small strength) was produced only in low-H sample. This is contrasted with unswept quartz where the 3625 and 3618 wavenumber bands were produced only in high-H sample upon 77K-irradiation.
	3585	Almost Unaffected	Reduced to 50%	
	3555	Reduced to 50%	Reduced to 50%	
	3520	Almost Unaffected	See Comments	
	3507	Reduced to 70%	Reduced to 70%	
	3476	Increased to 200%	Increased to 230%	(ii) The 3520 cm^{-1} band was not present in the as-Na-swept Low-H quartz but appeared in low strength after 77K-irradiation.
	3468	Increased to 150%	Unaffected	
	3451 (Na-n2a)	Reduced to 66%	Reduced to 60%	(iii) There was an appearance of a shoulder band at 3470 wavenumber on the lower frequency side of 3476 wavenumber band in both samples.
	3437	Reduced to 60%	Reduced to 60%	
	3395	Reduced to 70%	Reduced to 50%	
	3385	Reduced to 65%	Reduced to 45%	(iv) In the as-Na-swept condition the band at 3476 wavenumber was more intense than the 3468 wavenumber band. This situation is opposite of the unswept state of natural quartz. These bands increased upon first 77K-irradiation.
	3375	Reduced to 80%	Reduced to 70%	
	3350	Slightly Increased	Slightly Increased	(v) A band at 3240 cm^{-1} started appearing in low-H sample while it was there in the as-Na-swept high-H sample.
	3367 Al_1 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE -V.		

Warmed to room temperature		Comments: (i) The bands at 3625 and 3618 cm^{-1} became more clear in low-H sample. In high-H sample both of these bands were also clearly observable. (ii) The bands at 3555 and 3507 cm^{-1} were recovered to their as-Na-swept strength. (iii) The band at 3476 and 3468 cm^{-1} showed a reversal in their strength relative to their as-Na-swept values. (iv) The 3437 cm^{-1} band grew to double i.e. to its as-Na-swept strength in both the samples. (v) All other bands showed a recovery between 80-90% of their as-Na-swept strength.		
Treatment given	Band positions (cm^{-1})	Resultant change in relation to the as-Na-swept band strength of		Comments
		High-H Quartz	Low-H Quartz	
Irradiation at room temperature	3595	Reduced to 55%	Reduced to 55%	(i) The band at 3595 cm^{-1} showed an increase from its 77K-irradiated value. (ii) A doublet of two new bands at 3545 and 3532 wavenumbers appeared at this stage in both the samples but more prominently in the high-H quartz. (iii) A new small band at 3472 cm^{-1} appeared in both samples. (iv) Although the bands at 3476 and 3468 wavenumbers are small in strength but but they seemed to be growing in almost in same proportion in both samples. (v) A small and sharp new band at 3330 cm^{-1} appeared in both samples. (vi) The 3240 cm^{-1} band in high-H sample remained unaffected while it got lost in the background in the low-H sample.
	3585	Removed	Removed	
	3555	Reduced to 20%	Lost to Background	
	3520	Lost to Background	Lost to Background	
	3507	Reduced to 43%	Reduced to 25%	
	3476	Increased to 150%	Increased to 150%	
	3468	Increased to 200%	Increased to 200%	
	3451 (Na-n2a)	Reduced to 50%	Reduced to 48%	
	3437	Increased to 125%	Increased to 125%	
	3395	Reduced to 66%	Reduced to 50%	
	3385	Lost to Background	Lost to Background	
	3375	Reduced to 55%	Lost to Background	
	3350	Removed	Removed	
	3367 Al_l 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE-V.		
Final irradiation at 77K	3595	Reduced to 40%	Lost to Background	(i) The band at 3545 cm^{-1} produced after room temperature-irradiation was reduced to about half of its value in both samples and the 3532 wavenumber band was removed. (ii) The band at 3476 cm^{-1} in high-H sample was reduced while it was increased in the low-H quartz. (iii) The 3468 cm^{-1} band in high-H sample was increased to double while in the low-H quartz it was lost to background beyond recognition. (iv) The 3472 cm^{-1} band produced at room temperature irradiation reduced to half in the high-H sample while it was increased to nearly double in low-H quartz. (v) The 3330 cm^{-1} band produced at room temperature irradiation remained almost unaffected in both the samples. (vi) The 3350 cm^{-1} band removed on room temperature irradiation did not recover at final 77K-irradiation. (vii) The 3240 cm^{-1} band increased in high-H sample increased from its value attained after room temperature irradiation and appeared distinctly in the low-H quartz.
	3585	Lost to background	Removed	
	3555	Removed	Removed	
	3520	Lost to Background	Lost to Background	
	3507	Reduced to 25%	Lost to Background	
	3476	Reduced to 75%	Increased to 175%	
	3468	Increased to 250%	Lost to Background	
	3451 (Na-n2a)	Reduced to 35%	Reduced to 25%	
	3437	Reduced to 66%	Reduced to 36%	
	3395	Reduced to 45%	Reduced to 45%	
	3385	Lost to Background	Lost to Background	
	3375	Reduced to 55%	Lost to Background	
	3350	Removed	Removed	
	3367 Al_l 3306 Al_s	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE-V.		

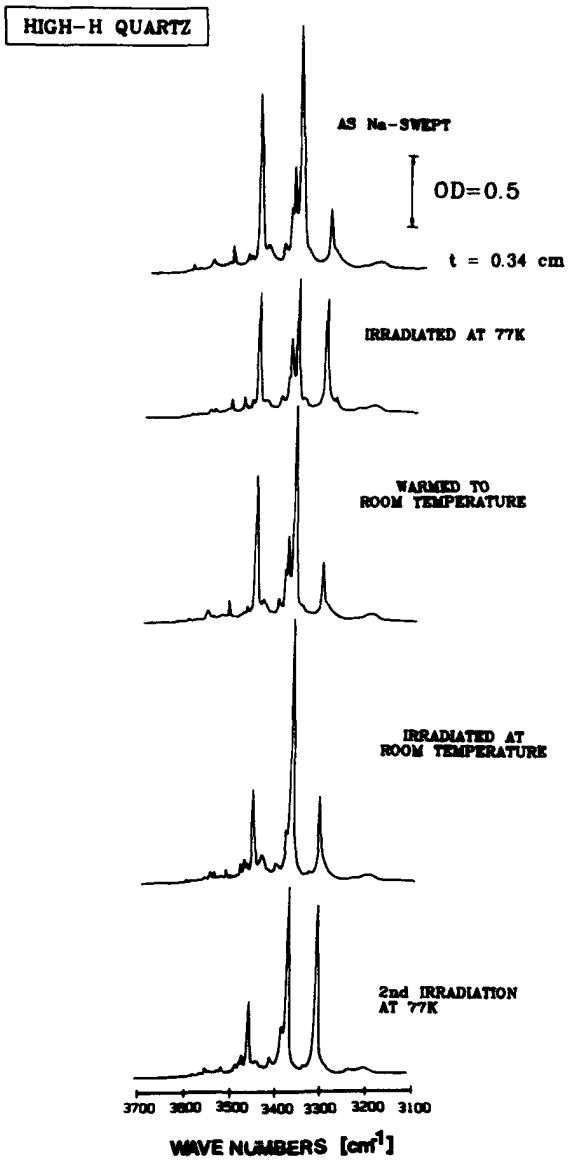


Figure 5: Infrared absorption spectra showing the effect of irradiation sequence on the Na-swept high-H natural quartz.

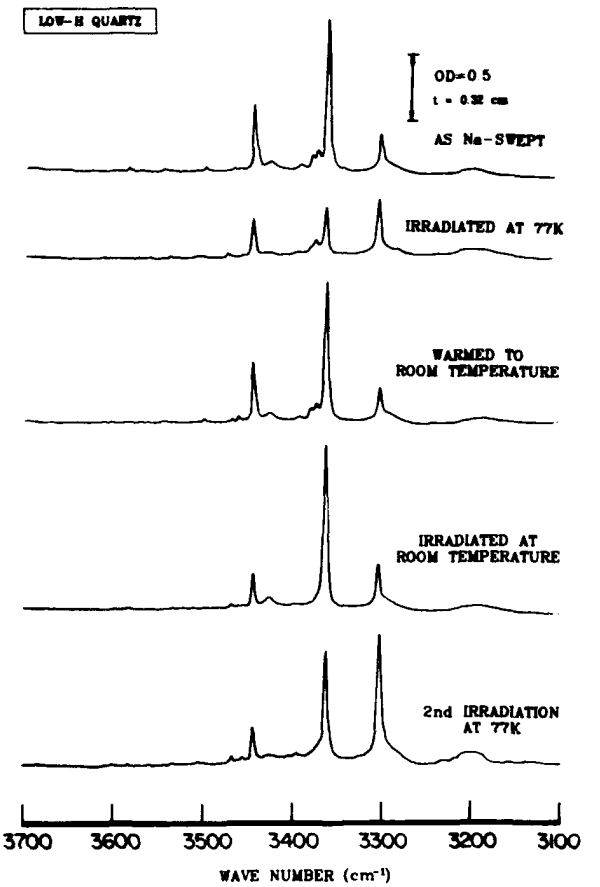


Figure 6: Infrared absorption spectra showing the effect of irradiation sequence on the Na-swept low-H natural quartz.

as-Na-swept low-H sample appeared in low strength upon 77K irradiation. It may be noted from the Figures 5 and 6 and Tables (I) and (II) that during the course of irradiation sequence carried out, sometimes appearance of new low strength infrared OH^- bands may be observed in unswept and Na-swept quartz. This generally occurred at the expense of already existing defect bands.

Comparative Studies of Irradiation Characteristics of H-swept High-H and Low-H quartz:

Figures 7 and 8 respectively show the spectra upon irradiation sequence of hydrogen swept high-H and low-H quartz. At the first 77K irradiation all the bands except the one at 3535 cm^{-1} and the other the Al_s at 3306 cm^{-1}, showed a significant reduct-

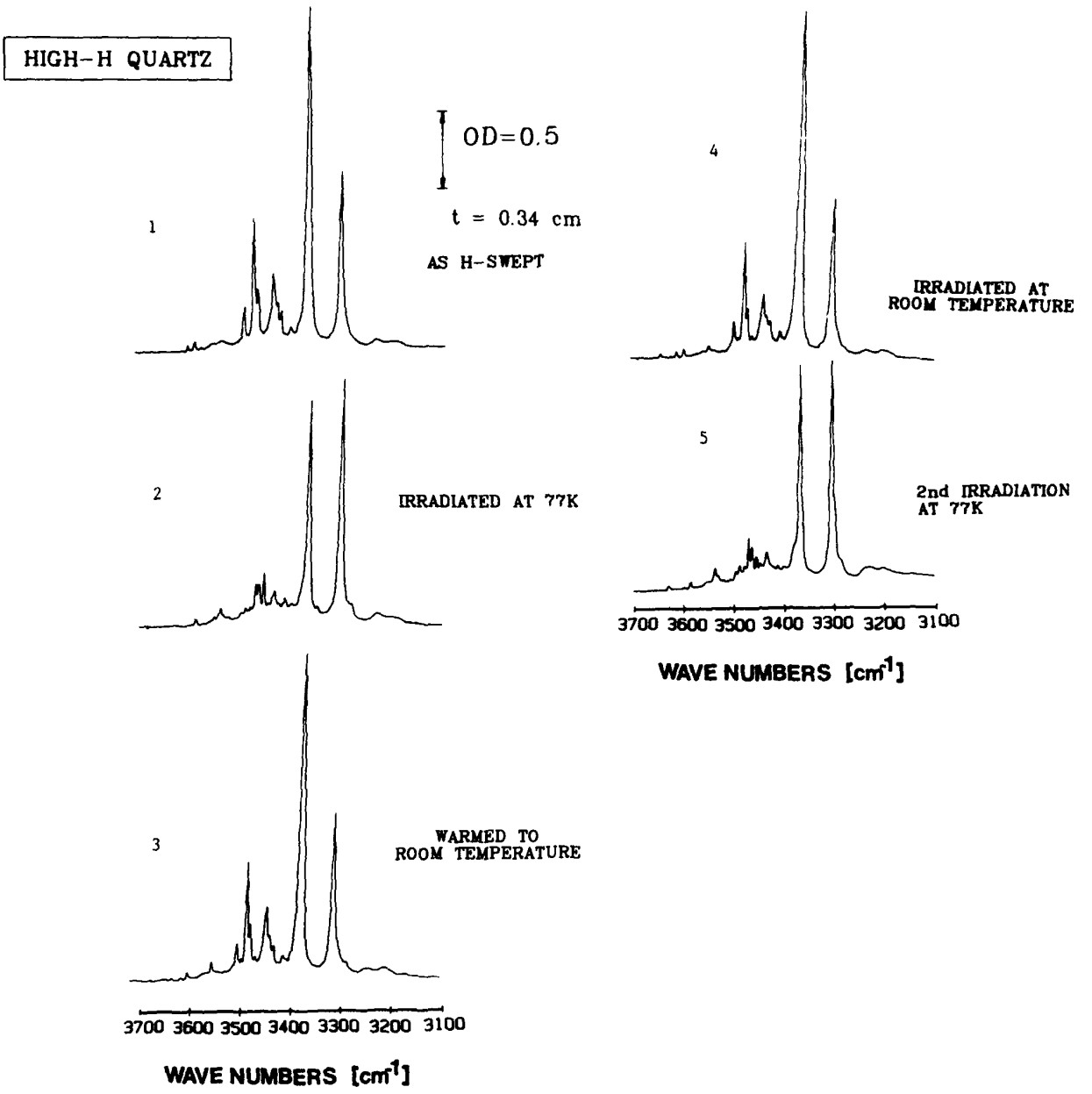


Figure 7: Infrared absorption spectra showing the effect of irradiation sequence on the H-swept high-H natural quartz.

TABLE - III

Summary of resultant percentage change in the infrared band strength of H-swept natural quartz crystals of high-H and low-H concentration as a function of irradiation at low-, room- and again low-temperature irradiation (radiation dose at every step = 2 Mrad).

Treatment given	Band positions (cm^{-1})	Resultant change in relation to the as-H-swept band strength of		Comments
		High-H Quartz	Low-H Quartz	
First irradiation at 77K	3595	Removed	Removed	(i) Four new very small bands at 3548, 3520, 3495, and 3476 wavenumbers were observed in both high-H and low-H samples.
	3583	Unaffected	Unaffected	
	3570	Lost to Background	Lost to Background	
	3535	Increased to 235%	Increased to 400%	(ii) Originally, the 3535 cm^{-1} band was a broad one (stronger in high-H sample by about four times of that in low-H sample) in both samples. After 77K-irradiation a sharp peak appeared in both samples at 3535 wavenumber superposed on the broad background in bothe samples.
	3483	Reduced to 20%	Reduced to 30%	
	3468 (H-n2a)	Reduced to 28%	Reduced to 28%	
	3460	Reduced to 75%	Increased to 130%	
	3430	Reduced to 30%	Reduced to 50%	
	3420	Lost to Background	Lost to Background	(iii) In both samples a new band at 3452 wavenumber appeared in significant strength. The band shape was different from the Na-3451 wavenumber band.
	3415	Lost to Background	Lost to Background	
	3395	Reduced to 50%	Reduced to 50%	
	3240	Unaffected	Slightly increased	(iv) A new small band at 3410 cm^{-1} was observed in both samples.
	3367 Al_1	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE-V.		(v) A small band at 3350 cm^{-1} was present in the as-H-swept low-H sample which was lost to background beyond recognition after 77K-irradiation.
	3306 Al_s			
Warmed to room temperature		Comments: (i) The very small bands at 3548 and 3520 cm^{-1} produced at 77K irradiation vanished. (ii) The band at 3476 cm^{-1} produced after 77K-irradiation was reduced to about 80% in high-H sample and was lost to background in the low-H natural quartz. (iii) H-n2a band in the low-H sample was slightly (107%) increased from its as-H-swept value; in the high-H sample it recovered to about 95% of as-H-swept value. (iv) The 3452 cm^{-1} band produced at 77K-irradiation was reduced to about 15% of its value in both the samples. (v) The 3240 cm^{-1} band in the low-H sample got increased slightly from its 77K-irradiated value. (vi) All other bands were recovered to about 80-95% and the spectral features were restored to the as-H-swept condition.		
Irradiation at room temperature	3595	Recovered to as-H-swept value	Recovered to as-H-swept value	(i) A small band at 3430 cm^{-1} was produced in both samples; more prominent in high-H sample.
	3583	Unaffected	Unaffected	
	3570	Unaffected	Unaffected	
	3535	Reduced to 90%	Unaffected	(ii) The 3476 cm^{-1} band produced after 77K-irradiation was reduced to about 75% of its value.
	3483	Reduced to 80%	Reduced to 80%	
	3468 (H-n2a)	Slightly increased to 107%	Slightly increased to 107%	
	3460	Reduced to 80%	Reduced to 65%	(iii) The band at 3452 cm^{-1} produced after 77K-irradiation was reduced to 15% of its value i.e. as was upon warming to room temperature following the irradiation at 77K.
	3430	Reduced to 90%	Unaffected	
	3420	Not Observable	Not Observable	
	3415	Reduced to 80%	Not Observable	
	3395	Increased to 150%	Unaffected	
	3240	Unaffected	Reduced to 50%	
	3367 Al_1	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE-V.		
	3306 Al_s			

Treatment given	Band positions (cm^{-1})	Resultant change in relation to the as-H-swept band strength of		Comments
		High-H Quartz	Low-H Quartz	
Final irradiation at 77K	3595	Removed	Removed	(i) The small band at 3630 cm^{-1} produced on room-temperature-irradiation stayed and was reduced to 75% of its value.
	3583	Almost Unaffected	Almost Unaffected	
	3570	Recovered to as-H-swept value	Recovered to as-H-swept value	
	3535	Increased to 290%	Increased to 265%	(ii) The very small band at 3548 cm^{-1} which was observed at the first 77K-irradiation was observed at this stage also in about same strength.
	3483	Reduced to 20%	Reduced to 30%	
	3468 (H-n2a)	Reduced to 40%	Reduced to 20%	(iii) The 3535 cm^{-1} band in high-H sample increased while in low-H it decreased as compared with the first 77K-irradiation
	3460	Reduced to 55%	Reduced to 66%	
	3430	Reduced to 30%	Reduced to 50%	(iv) A small shoulder band at 3530 wavenumber on low frequency side of 3535 wavenumber band appeared in both samples.
	3420	Lost to Background	Lost to Background	
	3415	Reduced to 12%	Started Appearing	(v) The 3520 cm^{-1} band appeared in both samples as on first 77K-irradiation. Additionally, in high-H sample very small bands at 3510 and 3505 wavenumbers appeared.
	3395	Reduced to 50%	Reduced to 50%	
	3240	Unaffected	Removed	(vi) A new very small band at 3495 wavenumber appeared in both samples.
	3367 Al_1 3306 Al_2	EFFECT ON Al-OH^- BANDS IS DESCRIBED SEPARATELY IN TABLE-V.		(vii) The 3476 cm^{-1} produced at first 77K-irradiation remained to its room temperature-irradiated value.
				(viii) In high-H sample, the 3452 cm^{-1} band increased from the value it attained after room temperature-irradiation but the final value was one-third of its value obtained at first 77K-irradiation. On the other hand, in low-H sample this band was only reduced to about 80% of its first 77K-irradiated value.
				(ix) In low-H sample the 3415 cm^{-1} band started appearing at this stage while it was not observable at previous stages.

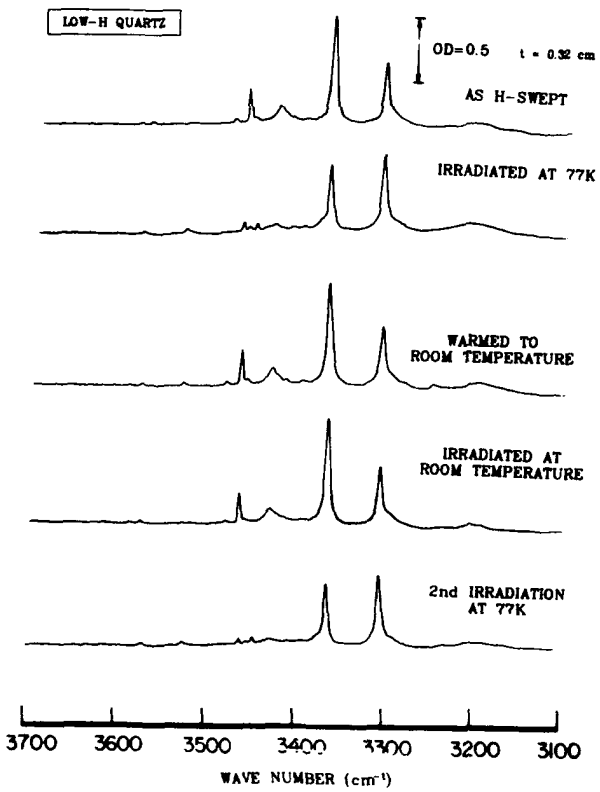


Figure 8: Infrared absorption spectra showing the effect of irradiation sequence on the H-swept low-H natural quartz.

ion. The increase of the 3535 cm^{-1} band was much high (400%) in low-H sample in relation to that in the high-H quartz (235%). Originally, the 3535 cm^{-1} band was a broad one in both the samples. It was nearly four times stronger in high-H sample than in the low-H one. After first 77K irradiation a sharp peak appeared at 3535 cm^{-1} superposed on the broad background in both the samples. At room temperature irradiation this band was slightly reduced in high-H sample while it remained almost unaffected in low-H quartz. Subsequent final 77K-irradiation resulted in increase of this band in high-H sample by about 125% of its value obtained upon first 77K irradiation. At this stage this band in low-H sample was reduced to about 66% of the value obtained upon first 77K irradiation.

The band at 3240 cm^{-1} remained almost unaffected in the high-H sample while in the case of low-H quartz it showed a slight increase. Another difference in the two samples was that while the 3460 cm^{-1} band in the high-H quartz was reduced (to about 75%) at the first 77K irradiation, that in the low-H quartz got increased to about 130%. All other bands in the two kinds of hydrogen-swept samples of natural quartz behaved qualitatively almost similarly. Table (III) summarizes the resultant percentage change in the band strength in both high-H and low-H hydrogen swept quartz. It was noted that in both the samples a new band at 3452 cm^{-1} appeared in significant strength at the first low temperture irradiation. The band shape was different from the 3451 cm^{-1} in unswept or Na-n2a in Na-swet samples. Upon warming to room temperature this band was reduced to about 15% of its 77K-irradiated value in both samples. Room temperature irradiation did not affect this band. Subsequent final irradiation at 77K gave interesting results. In the high-H sample it showed an increase from its value after room temperature irradiation but the final magnitude attained was one-third of its value obtained at the first 77K irradiation.On the other hand in the low-H sample this band was only reduced to about 80% of its 77K-irradiation value. However, extended irradiation (dose = 6 Mrad) at this stage resulted in increase of this band to about double of its strength.

Some new and very small bands were observed at the first 77K irradiation. These were positioned at 3548, 3520, 3495, 3476 waveenumbers. Warming to room temperature resulted either in disappearance or reduction in strength of these bands. Room temperature irradiation also resulted in some bands reduction of their strength (Table III).

The behavior of H-n2a band under the influence of radiation is very different from the major bands in alkali swept versions of natural quartz. Table (IV) shows a comparison of irradiation characteristics of Li-n3a , Na-n2a and H-n2a bands in unswept, Na-swept and H-swept natural quartz crystals of high- and low-H concentrations. In the present case only the low temperature irradiation decreased the intensity of this band. Irradiation at room temperature did not decrease the band intensity in the low-H sample while in the high-H quartz it was increased a little from the value attained upon warming to room temperature following the first irradiation at 77K. On the other hand in the alkali swept crystals both Li-n3a and Na-n2a bands showed a decrease in their intensities at all stages of irradiation. Since in the H-swept crystals there are no alkalis the room temperature irradiation will dissociate only protons from different OH^- related or point defects. The fact that upon room temperature irradiation the H-n2a band shows a slight increase indicates that this defect center preferes the accumulation of protons. This fact is also clearly shown in Figure 9. In the case of low-H quartz the H-n2a band is stronger than either Li-n3a or Na-n2a. In the high-H sample the sweeping picture is not so clear as the crystal has a large number of proton trapping centers in its as-received state. However, in this case the increased strength of this band at room temperature irradiation suggests a similar situation occuring in high-H sample also. It is known that for completely H-swept crystals all the aluminum ions are charge compensated by protons. The irradiation process then carried out does not form stable Al-hole centers [5,7]. The formation of stable Al-hole centers in such cases is prevented because the protons released from $Al\text{-}OH^-$ centers do not find trapping sites and therefore quickly return to the aluminum sites. It seems that an almost similar situation seems to be taking place with regard to release of protons from the center representing the H-n2a band at room temperature irradiation.

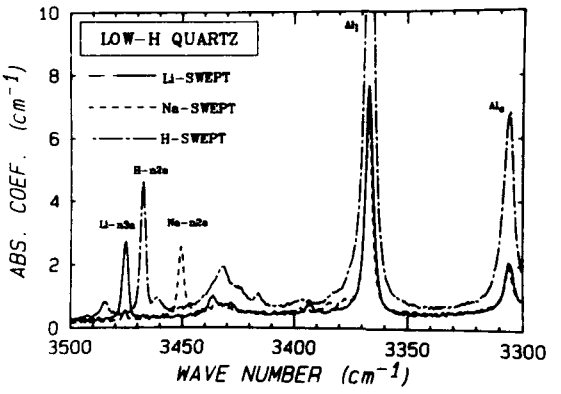


Figure 9: Infrared absorption spectra depicting all the major bands in low-H natural quartz in Li-swept, Na-swept and H-swept conditions. The H-n2a bandis stronger than either Na-n2a or Li-n3a.

TABLE - IV

Comparison of irradiation characteristics of Li-n3a, Na-n2a and H-n2a bands in unswept, Na-swept, and H-swept natural quartz crystals of high-H and low-H concentration.

Band position (cm^{-1})	Treatment given	Percentage change in the band strength					
		Unswept		Na-swept		H-swept	
		High-H	Low-H	High-H	Low-H	High-H	Low-H
3476	First LT-irradiation	-14	-30	+140	+230	Li-n3a band is not observed in H-swept quartz	
	Room Temperature irradiation	-27	-48	+150	+150		
	Second LT-irradiation	-40	-35	-75	+175		
3451	First LT-irradiation	Lost to Background	Lost to Background	-34	-40	Na-n2a band is not observed in H-swept quartz	
	Room Temperature irradiation	Lost to Background	Lost to Background	-50	-48		
	Second LT-irradiation	+400	Starts appearing	-35	-25		
3468	First LT-irradiation	Lost to Background	Lost to Background	+150	Unaff-ected	-72	-72
	Room Temperature irradiation	-27	-48	+200	+200	+107	+107
	Second LT-irradiation	Lost to Background	Lost to Background	+300	Lost to Background	-60	-80

Radiation Effects on Al-OH^- Centers in Unswept, Na-swept and H-swept High-H and Low-H Natural Quartz:

Table (V) presents the results on radiation effects on Al-OH^- centers in unswept, Na-swept and H-swept natural quartz of high-H and low-H concentration. The general feature was that on low temperature irradiation the Al_1 band decreased and Al_s increased in strength such that their intensities became almost equal. This is in accordance with the results on cultured crystals [5,25]. Since at low temperature irradiation only protons move and not alkalis, it would mean that protons from Al_1 released get into Al_s which eventually get more populated. This process continues till the equalization of proton population in the two Al-OH^- bonds is reached. Upon warming the low temperature irradiated quartz to room temperature both the Al-OH^- bands recover to nearly 85-95% of their original strength. Room temperature irradiation increased the combined strength of the Al-OH^- bands. The behavior of these bands at the final low temperature irradiation is again of similar character except a little quantitative difference which would be there due to the effect of room temperature irradiation in between the two low temperature irradiations. This shows that Al_s acts like a shallow traps for the protons but is actually a deep trap and shallower than Al_1 with a barrier above 77K.

Important features with regard to radiation-induced changes in case of Al-OH^- bands may be noted here. First, it may be observed that for the unswept samples the percentage change in Al_1 and Al_s at the first and the final 77K-irradiation steps is the same between the as-received ⟶ first 77K irradiation and room-temperature irradiated ⟶ final 77K-irradiation. Additionally, these bands at 77K irradiation decreased relatively less in comparison with low-H quartz. On the other hand, at room temperature irradiation the Al_1 in high-H sample increased more that what in low-H sample.

Some prominent observations on radiation effects on Al-OH^- centers of differently swept low- and high-H natural quartz are the following. It may be seen from the Table (V) and Figures 4 and 6 that in the low-H sample the percentage change of intensity of both the Al-OH^- bands was the same on Na-sweeping as it was in the unswept condition. After room temperature irradiation the Al_s band intensity increased by about 13.5%. There was practically no difference in the percentage change of Al_1 band intensity at this stage between the unswept and Na-swept quartz. Final low temperature irradiation led to much more reduction in the Al_1 band strength in Na-swept condition compared to unswept quartz. On the other hand, at this stage Al_s band in Na-swept quartz increased proportionally to about double in comparison with the unswept condition.

In the case of high-H quartz the percentage change in the strength of these bands at final low temperature irradiation was higher in Na-swept condition in comparison with unswept quartz. Room temperature irradiation did not increase the Al_1 band strength much (only by about 7%) in Na-swept condition while the Al_s band intensity increased significantly much in comparison with the unswept condition. Also, the final low temperature irradiation in Na-swept condition resulted in proportionally higher changes of these bands in comparison with the unswept quartz.

In the H-swept condition the Al_1 band decreased much faster at low temperature irradiation in comparison with unswept or Na-swept states while there was a relatively less increase in the strength of Al_s band. There was, as expected, no change in the band strength of Al-OH^- centers at room temperation irradiation. It may be mentioned here that Brown and Kahan [23] have found that an intense room temperature irradiation (dose of 6700 Mrad) reduced the strength of Al-OH^- in H-swept natural quartz. In the present study it was noticed that for the H-swept natural quartz the final low temperature irradiation reduced the extent of decrease of Al_1 and increase of Al_s in both the samples as compared to that at the first low temperature irradiation. This would be due to changes in the point defect configuration brought about due to intermediate room temperature irradiation.

Comparison with Cultured Quartz;

At the outset it may be mentioned that the Al-OH^- centers in natural quartz show similar irradiation characteristics as those observed in cultured quartz [5,25]. While these bands are present in as-received unswept natural quartz [29], they are produced in cultured crystals only after irradiation at temperatures beyond 200K or upon sweeping in hydrogen atmosphere [5,25,26,29,46]. The process of sweeping or irradiation in natural crystals increases the Al-OH^- band strength in the limit to saturate all the aluminum with hydrogen [29]. This is surmized in view of the findings of Halliburton et. al [5] that in cultured H-swept crystals there were no Al-hole centers observed in the as-swept condition; only Al-OH^- bands were observed which did not show any increase upon room temperature irradiation indicating that all the aluminum was charge compensated with protons. Low temperature irradiation of H-swept natural quartz in the present study has shown that like what happens in cultured quartz equalizes the strength of the Al-OH^- bands indicating that both short and long bonds in the AlO_4 tetrahedra become equally populated. Our isochronal anneal date [29] following a strong low temperature irradiation on these crystals also showed that like in the case of cultured crystals [25] the Al-OH^- bands grow in their relative strength until 230K. Above 230K, they quickly recover to their original strength.

TABLE - V

Comparison of irradiation characteristics of $Al\text{-}OH^-$ centers in unswept, Na-swept and H-swept natural quartz crystals of high-H and low-H concentration.

Treatment given	$Al\text{-}OH^-$ bands (cm^{-1})	Resultant change in relation to as-received or as-swept band strength of natural quartz crystals					
		Unswept		Na-swept		H-swept	
		High-H	Low-H	High-H	Low-H	High-H	Low-H
First irradiation at 77K	3367 Al_1	-40%	-30%	-51%	-30%	-63%	-60%
	3306 Al_s	+210%	+166%	+260%	+166%	+144%	+135%
Room Temperature Irradiation	3367 Al_1	+130%	+108%	+107%	+107%	unaffected	unaffected
	3306 Al_s	+125%	+103%	+178%	+120%	unaffected	unaffected
Final irradiation at 77K	3367 Al_1	-53%	-32%	-74%	-70%	-62%	-53%
	3306 Al_s	+316%	+208%	+386%	+400%	+122%	+122%

Sibley et.al.[25] interpreted that these bands arise from OH^- bonds in two different but closely related sites relative to aluminum ion and that a transfer from one site to the other takes place beyond 230K such that long bonds in AlO_4 tetrahedra become more populated than the short bonds.

It has been shown [5,26] that in cultured quartz at temperature about and above 200K irradiation leads to removal of alkali ions compensating the negative charge at aluminum sites. On the other hand hydrogen is mobile even at low temperatures [25]. The present studies on irradiation characteristics of Li-n3a, Na-n2a and H-n2a bands have shown that irradiation effects on these bands depend upon the presence of alkalis or hydrogen present at these defect centers. Both Li-n3a and Na-n2a bands which incorporate alkali ions show a relatively large decay at room temperature irradiation than at low temperature irradiation. In sharp contrast to the Li-n3a and Na-n2a, the H-n2a center did not show any effect at room temperature irradiation. On the other hand, this center showed a much larger reduction at low temperature irradiation than either Li-n3a or Na-n2a in the alkali swept versions. The present studies suggest that in the case of Li-n3a and Na-n2a centers the processs of radiation-induced dissociation of alkali ions from these centers is similar to the temperature dependence of radiation-induced mobility of alkalis from aluminum. From the H-n2a band also, like the case of H-swept or irradiated cultured quartz the dissociation of protons from $Al\text{-}OH^-$, the protons are removed on irradiation at low temperatures.

Interestingly, all the major bands in H-swept natural quartz following a low temperature irradiation show almost complete recovery upon warming toroom temperature . This suggests that like cultured quartz [25] the mobility of hydrogen is not of long range order at low temperature irradiation. It is known [5,7] that in case of H-swept cultured crystals the room temperature irradiation does not form stable Al-hole centers. Instead, the hydrogen which is knocked from the aluminum under the influence of radiation diffuses back to the aluminum. In the process of back diffusion the transient frequency offsets and acoustic loss are observed [33-41]. It seems that in the H-swept natural crystals an almost similar situation exists with regard to radiation-induced removal of protons from H-n2a centers at room temperature irradiation which does not allow the formation of $[H\text{-}n2a_e+]^0$ - centers from their precursors. Thus, like $Al\text{-}OH^-$ centers the H-n2a centers also hold hydrogen firmly so that no post-room temperature-irradiation steady effects are observable.

SUMMARY AND CONCLUSIONS

The present investigations have made use of infrared absorption measurement technique to monitor the OH^- related point defects in different versions of high-H and low-H natural quartz viz. unswept or Li-swept, Na-swept and H-swept. The description includes irradiation effects at low and room temperature on the major as well as smaller but prominent bands. A number of interesting features have emerged from the present work. They are:

[1] Irradiation at low temperatures (~77K) reduces the strength of almost all the infrared bands. Most of the low strength bands disappear in the background. Subsequent warming to room temperature results in significant recovery (about 85-95%) of the reduced band strength. This suggests that protons released from these low strength bands under the influence of irradiation at low temperature go to shallow traps. On the other hand major bands show a complete recovery by 760K [29] showing that protons from these defects go to deep traps.

[2] The effect of irradiation at 77K as well as 300K on the $Al\text{-}OH^-$ centers in natural quartz have similar characteristics as are shown up by all grades of cultured quartz.

[3] For the same radiation dose (2 Mrad used at every stage) the Li-n3a and Na-n2a centers showed a relatively large decay upon irradiation at 300K than at 77K. On the other hand the H-n2a centers do not show any depletion at 300K while at 77K irradiation their strength is significantly reduced to about 10% of their original value.

[4] A small band at 3595 cm^{-1} is observed in all unswept and swept conditions of natural quartz. This band in most cases, except Na-swept high-H sample, is removed by irradiation at 77K and recovered upon warming to 300K. Also, 300K irradiation did not reduce the intensity of this band attained upon warming to room temperature after the first 77K irradiation. In fact it increased by a small amount in almost all cases.

[5] During the course of irradiation sequence carried out, sometimes appearance of new low strength infrared OH^- bands may be observed in all natural quartz; unswept or swept. This generally occurs at the expense of already existing defect bands.

ACKNOWLEDGMENTS

The author expresses his gratitude and deep appreciation to Professor J.J. Martin for supporting this work. He is also grateful and appreciative to Dr. J.C. King of Sandia National Laboratories for his encouragement to investigate natural crystals also. The Arkansas (low-H) quartz of the present study was given by Professor Martin and the Brazilian (high-H) came from Dr. King. This work was supported under contracts with Sandia National Laboratories, Albuquerque, NM and with the Solid State Sciences Division, Rome Air Development Command of the U.S. Air Force at Oklahoma State University under the principal investigator Professor J.J. Martin.

REFERENCES

1. J.C. King and D.R. Koehler "Radiation Effects on Resonators" in Precision Frequency Control, E.A. Gerber and A. Ballato, eds., Orlando, FL: Academic Press, Vol.1. pp.147-159 (1985).

2. L.E. Halliburton, J.J. Martin and D.R. Koehler "Properties of Piezoelectric Materials" in Precision Frequency Control, E.A. Gerber and A. Ballato, eds., Orlando,FL: Academic Press, Vol.1 pp.1-46 (1985).

3. J.J. Martin., Radiation-induced Frequency Offsets and Acoustic Loss in AT-cut Quartz Crystals" J. Appl. Phys. 68, 5095-5104 (1990).

4. L.E. Halliburton, "Defect Models and Radiation Damage Mechanisms in Alpha-Quartz", Proc. International Conference on Insulators, Univ. of Utah, Salt Lake City, UT, Aug.1984.

5. L.E. Halliburton, N. Koumvakalis, M.E. Markes and J.J. Martin, "Radiation Effects in Crystalline SiO_2: the Role of Aluminum", J. Appl. Phys. vol.57 pp. 3565-3674 (1981).

6. J.A. Weil. "The Aluminum Centers in Alpha Quartz", Radiat. Eff. Vol. 26, pp. 261-265 (1975).

7. L.E. Halliburton, M.E. Markes and J.J. Martin, "Point Defects in Synthetic Quartz: A Survey of Spectroscopic Results" in Proc. 34th Ann. Frequency Control Symposium, EIA, Pp.1-8, May 1980.

8. J.J. Martin "Aluminum-related Acoustic Loss in AT-cut Quartz Crystals", J. Appl. Phys. vol.56, pp. 2536-2540 (1984).

9. S.W.S. McKeever "Thermoluminescence in Quartz and Silica" Radiation Protection Dosimetry, Vol.8, pp.81-98 (1984).

10. L.E. Halliburton, J.J. Martin and W.A. Sibley, "Study of Defects Produced by the Growth, Post Treatment and Fabrication of Quartz", Oklahoma State University, RADC-TR-81-276, Oct. 1981. Copies available from Rome Air Development Command, Hanscom, AFB, MA 01731.

11. L.E. Halliburton, J.J. Martin and W.A. Sibley, "Study of Defects Produced by the Growth, Post Treatment and Fabrication of Quarzt", Oklahoma State University, RADC-TR-83-309, Jan. 1984. Copies available from Rome Air Development Command, Hanscom, AFB, MA 01731.

12. J.A. Weil, "A Review of Electron Spin Resonance Spectroscopy and Its Applications to the Study of Paramagnetic Defects in Crystalline Quartz", Phys. Chem. Minerals, Vol. 10, pp. 149-165 (1984).

13. A. Kats," Hydrogen in Alpha Quartz", Philips Research Repts. Vol. 17, pp. 133-195; 201-279, (1962).

14. D.M. Dodd and D.B. Fraser, "The 3000-3900 cm^{-1} Absorption Bands and Anelasticity in Crystalline Alpha-Quartz", J. Phys. Chem. Solids, Vol. 26, pp. 673-686 (1965).

15. J.J. Martin and A.F Armington, "Effect of Growth Rate on Quartz Defects", J. Crystal Growth, Vol.62, pp. 203-206 (1983).

16. H.G. Lipson, F. Euler and A.F. Armington, "Low Temperature Infrared Absorption of Impurities in High Grade Quartz", Proc. 32nd Ann. Frequency Cont. Symposium, EIA, pp. 11-23 (1978).

17. F. Euler, H.G. Lipson and P.A. Ligor, "Radiation Effects in Quartz Oscillators, Resonators and Materials", Proc. 34th Ann. Frequency Control Symp. EIA, pp. 72-80 (1980).

18. H.G. Lipson, A. Kahan, R.N. Brown and F. Euler, "High Temperature Resonance, Loss and Infrared Characterization of Quartz" Proc. 35th Annual Frequency Control Symp., EIA, pp. 329-334 May 1981.

19. F. Euler, H.G. Lipson, A. Kahan and A.F. Armington,, "Characterization of Alkali Impurities in Quartz", Proc. 36th Annual Frequency Control Symposium, EIA pp. 115-123. (1982).

20. H.G. Lipson, A. Kahan and J.O'Connor, "Aluminum and Hydrogen Defect Centers in Vacuum Swept Quartz", Proc. 37th Annual Frequency Control Symp. IEEE, pp. 169-173 (1983).

21. H.G. Lipson and A. Kahan, "Effects of Vacuum Sweeping and Radiation on Defect Distribution in Quartz" IEEE Trans. Nucl. Sci., Vol. NS-31 pp.1223-1229 (1984).

22. H.G. Lipson and A. Kahan, "Infrared Characterization of Aluminum and Hydrogen Defect Centers in Irradiated Quartz", J. Appl. Phys., Vol. 58, pp. 963-970 (1985).

23. R.N. Brown and A. Kahan, "Optical Absorption of Irradiated Quartz in the Near I.R.". J. Phys. Chem. Solids vol. 36, pp. 467-476 (1975).

24. G.B. Krefft "Effects of High-Temperature Electrolysis on the Coloration Characteristics and OH-Absorption Bands in Alpha-Quartz", Radiation Effects. Vol. 26, pp. 249-259 (1975).

25. W.A. Sibley, J.J. Martin , M.C. Wintersgill and J.D. Brown, "The Effect of Radiation on the OH Infrared Absorption of Quartz Crystals", J. Appl. Phys. Vol. 50, pp. 5449-5452 (1979).

26. L.E. Halliburton, M. Markes, J.J. Martin, S.P. Doherty, N. Koumvakalis, W.A. Sibley, A.F. Armington, and R.N. Brown, "Radiation Effects in Synthetic Quartz: The Role of Electrodiffusion and Radiation-induced Mobility of Interstitial Ions", IEEE Trans. Nucl. Sci., Vol. NS-26, pp. 4851-4856 (1979).

27. H.G. Lipson and A.F. Armington, "Aluminum and Hydroxide Distribution in Cultured Quartz Grown From +X Seeds", J. Crystal Growth, Vol. 80, pp.51-59 (1987).

28. B. Subramaniam, L.E. Halliburton and J.J. Martin, "Radiation Effects in Crystalline SiO_2 : Infrared Absorption from OH -Related Defects", J. Phys. Chem. Solids Vol. 45, pp. 575-579 (1984).

29. Harish Bahadur, "Infrared Characterization of Natural and Cultured Quartz: The Effect of Electrodiffusion and irradiation", J. Appl. Phys., Vol.66, pp. 4973-4982 (1989).

30. J.J. Martin,"Electrodiffusion (Sweeping) of Ions in Quartz - A Review", IEEE Trans. Ultrasonics, Ferroelctrics and Frequency Control, Vol. 35,pp. 288-296 (1988).

31. J.J. Martin, Ho B. Hwang, H. Bahadur and G.A. Berman, "Room Temperature Acoustic Loss Peaks in Quartz" , J. Appl. Phys, Vol. 65, pp. 4666-4671 (1989).

32. J.C. King, "The Anelasticity of Natural and SyntheticQuartz at Low Temperatures", Bell Syst. Tech. J., Vol. 38, pp. 573-602 (1959).

33. B.R. Capone, A. Kahan, R.N. Brown and J.R. Buckmelter, "Quartz Crystal Radiation Effects", IEEE Trans. Nucl. Sci., Vol. NS-17, pp. 217-221 (1970).

34. R.A. Poll and S.L. Ridgway, "Effects of Pulsed Ionizing Radiation on Some SElected Quartz Oscillator Crystals", IEEE Trans. Nucl. Sci., Vol. NS-13, pp. 130-140 (1965).

35. T.J. Young, D.R. Koehler adi R.A. Adams, "Radiation-induced Frequency and Re istance Changes in Electrolyzed High Purity Quartz" Proc. 32nd Ann. Frequency Control Symp., EIA May 1978, pp. 34-42.

36. A. Kahan, F. Euler, H.G. Lipson, C. Chen and L.E.Halliburton, "Radiation Effects in Vacuum Swept Quartz", Proc. 41st Ann. Frequency Control Symp. IEEE, May 1987, pp. 216-221.

37. P. Pelligrini, F. Euler, A. Kahan, T.M. Flanagan, T.F. Wrobel, "Steady State and Transient Radiation Effects in Precision Quartz Oscillators" IEEE Trans. Nucl. Sci., Vol. NS-25, pp.1267-1273 (1976).

38. J.C. King and H.H. Sander, "Transient Changes in Quartz Resonators Following Exposure to Pulse Radiation", Radiat. Eff. Vol. 26, pp. 203-212 (1975).

39. J.C. King and H.H. Sander, "Rapid Annealing of Frequency Change in Crystal Resonators Following Exposure to Pulse X-rays", IEEE Trans. Nucl. Sci, Vol. NS-19, pp. 23-32 (1972).

40. J.C. King and H.H. Sander, "Transient Change in Q and Frequency of AT-cut Crystal Resonators Following Exposure to Pulse X-rays", IEEE Trans. Nucl. Sci., Vol. NS-20, pp. 117-125 (1973).

41. D.R. Koehler and J.J. Martin, "Radiation-induced Transient Acoustic Loss in Quartz Crystals", J. Appl. Phys. Vol. 57, pp. 5205-5209 (1985).

42. J.R. Norton, J.M. Cloren and J.J. Suter "Results form Gamma Ray and Proton Beam Testing of Quartz Resonators" Proc. 38th Annual Frequency Control Symp., IEEE May 1984, pp. 63-72.

43. J. Suter and R. Maurer, "Low and Medium Dose Radiation Sensitivity of Quartz Resonators with Different Al-impurity Content", Proc. 40th Annual Frequency Control Symp.,IEEE, May 1986 , pp. 134-139.

44. J.J. Suter and R.H. Maurer, "Low-and Medium -Dose Radiation Sensitivity of Quartz Resonators with Different Aluminum Impurity Content", IEEE Trans. Ultrason, Ferroelect. and Frequency Control Vol.UFFC-34, pp. 667-673 (1987).

45. T. Flanagan, R. Leadon and D. Shanon, "Evaluation of Mechanisms for Low-Dose Frequency Shifts in Crystal Oscillators", Proc. 40th Annual Frequency Control Symposium, IEEE, May 1986 , pp. 127-133.

46. M.E. Markes and L.E. Halliburton, "Defects in Synthetic Quartz: Radiation-induced Mobility of Interstitial Ions", J. Appl. Phys., Vol. 50, pp. 8172-8180 (1979).

47. R.A. Weeks and M. Abraham, "Electron Spin Resonance of Irradiated Quartz: Atomic Hydrogen", J. Chem Phys. Vol. 42 pp. 68-71 (1965).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ASSESSMENT OF MATERIAL PERFECTION AND OBSERVATION OF VIBRATION MODES IN LITHIUM TANTALATE BY X-RAY TOPOGRAPHY

A. Zarka, B. Capelle, Y. Zheng, J. Détaint*, J. Schwartzel*, C. Joly* and D. Cochet-Muchy**

Laboratoire de Minéralogie-Cristallographie, Université Paris VI et Paris VII,
Associé au C.N.R.S. 4, place Jussieu 75252 Paris Cedex05 France.
*Centre National d'Etudes des Télécommunications dept. PAB/BAG/MCT/CMM,
196, Avenue H.Ravera 92220 Bagneux France
**CRISMATEC 2, rue des Essarts, ZI des Mayencins, 38610 Gières France

ABSTRACT: Using the white radiation delivered by the D.C.I. Synchrotron at the L.U.R.E. (Orsay France), the extended defects of lithium tantalate were studied by reflection and transmission topography. Some low angle subgrains were observed in slices cut perpendicularly to the **Y , Y-45°** and **Z** direction. The typical size of the subgrains is a few cm3. Dislocations were imaged in thin plates ; the density was found in the range of several hundred per cm2.

The large intensity of the synchrotron source has permitted an experimental study of the thickness vibration modes in this material which is highly absorbent for the X-rays. The two thickness shear modes existing in X-cut plates were observed using either conventional or stroboscopic topography. The fundamental fast shear mode and its anharmonics display strong trapping properties. In the contrary, the slow shear mode is not strongly trapped and extends far from the electroded region.

In most cases, the fast shear mode and its anharmonics are coupled to plate modes. Using the time structure of the radiation of the D.C.I. Synchrotron, we have shown that, in most instances, these coupled modes include progressive components with a large standing wave ratio. The existence of these progressive waves, characterized by a large amplitude was observed in all the examined samples. They carry a fraction of the acoustic energy toward the mounting structure of the resonators, and are the origin of the limited Q factors (some thousands) observed in conventional **X-cut** resonators.

I INTRODUCTION

Lithium tantalate was discovered to be piezoelectric and ferroelectric in 1949 in the course of a systematic research made to find new piezoelectric materials. The first single crystals of significant size were synthesized at the Bell Laboratories [1]. Then this material was developed in several other laboratories and countries. It is available in France since the researches made at the C.N.E.T. at the beginning of the years seventies [2]. Like many other oxygen-octahedral ferroelectrics, it has very large electromechanical coupling coefficients, but it is the only one known to have thermally compensated cuts for the extension and flexural modes of bars [3] and for the thickness shear modes of plates [4]. This material is thus very interesting to obtain filters with large bandwidth or oscillators with large shifts.

Since the end of the seventies, this material has received , mostly in Europe, applications to filtering and to frequency generation in professional electronics. More recently, an important development of the applications of lithium tantalate appeared with its use in consumer electronics (S.A.W. Video filters in television). A further important step of development it expected to occur, in the forthcoming years, with intensive applications to filtering in the new pan-european numerical radiotelephone.

Lithium tantalate is obtained by the Czochralsky growth method at very high temperature (1680°C). It can present a quite large non stoechiometry with an usual Li/[Li+Ta] ratio equal to 0.485 which corresponds nearly to the congruent melting point. Very few studies of the crystal perfection were made [5] [6], in particular using X-ray diffraction techniques. They have found that the most common defects are ferroelectric domains, subgrain boundaries, and dislocations.

The use of synchrotron radiation permits now to investigate in good conditions the crystalline quality of this material by transmission X-ray topography. These investigations were extremely difficult, or even not possible with conventional X-ray sources because of the very high X-ray absorption that results of the large Z of tantalum. In this study, using the synchrotron radiation of the D.C.I. storage ring at the L.U.R.E. (Orsay France), we have characterized the crystalline defects of experimental and commercial LiTaO3 crystals using either transmission or reflection topography. The new possibilities given by the intensity and the time structure of the X-ray radiation of the D.C.I. Synchrotron, have permitted the first important experimental study of the thickness vibration modes in lithium tantalate using either the time resolved (stroboscopic) or the conventional (time integrated) X-ray topography techniques [7]. In a recent contribution [8], we have shown that the thickness shear modes of LiTaO3 **X-cut** resonators are strongly coupled to plate modes. In this communication, we presents the results of a detailed experimental study of the influence of the resonator design on the vibration modes of **X-cut** plates. The most important features observed concern the weak trapping characteristics of the slow shear mode, and the fact that the fast shear mode and its anharmonics are nearly in all cases strongly coupled to plate modes. In most instances, these plates modes have progressive components that carry a fraction of the acoustic energy out of the resonator.

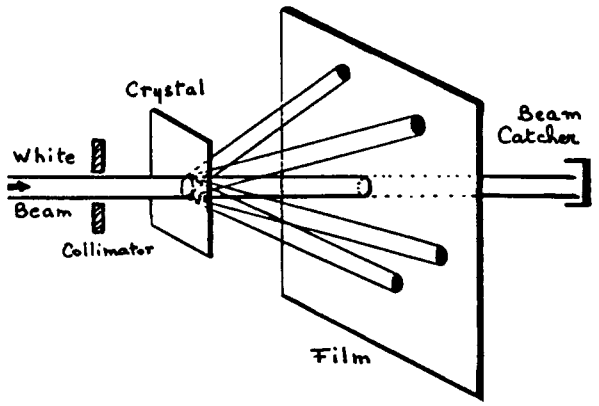
CH2965-2/91/0000-058 $1.00 © 1991 IEEE

II CHARACTERIZATION OF LiTaO3 CRYSTALS.

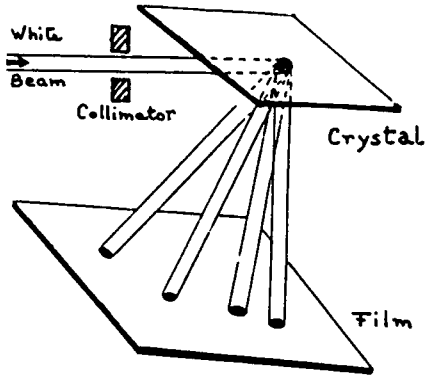
To characterize crystals having a 2 inches diameter pulled in the **Y** direction we have cut **Y** , **Y-45°**and **Z** slices with a thickness of about 2mm. Some of these slices were cut in the middle in the z (or z') direction. Some samples were thinned to .1 mm to make transmission observations. Other plates were cut in commercial or experimental crystals, normal to the x direction to make resonators having thickness ranging from .2mm to 0.07mm.

II.1 Experimental settings.

For crystals too thick to be examined in a transmission Laue setting (figure 1a), a reflection Laue setting using an oblique incidence was used (figure 1b). In this case the film is adjusted parallel to the sample at the distance (10 to 20 cm) choosen to obtain in each case a good compromise between the resolution and the separation of the diffraction spots . For the thinner plates the conventional transmission Laue setting was used (figure 1a).



1a:Laue transmission setting



1b:Laue reflection setting

Figure 1: Laue transmission setting (1a) and oblique incidence setting (reflection) (1b).

II.2 Experimental results.

In the following figures we present typical examples of the observations made. In figure 2 is represented a reflection topograph of nearly one half of a Y-cut slice of a 2" crystal. In this figure corresponding to a diffraction vector having a direction close to the normal to the surface, we can observe a lattice of dislocations and the either black or white contrasts of the subgrain boundaries. The dislocation density in this sample is quite large (several hundred by cm2 on the topograph) since the explored thickness is very small. The contrasts corresponding to the subgrain boundaries are white lines if the relative orientations of the two adjacent subgrains are such that the corresponding diffracted beams are divergent which gives a white region without image. In the case of relative orientations giving a convergence of the diffracted beams a darker contrast (black) is obtained due to the overlapping of the images. It is possible, from the exact knowledge of the experimental setting, and the measurements of the width of the black and white lines, to calculate the relative orientations of different subgrains. The examined part of this plate (3.7x1.7cm) contains about ten subgrains, the larger having an area exceeding two square centimetres.

In figure 3 is represented a reflection topograph covering 3.2cm x 1.7cm of a **Y-45°**slice. About 4 subgrains can be observed in this topograph. We can also observe several indications of the possibility of a continuous variation of the crystalline orientation inside several subgrains.



Figure 2: Reflection topograph of an Y-cut slice.



Figure 3: Reflection topograph of an Y-45° cut slice.



Figure 4: Transmission Laue topograph of a thin (0.1mm) Y-45° plate.



Figure 5: Transmission Laue topograph of another Y-45° plate.

Figure 4 is a transmission topograph of a **Y-45°** thin slice (0.100mm). The area explored in the plate is very nearly 0.8cm x1.6cm. This sample presents a quite large density of dislocations homogeneously distributed across the surface. Only one subgrain boundary can be observed.

In figure 5 is represented an other transmission topograph of a **Y-45°** plate. The size of the area explored is the same as in figure 4. This topograph presents an interesting feature of the dislocation lattice which can be seen to converge, probably to form subgrain-boundaries.

In this topograph , it seems to appear two families of dislocations, the first one corresponding to the dislocations already observed in figure 4 that give images roughly directed in the pulling (y) direction; the second family corresponds to dislocations organized in convergent lattices situated in a plane closer to that of the plate. It is probable that this second family has not the same origin as the first one. This point, which is probably important to understand the mechanism of apparition of the subgrain and also to optimize the growth and cooling conditions of the crystals, need further investigations.

In figure 6 is given a reflection topograph of a large **Z cut** slice of an experimental crystal. Some subgrains boundaries and emergences of the dislocations can be observed.



Figure 6: Reflection topograph of an Z-cut slice.



Figure 7: Reflection Laue topograph of another Z-cut slice. (same scale as in figure 2 and 3).

An other example of a **Z-cut** slice is presented in figure 7. Again, several subgrains can be observed. But, as in figure 6, their surfaces are much larger than what was observed in figures 2 and 3. Moreover, their orientations are generally also much similar. These crystals are thus more adapted to obtain piezoelectric devices having highly reproducible qualities.

Other characterizations were made on the thin **X-cut** plates used as resonators, the corresponding topographs are discussed in the following paragraph.

On the whole, the crystal quality of TaLiO3 seems to be sufficiently good to make bulk and surface wave devices owing to the large dimensions of the observed subgrains.

III EXPERIMENTAL STUDY OF THE VIBRATION MODES IN X-CUT RESONATORS .

III.1 The thickness shear modes of X-cut plates.

The one dimensional piezoelectric theory of thickness modes [4] indicates that only the two shear modes can be exited in **X-cut** plates. A slow shear mode which has a phase velocity of about 3368 m/s , a small coupling coefficient (k=3.5%) a large negative temperature coefficient (-44.10-6/°C) and a fast shear mode having a phase velocity of 4212 m/s, a large coupling coefficient (k=45 %) and a small first order temperature coefficient which vanishes at a temperature depending on the stoechiometry of the crystal. The shear displacements of these modes which are orthogonal, are not exactly situated in directions contained in the plane of the plate, but, they make a small angle with this plane (quasi-shear modes). By convention, we will refer in the following the displacements to a system of axis constituted by the projection of the displacement of the fast shear mode on the plane of the plate (x'_1), the normal to the plate(x'_2=x), and the projection of the slow shear mode on the plane of the plate (x'_3). The difference between this system and that which is constituted by the three one-dimensional eigen-modes of an **X-cut** plate is small, and, in practice, lower than the experimental precision obtained in the present investigations.

On figure 8, a typical response curve of an **X-cut** resonator (figure 9) is represented. On this figure we can observe the fundamental slow shear mode and the fundamental fast shear mode. The shape of the resonance curve of the fast shear mode is typical of the large coupling coefficient (large frequency difference between the resonance and the antiresonance), it indicates also that this mode is trapped under the electrodes. This fact is observed in most cases, for resonators with electrodes having dimensions much lower than the plate ones, that are made using the thickness shear modes (fundamental, overtones, and their anharmonics up to a certain rank).

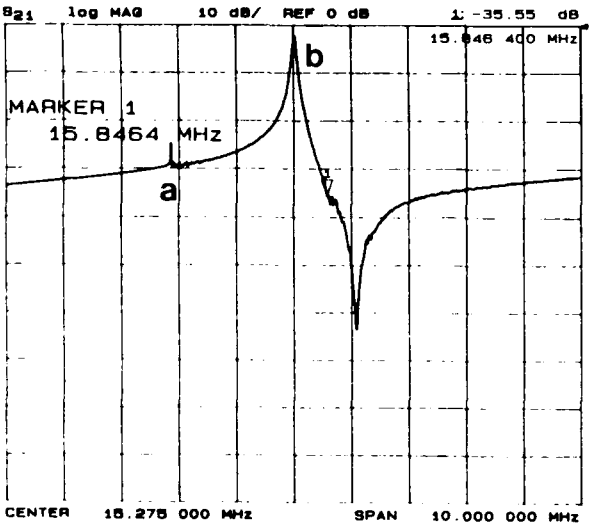


Figure 8: Typical response curve of an X-cut resonator *with small electrodes: (a) fundamental of the slow shear.* (b) fundamental fast shear.

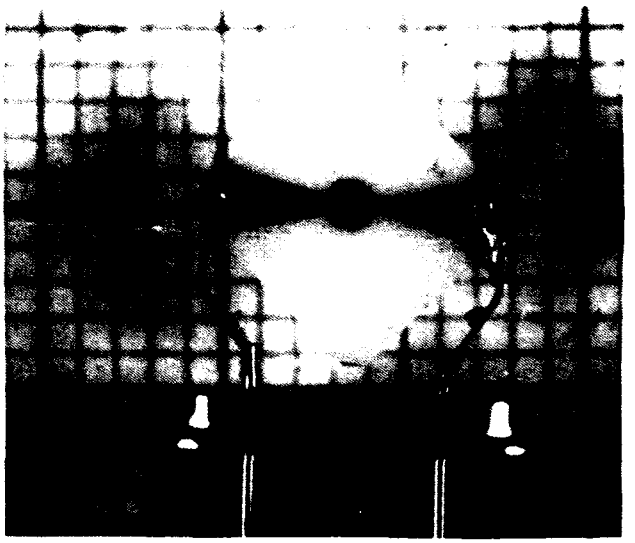


Figure 9: Photography of an X-cut resonator designed to have no anharmonics of the fast shear.

This energy trapping of the modes in the electroded region results of the frequency lowering due to the mass loading and to the electrical effect of the electrodes.

The very large coupling coefficient of the fast shear induces a very important electrical frequency lowering (of the order of $4k^2/\pi^2$ in a crude approximation) and an important difference between the cut-off frequency of the electroded region and that of the unelectroded region [9] [10]. This fact leads to a very fast attenuation of the fundamental shear mode outside the electroded region and also to the need to use very small electrodes to have no anharmonics of the fast shear.

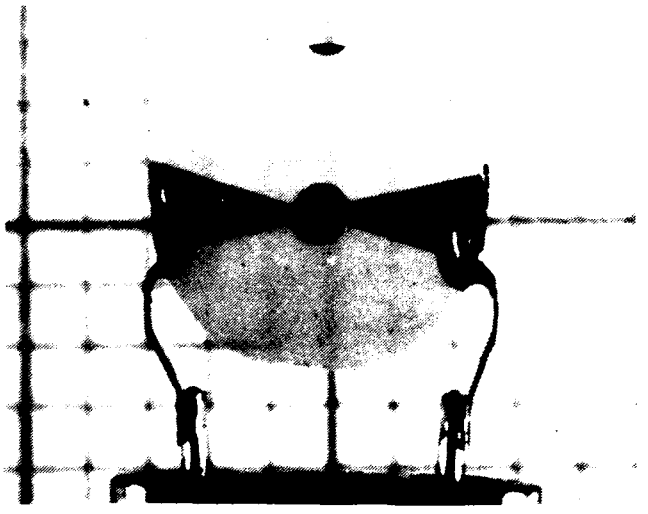


Figure 10: Photography of an X-cut resonator designed to have two symmetrical anharmonics of the fast shear.

The resonator represented in figure 10 was made with larger electrodes (having a much larger ratio of their lateral dimensions to the thickness of the plate) so that the fast shear mode has two anharmonics modes (figure 11). This was done in order to study the properties of these modes that presents an interest in some devices (the monolithic filters) and are in other cases very undesirable (discrete resonators for filter applications). The anharmonics modes that can be predicted by the two [11] or three dimensional theories [12] , have a displacement similar to that of the corresponding shear modes in the thickness direction (x'_2), but this shear component (essentially u_1) have lateral variations presenting one or several sign reversal (nodal lines). In lithium tantalate like in other materials the apparition of the anharmonics is mainly governed by the lateral dimensions of the electrodes, the electrical lowering (constant for a given mode and a given orientation), and, the mass-loading of the electrodes.

The slow shear mode that can be observed in figures 8 and 11 has a quite surprising frequency response presenting several anharmonics.

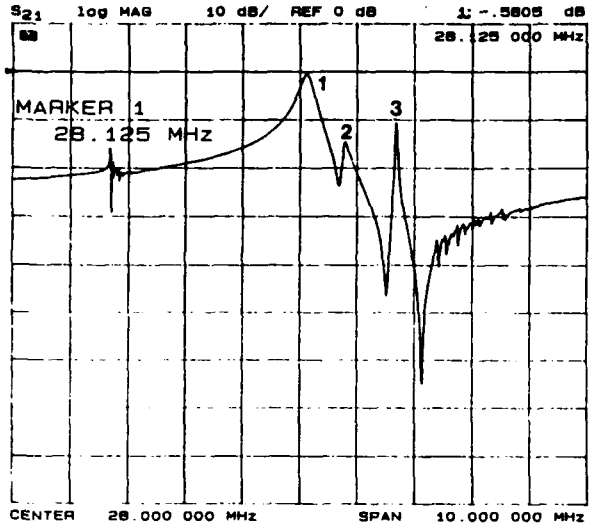


Figure 11: Electrical response of the resonator represented in Figure 10.

III.2 Experimental setting.

The observations were made using a transmission Laue setting. With this setting, in the case of a normal incidence, Laue diagrams similar to that represented in figure 12 are obtained. This diagram is relative to a resonator similar to that represented in figure 9 , vibrating on the fundamental fast shear mode near 15.846400 MHz . In this Laue pattern, we can observe an extinction of the component u_1 of the fast shear, in a direction (x'_3) which make an angle of about 45° with the edges of the rectangular **X-cut** plates. The edges of this plate are in the y and z directions , so that this experiment is a confirmation of the theoretical value (-47° from the y axis) of the direction of the polarization (u_1) of the fast shear mode. The extinction of the contrast relative to the u_1 component result of the cancellation of the product **g.u** in the x'_3 direction . Another Laue diagram of the same resonator is represented in figure 13. In this case, the incident beam was delimited using a circular diaphragm and the distance between the sample and the film was reduced. This permits to obtain a better accuracy in the measurements and an easier comparison with computed simulations of the diagrams made to identify the diffraction vectors of the different spots.

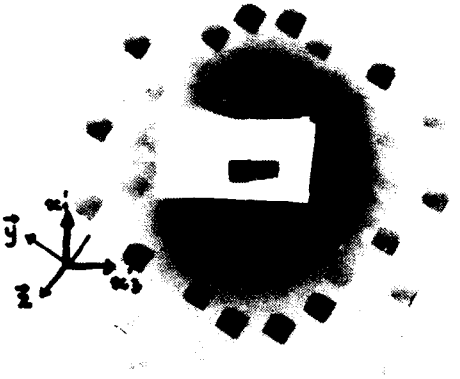


Figure 12: Laue pattern of an X-cut resonator (normal incidence). The projections of u_1 (axe x'_1) and of u_3 (axe x'_3) are drawn on the diagram.

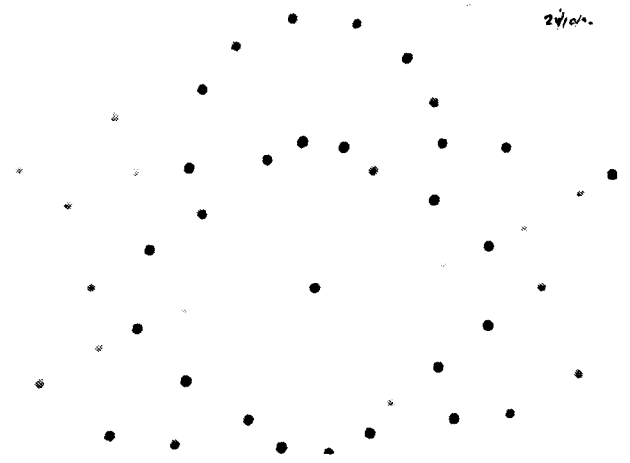
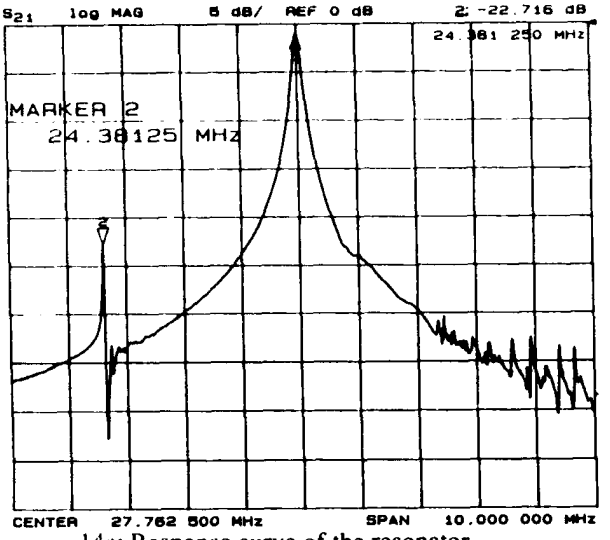


Figure 13: Laue pattern of an X-cut resonator (normal incidence and reduced sample- film distance).

III.3 X-cut resonator with small and symmetrical electrodes.

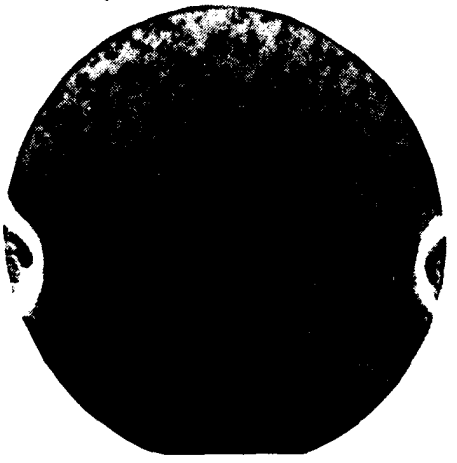
This resonator has a thickness of about 70 microns and 5 mm of diameter . In figure 14a is represented the response curve . Two modes are exited .The slow shear mode appears on the low frequency part (left) of the diagram and the fast shear mode is on the right.

Conventional topographs of this resonator have been obtained using the transmission Laue setting .



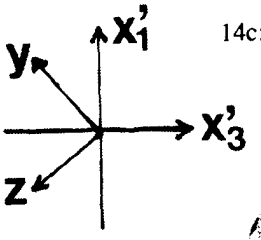
14a: Response curve of the resonator .

Respectively in Figures 14b and 14c are shown the topographs that visualise the u_1 component of the thickness vibration . In figures 14d and 14e are given the topographs which visualise the u_3 components of the slow and the fast shear modes . The slow shear mode shows a u_3 components elongated in the x'_3 direction (Figure 14d) and a little component u_1 (Figure 14b). The fast shear mode u_1 is very much confined (Figure 14c) but it is coupled with the plate mode (probably of a flexure type) which extend up in the x'_1 direction . In Figure 14e a little u_3 component can be observed in the case of the fast shear mode . The quality of the sample can be observed in the first topograph (Figure 14b) . In fact ,the density of the dislocations is not so large and it may be observed that the mounting clips are in the x'_3 direction .



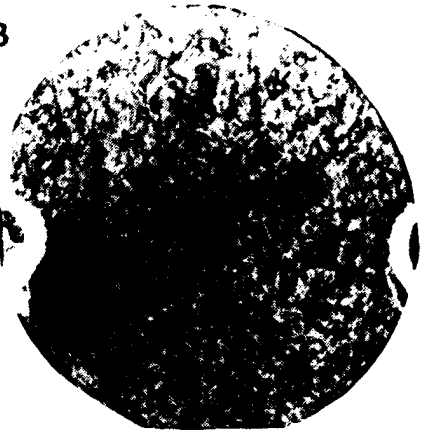


14c: u_1 component of the fast shear mode .





14d: u_3 component of the slow shear mode .



14e: u_3 component of the fast shear mode .

Figure 14: X-cut resonator with small and symmetrical electrodes .

III.4 Fast shear mode of a resonator with small and symmetrical electrodes.

On figure 15 is represented the responses curves of a resonator having small rectangular electrodes elongated in the x'_3 direction . The two curves corresponds respectively to the response of the resonator alone and of this resonator with a serial capacitance adjusted so that the resonance frequency was exactly equal to 9 times the recurrence frequency of the synchrotron (28.523520 MHz). This resonator was examined using the stroboscopic and the conventional technique [7] [13] [14]. On figure 16.a is represented the u_1 component of fundamental fast shear mode observed in stroboscopic conditions (observation during a 1ns time slot of the vibration mode at a fixed time of the vibration period), The figure 16.b represents the same component observed using conventional topography (i.e. with a continuous shift of the 1ns observation time slot over the period , an integration being performed by the film). From these figures, it can be concluded that the conventional trapped shear mode is coupled to a progressive component that extends over the total surface of the plate. Using the stroboscopic setting , and a slow variation of the phase shift between the resonator excitation and the synchrotron current, it is possible to observe with an X-ray sensitive video camera, that the non-stationnary component is progressing towards the mounting clips (true stroboscopic observation). Thus this progressive component , which has a large standing wave ratio, carries a fraction of the energy outside of the resonator. This fact was observed in all the examined resonators and is most probably responsible of the rather low Q factor of the fundamental fast shear mode.

The mode shape of the u_1 component is elongated in the x'_1 direction whereas the electrodes are elongated in the x'_3 direction; this indicates that the **X-cut** of lithium tantalate presents a large in plane anisotropy [10]. On figures 16c and 16d are represented, very nearly, the u_3 components, observed respectively in stroboscopic and conventional conditions. Again, we can observe a weak stationary trapped part (antisymmetrical with one nodal line in both the x'_1 and x'_3 directions) and a progressive part. The fact that an essentially shear mode has a small u_3 component results of the finite dimensions of the electroded region. This type of mode can also have an u_2 component, (this was observed in several instances).

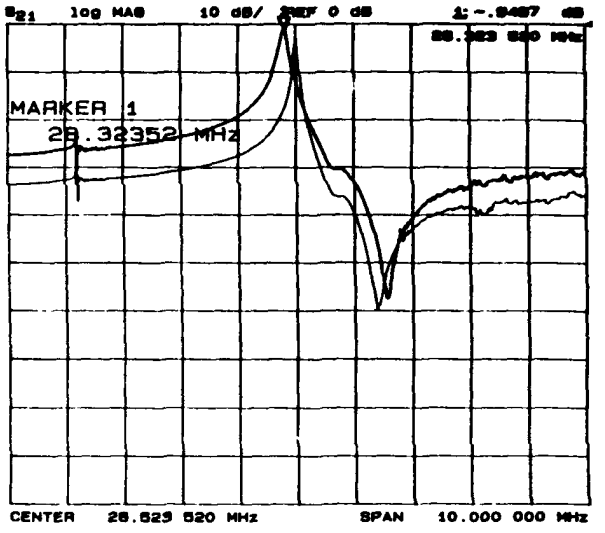
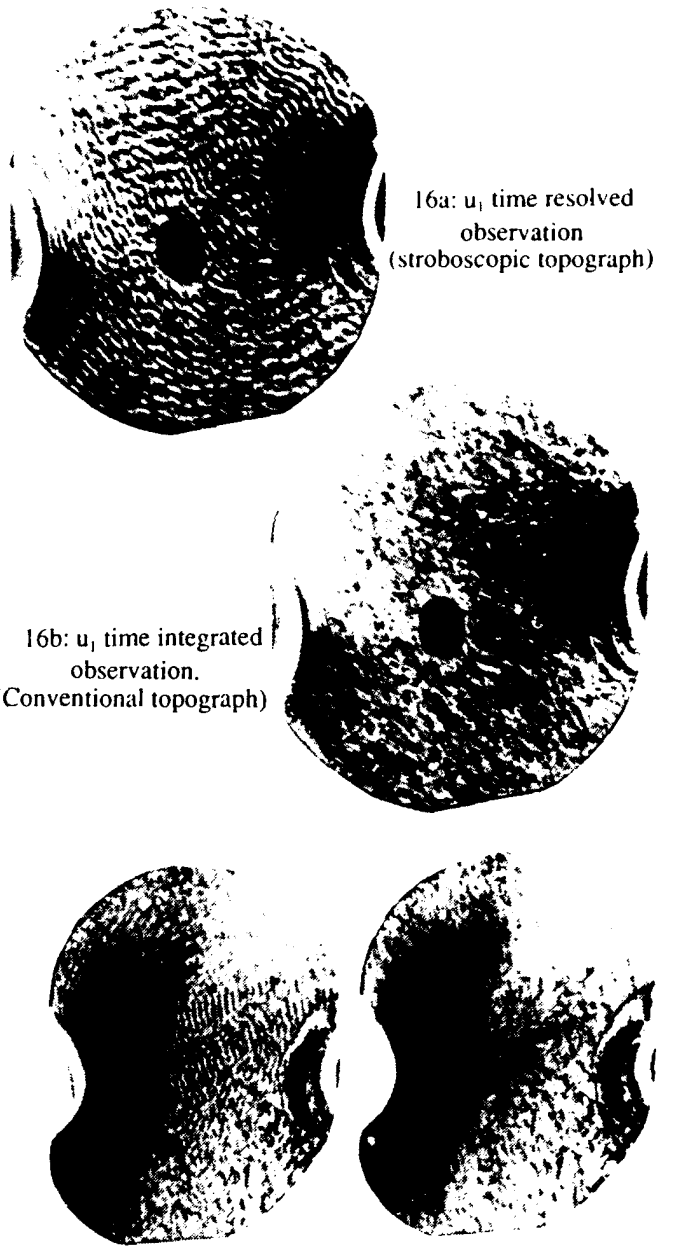


Figure 15 : Electrical response of a resonator with small and symmetrical rectangular electrodes.



16a: u_1 time resolved observation (stroboscopic topograph)

16b: u_1 time integrated observation. (Conventional topograph)

16c: u_3 time resolved observation. (Stroboscopic topograph)

16d: u_3 time integrated observation. (Conventional topograph)

Figure 16: Observed fast shear mode of a resonator having symmetrical small electrodes .

III.5 Mode of a resonator with symmetrical larges electrodes (plate modes coupled to the fast shear).

A resonator having large elliptical electrodes (.9x.5mm) elongated in the x'_3 direction with the electrode tabs in the x'_1 direction was examined using the stroboscopic and conventional technique. The results are given in figure 17 for the fundamental fast shear mode. Three diffraction vectors are considered in this figure (one on each rank of the figure) . The first one (Figures 17a, 17b, 17c) is such that nearly only the u_1 component of the vibration mode can be visualized if the plate is vibrating (a very small fraction of u_2 may also appear). The second and the third ones are such that nearly only the u_3 component is imaged. Respectively figures 17d, 17e, 17f and figures 17g, 17h, 17i). The three columns of this figure correspond respectively to topographs obtained without excitation, stroboscopic topographs and conventional topographs. On the topograph taken without vibration

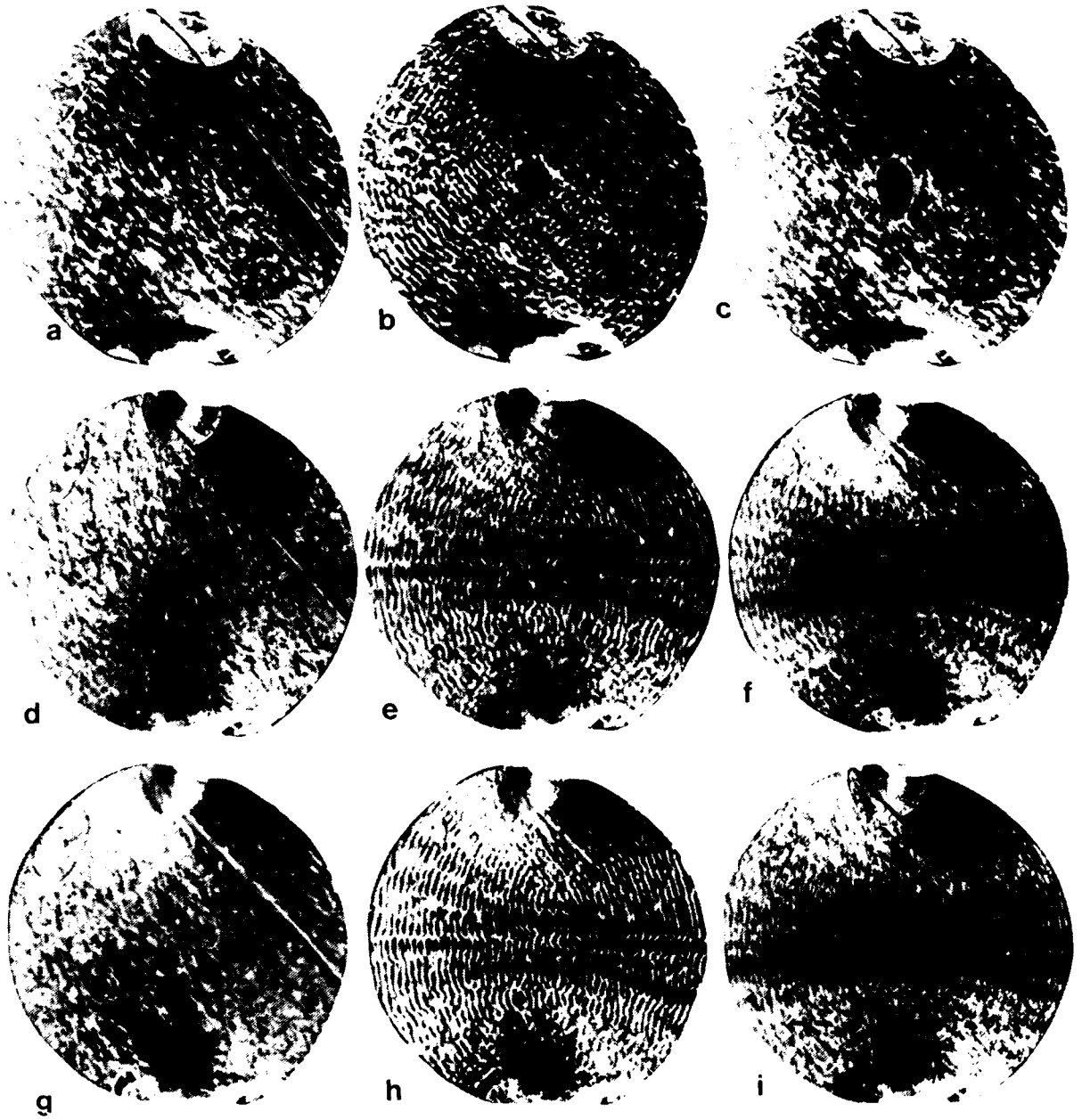


Figure 17: Observation of the fundamental fast shear mode in a resonator with large symmetrical elliptical electrodes.

17a, b, c: images obtained with a diffraction vector nearly in the u_1 direction.
17d, e, f: images obtained with a diffraction vector nearly in the u_3 direction.
17g, h, i: images obtained with a diffraction vector close to the u_3 direction.
(In the columns: without excitation, stroboscopic topographs, conventional topographs).

(first column), it is possible to observe the characteristic contrasts of the dislocations existing in the plate and one subgrain boundary. In figure 17a and partially in figure 17g, these contrasts are inverted. This results of the very important value of the μ.t product (specific absorption x thickness) even for this very thin sample (70 microns).

On the conventional topographs of the third column that were extracted of the same Laue diagram, the shape of the u_1 component (figure 17c) is quite conventional except that the mode shape is elongated in the x'_1 direction whereas the electrodes are elongated in the x'_3 direction (effect of the lateral anisotropy). In this case the u_3 component is very important and extends, mainly in the direction of the x'_3 axis, up to the edge of the plate (figures 17f and 17i). On the stroboscopic topographs of the second column, taken also from the same film, there are formal evidences of the existence of strong u_1 (figure 17b) and u_3 (figures 17e and 17h) progressive components. Although no separate examination of the u_2 component have been made (it requires reflection topography), the examination of the diffractions spots on the transmission topographs seems indicate that this component is also non negligible.

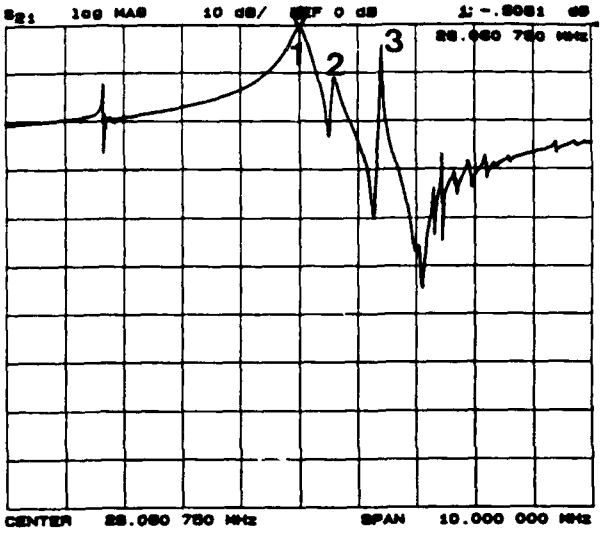


Figure 18a:Response curve of the resonator with large electrodes.

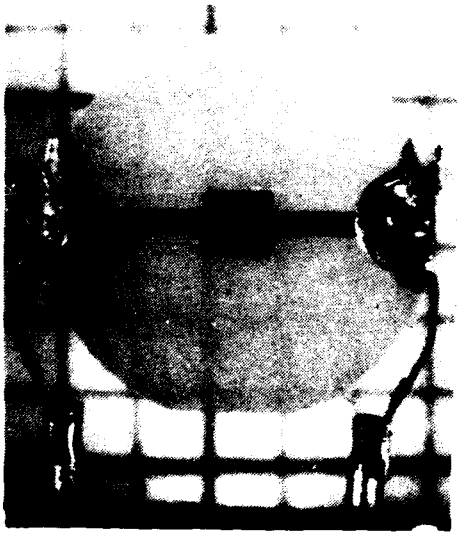


Figure 18b:Photography of the resonator (large square electrodes).

III.6 Mode of a resonator having symmetrical larges electrodes (anharmonics of the fast shear).

On Figure 18 are represented the u_1 and u_3 components of the three fast shear modes (figure 18a) of a resonator having large square electrodes (figure 18b). These components were observed using the conventional topography technique. Figures 18c, 18d, 18e are respectively relative to the u_1 components of the fundamental mode and of the two symmetrical anharmonics having two nodal lines (the only ones that are trapped in these conditions). The first anharmonic appearing after the main mode has nodal lines in the x'_1 direction whereas the second has its nodal lines in the x'_3 direction. A very weak stationary coupled u_1 component, extending largely beyond the electroded region, can be observed in figures 18d and 18e. The fundamental mode has a weak trapped u_3 component (figure 18f) whereas the two anharmonics have large u_3 stationary components extending up to the fixations of the plate (figures 18g and 18h).

III.7 Modes of a resonator having large unsymmetrical electrodes. (Slow shear and fast shear)

This resonator was electroded using two rectangular masks that have a large pentagonal overlapping region which constitutes an electrode system not symmetrical with respect to the x'_1 and x'_3 axis and that is not centred in the plate (figure 19). The response curve of this resonator is given on figure 20. On this curve it is possible to observe the fundamental slow shear mode and several of its anharmonics and the fundamental fast shear mode followed by three strong anharmonics. A detailed view of the response near the slow shear mode is given in figure 21.

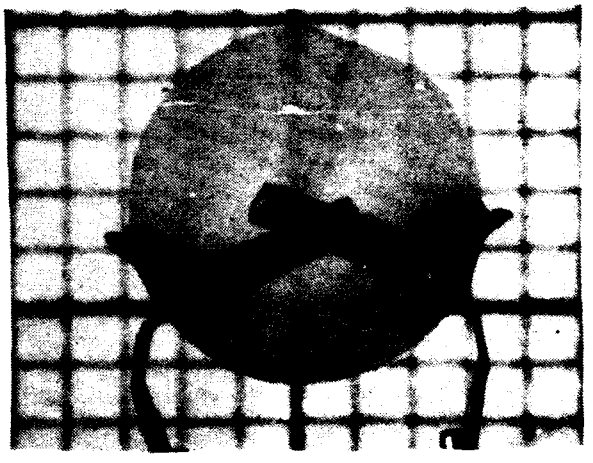
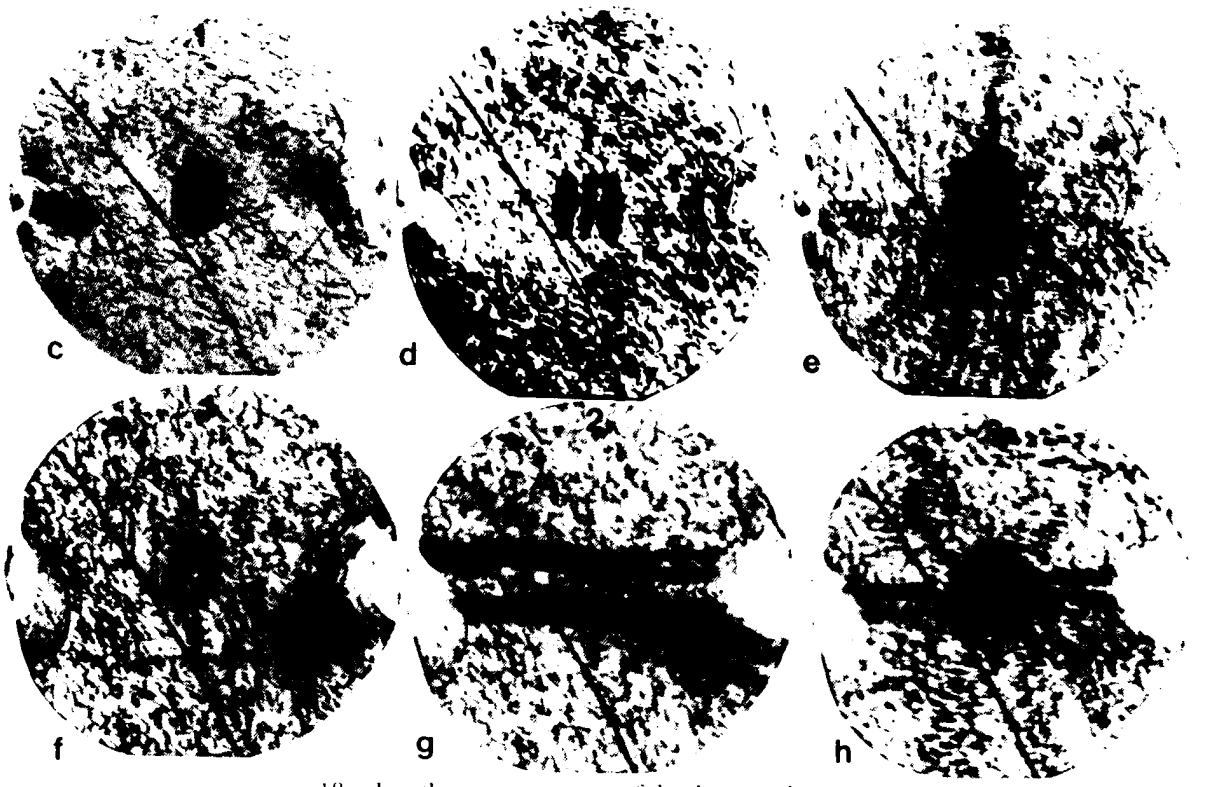


Figure 19: Photography of a resonator having unsymmetrical electrodes not centred in the plate.



18c, d, e: the u_1 components of the three modes.
18f, g, h: the u_3 components of the three modes.

Figure 18: The fundamental and the two trapped anharmonics of the fast shear.

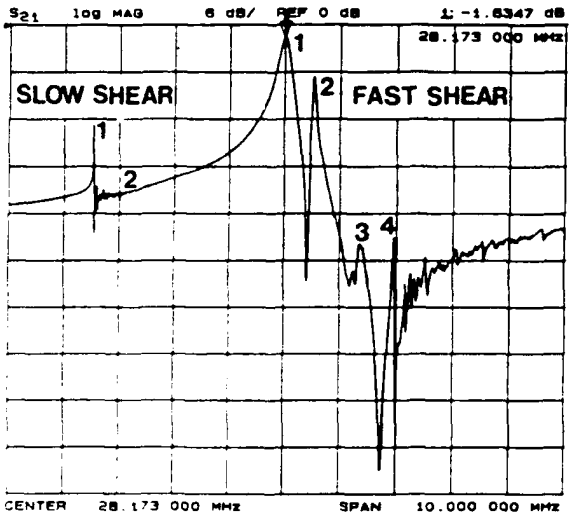


Figure 20. Electrical response of the resonator represented in Figure 19. (fundamental slow shear, fundamental fast shear and theirs anharmonics).

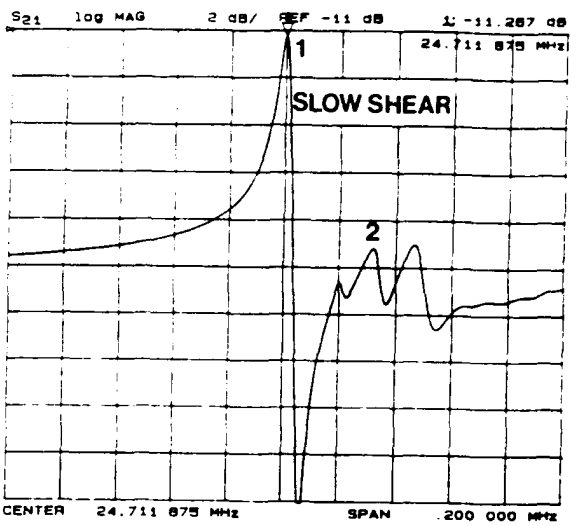
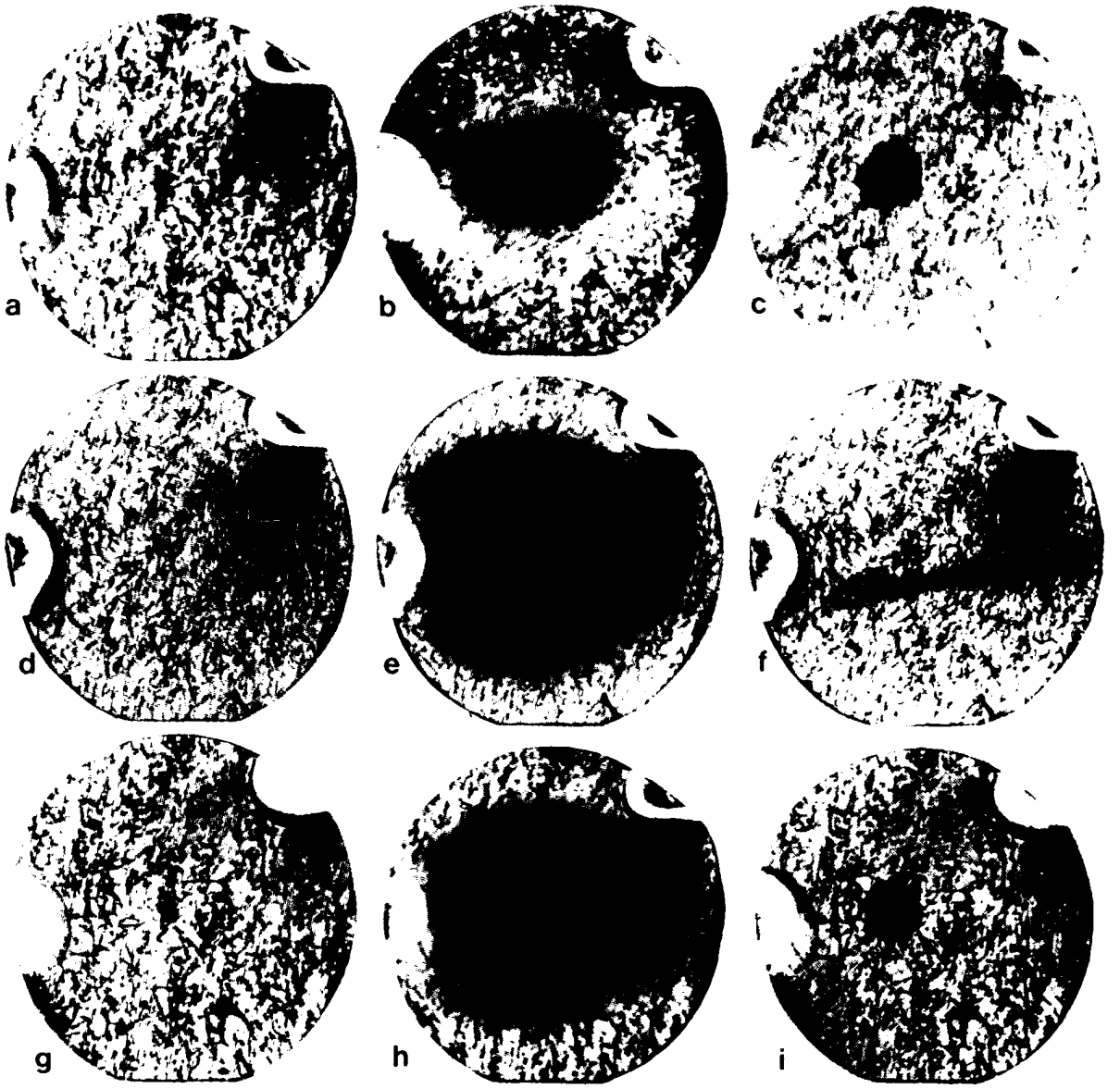


Figure 21: Electrical response near the slow shear.

In figure 22 are given conventional (time integrated) topographs taken with three diffraction vectors respectively in the u_1, in the u_3 and, in a composite direction (one diffraction vector on each line). The topographs of the first column concern the resonator with no excitation. In these topographs the dislocations existing in the material are imaged. In the second column of figure 22 are respectively represented the u_1 component (22b), the u_3 component (22e) and a mix of the three components (22i) of the fundamental slow shear mode. From these observations we can induce that this mode is very weakly trapped and also that it is strongly coupled to the fast shear one (large value of the u_1 component). The topographs of the last column are relative to the fundamental fast shear mode. As expected from theoretical considerations, the u_1 component of this mode is very strongly trapped under the electrodes and displays their pentagonal shape. The u_3 component of the fast shear extends up to the edge of the plate.

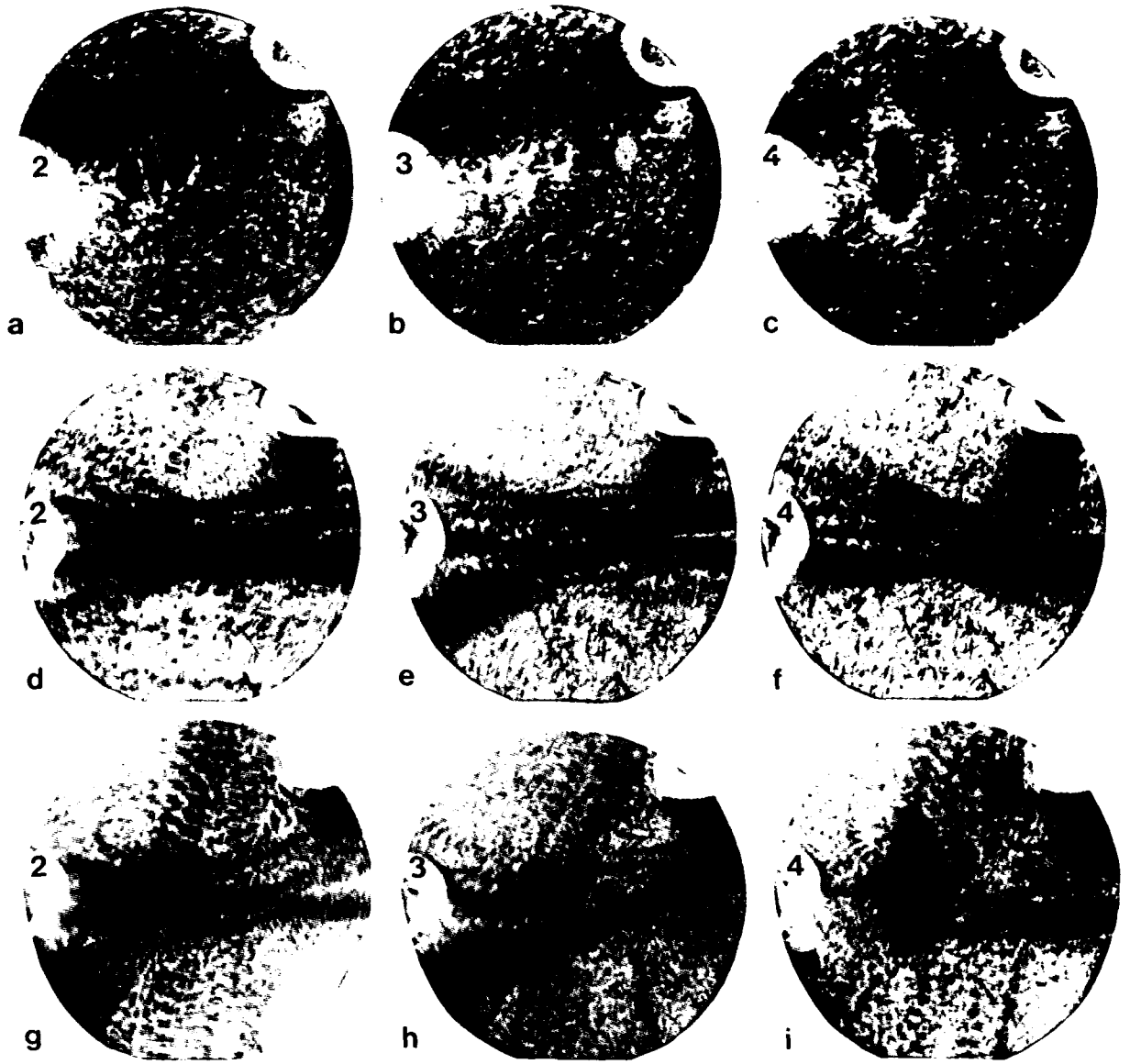


22a, b, c: images obtained with a diffraction vector nearly in the u_1 direction.
22d, e, f: images obtained with a diffraction vector nearly in the u_3 direction.
22g, h, i: images obtained with a diffraction vector in a composite direction .
(in the columns: without excitation, slow shear, fast shear).

Figure 22: Topographs of the resonator represented in Figure 19, without vibration, fundamental slow shear mode and fundamental fast shear mode.

Conventional topographs of the three anharmonics of the fast shear are given in figure 23. The u_1 component of the first anharmonic which displays, as expected 2 nodal lines in the x'_3 direction , is symmetrical in the x'_1 and in the x'_3 direction. This mode has a similar nature as the first anharmonic excited by electrodes symmetrical with respect to the x'_1 and x'_3 axis (figure 18) . The second anharmonic has a very unconventional weak u_1 component with much probably one nodal line in one direction and two in the other. The u_1 component of the third anharmonic has two nodal line in the x'_3 direction and is very nearly symmetrical.

The u_3 components of these anharmonics are very large and again extend up to the edge of the plate. Their structure, in particular in the electroded region , is probably due to some coupling with the slow shear modes. The comparison of the topographs representing the u_1 and the u_3 components of these modes and those representing a mix of the u_1, u_2, and, u_3 components, indicate the existence in these latter topographs, of a non negligible u_2 component.



23a, b, c: modes imaged with a diffraction vector nearly in the u_1 direction.
23d, e, f: modes imaged with a diffraction vector nearly in the u_3 direction.
23g, h, i: modes imaged with a diffraction vector in a composite direction.

Figure 23: The fast shear anharmonics of the resonator with unsymmetrical electrodes.

III.8 Resonator with symmetrical large electrodes (anharmonics of the slow shear modes).

The response curve of this resonator is displayed in the figure 24 . On this curve we can observe, after the slow shear mode, several "spurious" resonances that are coupled to the fundamental fast shear mode and to the anharmonics of the fast shear. In 1 and 2 are indicated 2 modes of this type that where particularly investigated. The drawing presented in figure 25 shows that the mounting clips were put in the the x'_1 direction instead of the usual x'_3. In the figure 26a is given a conventional topograph of this resonator obtained without vibration . The black and white contrasts may correspond to ferroelectric domains. On figure 26b is presented the topograph of a spurious mode existing after the slow shear (u_3 component). On figure 26c is presented another spurious mode existing between the slow and the fast shear modes. Since it was observed that the u_1 component of these modes is very weak, it appears that these modes are untrapped anharmonics of the slow shear that extend up to the plate edge in the x'_3 direction . This type of mode has an important electrical response each time that the mounting clips are not in the x'_3 direction. They are very much attenuated when the mounting clips are in the x'_3 direction.

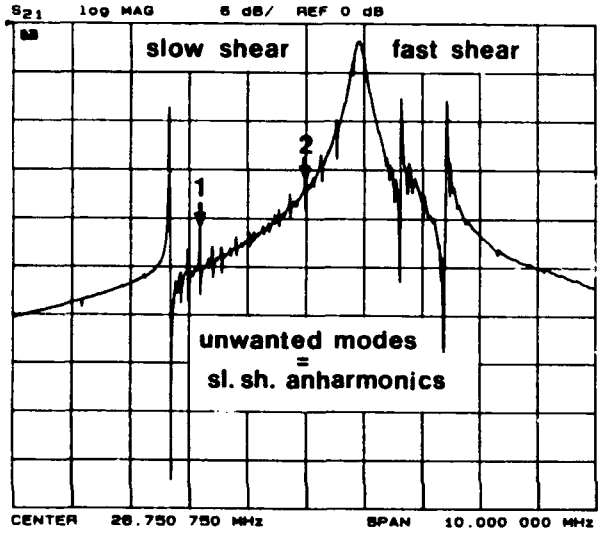


Figure 24: Resonator with symmetrical large electrodes : response curve .

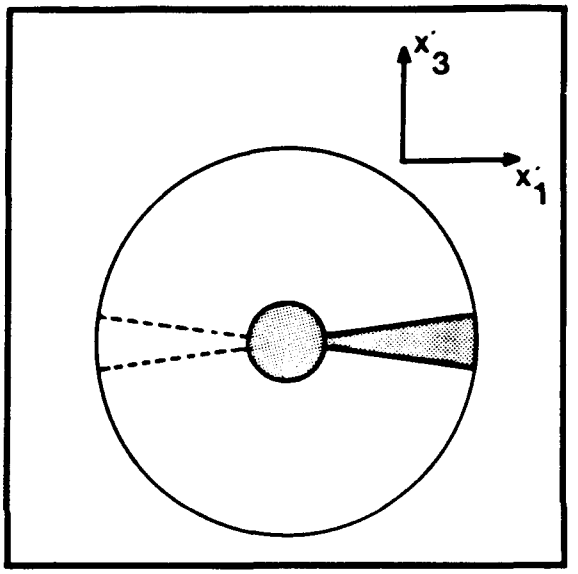
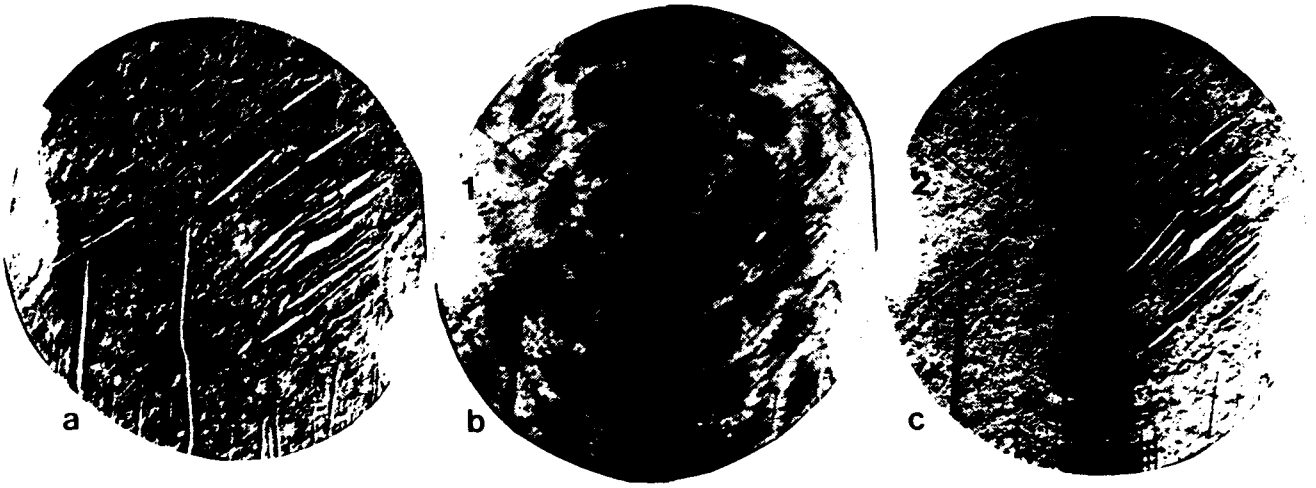


Figure 25: Geometry of the same resonator as in figure 24.



26a: Conventional topograph of this resonator obtained without vibration.
26b: Topograph of a spurious mode existing after the slow shear.
26c: Topograph of another spurious mode situated between the slow and the fast shear modes.

Figure 26: Topographs of the same resonator as in figure 24.

IV CONCLUSION.

In this study the crystalline perfection of LiTaO3 sample was examined using synchrotron radiation X-ray topography. In thick plates imaged in reflection, a quite large density of dislocations and the existence of subgrain boundaries were observed. The large intensity of the white synchrotron radiation has permitted to obtain good transmission topographs of thin plates. In these topographs individual dislocations were well resolved and, it has appeared that, at least, two types of dislocations can exist in this material. On the whole, the quality of the material appear to be sufficiently good for the piezoelectric applications owing to the large dimensions of the subgrains. The new experimental crystals have a greater perfection.

The large intensity and the time structure of the X-ray radiation of the D.C.I. Synchrotron of the L.U.R.E. (Orsay France) have permitted to make the first important experimental study of the thickness vibration modes of X-cut plates using either the conventional or the stroboscopic technique.

This study has revealed several features of the slow and of the fast shear modes excited in these plates. The most important one is the fact that the fundamental fast shear mode and its anharmonics were, in all the observed cases," coupled" to progressive components extending on the whole surface of the plate. These progressive components lead to the dissipation of a small fraction of the energy of the mode in the mounting structure of the resonator and are most probably at the origin of the limited Q factor of the resonators. In many instances , the fast shear modes, possess stationary components also extending all over the plate that may be due to a coupling to plates modes. The importance of the u_3 component of these modes and, its particular structure most often observed, indicate probably a strong coupling to the slow shear. The fundamental of the slow shear and its anharmonics are very weakly trapped by the small electrodes needed to have a good response for the fast shear.

On the whole, this study has shown the great complexity of the fundamental shear modes in a highly piezoelectric material and the great possibilities offered by the synchrotron radiations X-ray topography. These experiments indicate also the fundamental interest of using the modern theories of plate vibrations that include the possibilities of describing the coupling of several kinds of plate modes to the thickness ones. Only this kind of theories can permit a sufficiently accurate description of the actual vibration modes in highly piezoelectric materials and an optimization of the design of such devices.

Acknowledgments: The authors wish to acknowledge R.Coquillé, P.Auvray of CNET LAB/OCM/MPA and J.P.Aubry of CEPE for fruitful discussions about the lithium tantalate material, and B.Dubouis of CNET PAB/STC/ECG for its important contribution in the electronics permitting the stroboscopic experiments. They acknowledge the artful contribution of M. A.Jeanne-Michaud to the quality of the topographs. They thanks MM. Y.Bernard, G.Caumont and Mrs.A.Daniel for their technical assistance.

REFERENCES

[1] A.A.Ballman J. Amer. Ceram. Soc. vol.48 p.112 (1965).
[2]M. Nolf, R. Coquillé, A. Lebret, P. Auvray, R.Lançon, J. Détaint. Rapport interne CNET CPM/PMT39 EST/DEF1592.
[3] M. Onoe, T.Shinada, K.Itoh, S. Miyazaki, Proc.27th Annual Frequency Control Symposium p 42 (1973).
[4] J.Détaint, R.Lançon. Proc.30th Annual Frequency Control Symposium (1976).
[5] P. Auvray CNET/LAB/OCM in reference [2] and personnal communication.
[6] J.C.Doukhan, P.Cordier, N.Doukhan, M.Fragneau, J.P.Aubry. Proc. 43rd Annual Frequency Control Symposium p.497 (1989).
[7] B.Capelle, J.Détaint, A.Zarka, Y.Zheng, J. Schwartzel Proc.44th Annual Frequency Control Symposium p.416 (1990).
[8] J.Detaint, J.Schwartzel, H.Carru, R.Lefèvre, C.Joly, B.Capelle, Y.Zheng, A.Zarka Proc.44th Annual Frequency Control Symposium p.337 (1990).
[9] H.F.Tiersten J. Acoust. Soc. Amer. vol. 59 n°4 p.879 (1976).
[10] J.Détaint, J.Schwartzel, C.Joly, E.Philippot. Proc. 41st Annual Frequency Control Symposium p. 314 (1987).
[11] J.W. Burgess , M.C.Hales Proc. IEE vol 123 N°6 (1976).
[12] R.C.Peach I.E.E.E. trans. ultrason. ferroelectrics, frequency control v35 n°5 p.593 (1988).
[13] Y. Zheng, A.Zarka, B.Capelle, J.Détaint, J.Schwartzel Acta cryst. vol. A45 p 275 (1989).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

EVALUATION OF NONLINEAR ELASTIC COEFFICIENT CAUSING FREQUENCY SHIFTS IN AT-CUT RESONATORS

Yasuo Hirose and Yasuo Tsuzuki

Faculty of Engineering, Yokohama National University
156 Tokiwadai, Hodogaya-ku, Yokohama, Japan 240

Abstract

This paper describes an experimental evaluation scheme of the forth order nonlinear elastic constant of quartz C_{6666} using AT-cut resonators. The evaluation has been made from measured quadratic frequency changes against excitation currents and actual vibration amplitude distributions of thickness-shear vibrations measured by the optical fiber vibration sensor. We have obtained more reasonable estimates of C_{6666} compared with those previously reported with a wide variation.

Introduction

The AT-cut quartz resonators have been in extensive use as ultra-high stability resonators for selection and control of frequency. It has been known, however, that they exhibit nonlinear behaviors at high excitation levels. For example, their resonance frequencies slightly increase nearly in proportion to the squared values of excitation current amplitudes, known as the so-called amplitude frequency effect (the A-F effect)[1].

To elucidate these nonlinear behaviors, a number of theoretical and experimental work have so far been made. For instance, Gagnepain considered in his early work that this frequency change was due to the forth order nonlinear elastic constant C_{6666}. He obtained estimate values of this constant on the order of 10^{14} N/m^2 using three different AT-cut resonators and assuming uniform vibration distributions over the electrodes[1]. Later he showed, in his review paper, values of $3\sim8\times10^{13}$ N/m^2 obtained from a one dimensional model[2]. Tiersten made detailed theoretical nonlinear analyses of thickness-shear vibrations of quartz resonators[3]-[6], and Smythe made mainly experimental work on nonlinear resonance and intermodulation of quartz resonators[7],[8]. They gave estimate values of C_{6666} of the order of $10^{12}\sim10^{13}$ N/m^2. These estimate values, however, vary considerably from each other, by an order of magnitude in some cases, and consequently, it appears that this constant has not been well known. It would be of great significance, from a practical design point of view in particular, if we could estimate in advance the resonator nonlinear performance, by knowing this nonlinear constant.

With this in mind, we describe in this paper an evaluation scheme of this nonlinear elastic constant. The scheme uses both quadratic frequency changes against excitation and vibration amplitude distributions over the resonator surface, measured by the high sensitivity optical fiber vibration sensor reported earlier[9]. We have obtained more reasonable estimates of C_{6666} compared with those previously reported with a wide variation, although we can not preclude the possibility of the existence of other unknown nonlinear factors affecting the present results.

Amplitude-Frequency Effect of the AT-cut Resonator

Fig. 1 shows an example of the amplitude-frequency effect of a circular planer trapped-energy AT-cut resonator as shown in Fig.2. Relative frequency changes of the third and fifth overtone vibrations are plotted as functions of the squared value of excitation current amplitude I_e^2. Frequency measurements were performed by obtaining frequencies at which the resonator impedances become purely resistive. The resonator has a fundamental resonance frequency of nearly 10 MHz, a radius of 6.2 mm, a thickness of 0.165 mm, and gold evaporated electrodes with a radius of 3.0 mm. Great care was taken to support the resonator to avoid any additional nonlinear effects caused by stresses. It is seen that the relative frequency changes are almost perfectly proportional to the squared values of excitation.

Also shown by dotted lines are the resonance resistances R_S of the resonator for these overtone vibrations. Nearly constant resistances over a wide range of excitation indicate that the resonator would be of sufficient quality for the evaluation of the nonlinear elastic constant.

CH2965-2/91/0000-072 $1.00 © 1991 IEEE

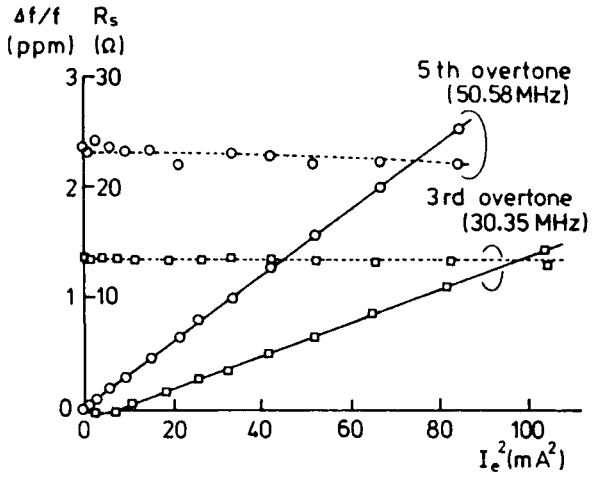


Fig.1. Relative frequency changes against squared excitation currents.

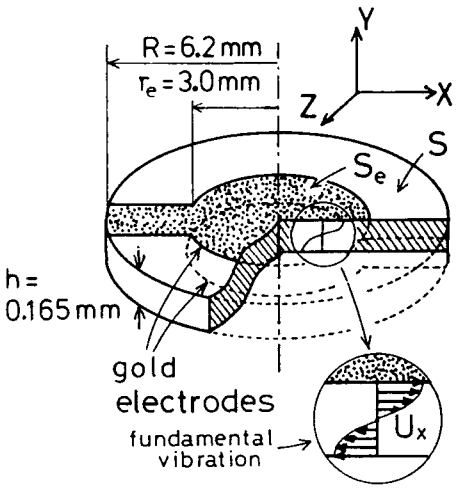


Fig.2. AT-cut quartz resonator.

Evaluation Method of the Nonlinear Elastic Constant

In this section we describe the evaluation method of the forth order nonlinear elastic constant C_{6666} which has been believed to be mainly responsible for the amplitude-frequency effect. In this method, the constant is evaluated from two experimentally measured characteristics of the resonator, i.e., frequency changes due to excitation level, and vibration displacement distributions measured by the optical fiber vibration sensor[9].

Since the dominant deformation in the AT-cut resonators is the thickness shear deformation S_6, the strain energy function W is approximately expressed as[10]

$$W - (1/2)C_{66}S_6^2 + (1/24)C_{6666}S_6^4 = W_0 + \Delta W \qquad (1)$$

In this equation, C_{66} is the linear elastic constant of quartz, and terms which become zero when time averaged, and higher order terms with negligible contributions are omitted.

The resonance frequency change of an AT-cut resonator when this energy is varied by changing excitation is given by

$$\frac{\Delta f}{f} = \frac{1}{2} \times \frac{\frac{1}{2\pi}\int_0^{2\pi}\int_V \Delta W \, dV \, d(\omega t)}{\frac{1}{2\pi}\int_0^{2\pi}\int_V W_0 \, dV \, d(\omega t)} \qquad \cdots\cdots\cdots\cdots (2)$$

where ω is the resonant angular frequency of the vibration, and V, the volume of the resonator.

The x direction vibration displacement u_x and the thickness shear deformation S_6 are approximately given by the following equations, with n being the order of an overtone vibration, and h, the thickness of the resonator.

$$u_x = U(x,z)\sin(n\pi y/h)\sin\omega t \qquad (3)$$

$$S_6 = \partial u_x/\partial y = U(x,z)(n\pi/h) \cdot \cos(n\pi y/h)\sin\omega t \qquad (4)$$

Substituting eq.(4) into the terms W_0 and ΔW in eq.(1) and then substituting these terms into eq.(2), we obtain

$$\frac{\Delta f}{f} = \frac{3}{128}\frac{C_{6666}}{C_{66}}(n\pi/h)^2 \frac{\int_S U^4 dS}{\int_S U^2 dS} \qquad (5)$$

where S is the resonator surface area. For convenience, we use the normalized vibration distribution F = F(x,z) defined as

$$U(x,z) = U_m F(x,z) \qquad (6)$$

where U_m is the maximum vibration amplitude. Substituting this distribution F into eq.(5) yields

$$\frac{\Delta f}{f} = \frac{3}{128}\frac{C_{6666}}{C_{66}}(n\pi/h)^2 U_m^2 \frac{\int_S F^4 dS}{\int_S F^2 dS} \qquad \cdots\cdots\cdots\cdots (7)$$

On the other hand, the induced total charge on an electrode is obtained by integrating the y direction electric displacement component over the electrode surface area S_e. Since the resonator current is the time derivative of this charge, the effective value I_e of the current is given as

$$I_e = (\sqrt{2}\,\omega\, e_{26}'/h) \int_{S_e} U\, dS$$

$$= (\sqrt{2}\,\omega\, e_{26}' U_a/h) \int_{S_e} F\, dS \qquad (8)$$

where e_{26}' is the piezoelectric constant of the AT-cut plate. Substitution of the vibration amplitude U_a obtained from eq.(8) into eq.(7) yields the following equation giving the quadratic frequency shift $\Delta f/f$ against the resonator current density j_e, in agreement with the experimental observations.

$$\frac{\Delta f}{f} = \frac{3}{256}\,\frac{C_{6666}}{C_{66}}\,(n\pi)^2\,\frac{I_e^2}{(\omega\, e_{26}' S_e)^2}\,J$$

$$= k\, j_e^2 \qquad (9)$$

where

$$J = \frac{S_e^2 \int_S F^4\, dS}{\left[\int_{S_e} F\, dS\right]^2 \times \left[\int_S F^2\, dS\right]} \qquad (10)$$

$$k = \frac{3}{256}\,\frac{C_{6666}}{C_{66}}\,\frac{(n\pi)^2}{(\omega\, e_{26}' S_e)^2}\,J \qquad (11)$$

$$j_e = I_e/S_e \qquad \text{(current density, A/m}^2\text{)} \qquad (12)$$

The value of k appearing in eqs.(9) and (11), the amplitude frequency effect coefficient (the A-F coefficient), can be determined from the inclination of a measured frequency change against squared excitation currents. The integration term J depends only on the shape of the vibration distribution F(x,z), and can be determined by measurements. If uniform vibration distributions over the resonator surface are assumed, i.e., F(x,z) = 1, then eq.(9) reduces to the following equation derived by Gagnepain et. al.[1].

$$\frac{\Delta f}{f} = \frac{3n^2}{256}\,\frac{C_{6666}}{C_{66}}\,\frac{\pi^2 I_0^2}{(e_{26}' S \omega_0)^2} \qquad (13)$$

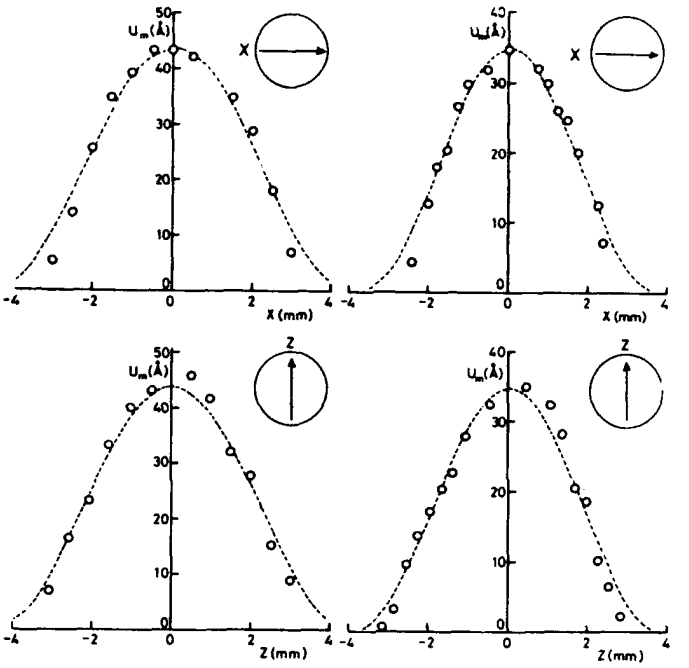
From eq.(11), the forth order nonlinear elastic constant C_{6666} is finally expressed as

$$C_{6666} = \frac{256}{3}\,C_{66}\,\frac{(2f\, e_{26}')^2}{n^2}\,\frac{k}{J} \qquad (14)$$

where f is the resonance frequency, $C_{66} = 0.2902\times10^{11}$ N/m^2, and $e_{26}' = -0.9474\times10^{-1}$ C/m^2. It should be noted that the forth order nonlinear elastic constant C_{6666} depends only on the shape of vibration distributions F(x,z) and not on their magnitudes.

Vibration Distributions of the AT-cut Resonator

As mentioned earlier, it is crucial for accurate evaluation of the nonlinear elastic coefficient to know the vibration distributions over the resonator surface.



(a) Third overtone
f_3 = 30.35 MHz
I_e = 10 mA

(b) Fifth overtone
f_5 = 50.58 MHz
I_e = 10 mA

Fig.3. Measured amplitude distributions of the third and fifth thickness-shear vibrations.

Recently quantitative vibration measurements of quartz resonators in the VHF frequency range have become possible by the optical fiber interferometric vibration sensor[9]. In this sensor, the amplitude of a vibration is determined by homodyne detection of phase modulation of interfering lights caused by the vibration to be measured.

Figs. 3(a) and (b) show amplitude distributions of the x direction vibration component of the third and fifth overtone thickness-shear vibrations along the x and z axes of the resonator, measured by this sensor. Measurements were performed by scanning focused laser beams over the resonator surface. Although the measured values have some dispersions, the nature of the vibrations is well revealed, i.e., similar vibration profiles along the x and z axes and narrower profiles for the fifth overtone compared with the third overtone vibration, for instance.

Resonator Currents and Motional Inductances

As pointed out previously, the normalized vibration distribution F(x,z) plays a key role in the evaluation of the nonlinear coefficient, and can be measured by the optical fiber vibration sensor. To assure that the measured vibration distributions

are of reasonable accuracy, resonator currents and motional inductances were calculated from the measured vibration distributions and compared with directly measured values.

The resonator current I_e is expressed by eq.(8), and is listed below again for the sake of convenience.

$$I_e = (\sqrt{2}\omega e_{26}' U_m h) \int_{S_e} F \, dS \qquad (15)$$

The motional inductance L_e is obtained by equating the time-averaged kinetic energy of the resonator to the electromagnetic energy stored in its equivalent circuit, and is expressed as

$$L_e = \frac{\rho h^3}{8 e_{26}'^2} \frac{\int_S F^2 \, dS}{\left[\int_{S_e} F \, dS \right]^2} \qquad (16)$$

where $\rho = 2.65 \times 10^3$ Kg/m^3 is the density of quartz. This equation indicates that the motional inductance L_e depends solely on the shape $F(x,z)$ of the vibration distribution, being independent of its magnitude.

Table 1 gives calculated and measured resonator currents and motional inductances for the third and fifth overtone vibrations. The calculated values were obtained by numerically evaluating the integral terms in eqs.(15) and (16) using the measured vibration distributions $F(x,z)$ as shown in Fig.3, and the measured values were obtained by use of a network analyzer. The results agree reasonably well with each other, and hence the measured vibration distributions are of reasonable accuracy.

Vibration mode (resonance freq.)	Resonator current I_e (calc.)	(meas.)	Motional inductance L_e (calc.)	(meas.)
3rd overtone (30.35 MHz)	10.9 mA	10.0 mA	7.03 mH	6.84 mH
5th overtone (50.58 MHz)	11.2 mA	10.0 mA	7.9 mH	7.57 mH

Table 1 Resonator currents and motional inductances.

Evaluation of the Forth Order Nonlinear Elastic Constant C_{6666}

As stated earlier, the A-F coefficient k in eq.(9) and the integration term J given by eq.(10) are needed for the evaluation of the forth order nonlinear coefficient C_{6666}, both of which are determined by measurements in this work.

The values of k were found to be 0.12×10^{-10} m^4/A^2 and 0.24×10^{-10} m^4/A^2 for the third and fifth overtone vibrations respectively from the slope of the frequency shift curves shown in Fig. 1. The electrode radius of $r_e = 3.0$ mm was used in this calculation.

The integration terms J were determined by numerically evaluating the three integrals in eq.(10) in terms of vibration distributions $F = F(x,z)$, obtained from the measured vibration distributions as shown in Fig. 3. The values of J were found to be 1.7 and 2.7 for the third and fifth overtone vibrations respectively.

Using these k and J values, the nonlinear elastic constant C_{6666} has been evaluated by eq.(14), and listed in Table 2 along with the values previously reported. The values, $C_{66} = 0.2902 \times 10^{11}$ N/m^2 and $e_{26}' = -0.9474 \times 10^{-1}$ C/m^2, were used in the calculation. Although there is still some difference between the values obtained from the third and fifth overtone vibrations, the difference is much smaller than the previously reported values with a wide variation, by an order of magnitude in some cases. Although the measured results are

Present Work	Previous Work
6.5×10^{13} N/m^2 (Evaluated from the 3rd overtone)	$7 \times 10^{13} \sim 1.5 \times 10^{14}$ N/m^2 (Ref. 1)
	$3 \times 10^{13} \sim 8 \times 10^{13}$ N/m^2 (From one dimensional model)(Ref. 2)
8.1×10^{13} N/m^2 (Evaluated from the 5th overtone)	$4 \times 10^{12} \sim 8 \times 10^{12}$ N/m^2 (From intermod. meas.) (Refs. 2, 3, 7)

Table 2 Evaluated estimate values of the forth order nonlinear elastic coefficient C_{6666}.

not extensive, and moreover, other unknown nonlinear factors could affect the present results for C_{6666}, we believe that we have obtained at least more reasonable estimates of C_{6666} compared with those previously reported.

Conclusions

We have described an experimental evaluation scheme of the forth order nonlinear elastic coefficient C_{6666}, using both the measured amplitude-frequency effect of the AT-cut resonator, and the actual thickness-shear vibration distributions measured by the optical fiber vibration sensor. We have obtained more reasonable estimates of C_{6666} compared with those with a wide variation previously reported in a number of work.

References

(1) W. P. Mason and R. N. Thurston;"Physical Acoustics", Vol.XI, Academic Press.

(2) J. J. Gagnepain ;"Nonlinear Properties of Quartz Crystal and Quartz Resonators", Proc. 35th Annual Freq. Control Symposium, pp.14-30 (1981).

(3) H. F. Tiersten ; "Analysis of Intermodulation in Rotated Y-cut Quartz Thickness-Shear Resonators", Proc. 28th Annual Freq. Control Symposium, pp.1-4 (1974).

(4) H. F. Tiersten; "Analysis of Nonlinear Resonance in Rotated Y-cut Quartz Thickness-Shear Resonators", Proc. 29th Annual Freq. Control Symposium, pp.49-53 (1975).

(5) H. F. Tiersten ; "An Analysis of Nonlinear Resonance in Electroded Contoured AT- and SC-cut Quartz Crystal Resonators", Proc. 38th Annual Freq. Control Symposium, pp.132-140 (1984).

(6) H. F. Tiersten and D. S. Stevens; "The evaluation of the Coefficient of Nonlinear Resonance for SC-cut Quartz Resonators", Proc. 39th Annual Freq. Control Symposium, pp.325-332 (1985).

(7) R. C. Smythe ; "Intermodulation in Thickness-Shear Resonators", Proc. 28th Annual Freq. Control Symposium, pp.5-7 (1974).

(8) R. C. Smythe and P. E. Morley;"Experimental Evaluation of the Effective Non-linear Elastic Constant for Trapped Energy and Contoured Resonators", Proc. 39th Annual Freq. Control Symposium, pp.317-324 (1985).

(9) Y. Hirose and Y. Tsuzuki ; "Measurement Method of VHF Elastic Vibrations by Optical Fiber Interferometric Sensing", Proc. 43rd Annual Symposium on Freq. Control, pp.360-364 (1989).

(10) K. Brugger ; "Thermodynamic Definition of Higher Order Elastic Coefficients, Physical Rev., Vol. 133, No. 6A, pp. A1611-A1612 (Mar. 1964).

Discussion

Question: K.K. Tuladhar, Oscilloquartz, Switzerland

Do you have any plans to extend current technique to cover contoured AT-cut (Fundamental - 5th overtone modes) crystals and SC-cut crystals?

Answer: Y. Hirose & Y. Tsuzuki

We have a plan to refine our evaluation technique and to extend it to cover contoured AT-cut crystals.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

THE AGING OF BULK ACOUSTIC WAVE RESONATORS, FILTERS AND OSCILLATORS

John R. Vig* and Thrygve R. Meeker**

* U.S. Army Electronics Technology and Devices Laboratory (LABCOM), Ft. Monmouth, NJ 07703-5601

** 2956 Lindberg Ave.,
Allentown, PA 18103

Abstract.

The aging of quartz crystal resonators, filters, and oscillators is reviewed, including the topics of: the impacts of aging, typical aging characteristics, aging specifications, aging mechanisms, temperature dependence of aging, frequency and overtone dependence of aging, drive level dependence of aging, the effects of aging interruptions, the dependence of aging on material and mode type, the state-of-the-art in low-aging oscillators, and aging acceleration effects. The aging mechanisms discussed include: contamination transfer effects, stress effects, electrode effects, diffusion effects, changes in the quartz material, and circuit and other electrical changes. Isothermal and thermal step stress aging acceleration methods are also reviewed.

Introduction.

The long term time dependence of a frequency source's frequency is often called frequency aging. Since other long term changes can occur, such as changes in the elements of the equivalent circuit of a crystal resonator, and in a crystal oscillator's input power ("input power aging" is defined in MIL-O-55310 [1]), clarity of expression requires that the particular parameter of interest be specifically stated. In the remainder of this paper, "aging" will mean frequency aging, unless otherwise specified.

"Aging" and "drift" have occasionally been used interchangeably in the frequency control literature. However, in 1990, recognizing the "need for common terminology for the unambiguous specification and description of frequency and time standard systems," the CCIR adopted a glossary of terms and definitions [2].

According to this glossary, aging is "the systematic change in frequency with time due to internal changes in the oscillator." Added to the definition is: "Note - It is the frequency change with time when factors external to the oscillator (environment, power supply, etc.) are kept constant." Drift is defined as "the systematic change in frequency with time of an oscillator." Drift is due to aging plus changes in the environment and other factors external to the oscillator. Aging is what one specifies and what one measures during oscillator evaluation. Drift is what one observes in an application. For example, the drift of an oscillator in a spacecraft is due to (the algebraic sum of) aging and frequency changes due to radiation, temperature changes in the spacecraft, and power supply changes.

The CCIR definitions of aging and drift are now incorporated into the military specifications for crystal oscillators, MIL-O-55310 [1], and have been recommended by the IEEE Standards Coordinating Committee 27 on Time and Frequency for inclusion in the next edition of the IEEE Standard Dictionary of Electrical and Electronics Terms.

Random changes of frequency with time, called short term stability (or, more correctly, short term instability), are characterized in the time domain by the two-sample deviation (also called the square-root of the Allan variance), and in the frequency domain by the various measures of phase noise, as defined in IEEE Standard 1139-1988 [3].

The very accurate and precise measurement of frequency allows the observation of very small changes in a resonator. It is generally true that the crystal resonance frequency is a more sensitive measure of the state of the resonator system than other measurements that can be made. It has therefore been very difficult to apply measurements other than frequency to studies of the nature of the aging process, particularly for low-aging devices.

Many aging measurements have been reported, but few have included a detailed scientific or statistical study of the aging processes. Our understanding of resonator aging processes is often based on indirect evidence gained from

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

other fields (such as material science, and the science of solid surfaces), and on process developments that seemed "sensible" for general reasons, and which were followed by an evaluation of the resulting aging. For high aging rate resonators, advanced surface science measurements seem to support the use of the "sensible" processes by qualitatively correlating with the measured aging.

The "sensible" processes generally include suitable crystal surface preparation, high level of cleanliness during assembly of the resonator, reasonably well controlled crystal mounting and processing, and hermetically sealing the resonator into a clean enclosure. High temperature in-process baking is often used at some stage of resonator fabrication, such as after mounting, or before final frequency adjustment or sealing. Sometimes a burn-in after sealing is used to "preage" the resonator before shipment, and to test for processing deviations.

Precision crystal resonators need to be protected from operating system environments by sealing in an appropriate package. Typically, these packages have been glass or metal. For these glass or metal packaged resonators the resulting resonator system is very complex, in different ways. Mechanical, chemical, and electrical interactions between the package, the enclosed materials, and the resonator all can cause aging.

In a crystal oscillator, in addition to the aging of the resonator, aging of some of the electrical elements (for example, series inductance or load capacitance) can also change the oscillator's frequency. Changes in the shape and configuration of the metal leads, and deformation of circuit boards and enclosures can also produce aging.

Some of the aging of crystal filters is probably associated with the aging of associated electrical elements. Otherwise, crystal filters are subject to the same aging processes as other types of resonators.

The following sections of this paper contain discussions and references to published reports on the topics that are relevant to our understanding of the aging of resonators, crystal oscillators, and crystal filters. The emphasis is on reviewing aging processes, and the literature since 1983. The pre-1983 literature was reviewed by Gerber [4].

The Impacts of Aging.

In most device applications, the frequency of the oscillator must remain within specified limits in order for the device to operate properly. When aging shifts the frequency beyond the limits, the oscillator must be recalibrated. Since crystal oscillators have a finite frequency adjustment range, oscillator aging can cause "end-of-life" when the aging rate is so high that it produces a frequency offset that exceeds the frequency adjustment range. (This rarely happens in properly designed oscillators.)

Soon after an oscillator's calibration, the frequency offset due to aging is usually small compared to the frequency offsets due to environmental (especially temperature) changes, however, eventually, the frequency offset due to aging becomes the dominant frequency offset. For example, a state-of-the-art oven controlled crystal oscillator is specified to have a frequency vs. temperature stability of 1×10^{-9}, an aging rate of 1×10^{-10} per day, and a frequency adjustment range of 4×10^{-7}. Assuming a worst-case linear aging rate at the specified rate, the frequency offset due to aging will equal the worst case temperature induced offset on the tenth day after calibration, and the oscillator will no longer be able to be calibrated after 4000 days (i.e., 11 years).

Similarly in clocks, aging eventually becomes the dominant time error source. Whereas the time error varies linearly with elapsed time due to the relatively constant clock rate errors (i.e., frequency offsets) caused by environmental effects, it varies as elapsed time squared when the clock error is due to (linear) aging. In filter applications, recalibration is usually not an option. Aging can gradually degrade performance until an end-of-life frequency shift is reached.

Aging Characteristics.

Aging can be positive or negative. Occasionally, a reversal in aging direction is observed. Typical (computer simulated) aging behaviors are illustrated in Fig. 1. The curve showing the reversal is the sum of the other two curves. So far, the simplest proposed aging model showing a reversal consists of two simultaneously acting aging mechanisms, as shown in Fig. 1.

The aging rate of an oscillator is highest when it is first turned on. Since the aging rate during the first few days to weeks is generally significantly higher than during subsequent intervals, the first part of the aging curve is sometimes referred to as "initial aging" or "stabilization period." At a constant temperature, aging usually has an approximately logarithmic dependence on time. When the temperature of a crystal unit is changed, e.g., when an OCXO is turned off, and turned on at a later time, a new aging cycle may start, as illustrated in Fig. 2.

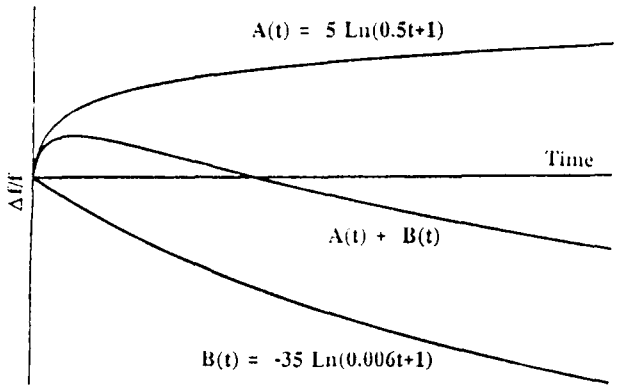


Figure 1 - Typical Aging Behaviors.

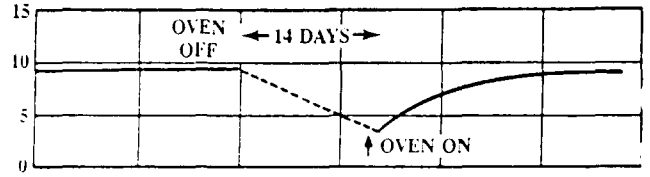


Figure 2 - Typical Aging Behavior, Including an Interruption (frequency change in parts in 10^9 vs. time in days).

The aging rates of typical commercially available crystal oscillators (in 1991) range from 5 to 10 ppm per year for an inexpensive clock oscillator, to 0.5 to 2 ppm per year for a TCXO, to 0.05 to 0.1 ppm per year for an OCXO. The highest precision OCXOs can age less than 0.01 ppm per year.

Aging Specifications.

Aging is expressed either as a normalized frequency change per unit time after a specified time period, where the unit of time is usually a day, or as a total normalized frequency change over a specified time period. It is necessary to specify a period because aging can be highly nonlinear, especially when a unit is newly fabricated or is first turned on. Examples of aging specifications are: 1) "5 X 10^{-10} per day 7 days after turn-on" for an oven controlled crystal oscillator (OCXO), 2) "1 ppm per year" for a temperature compensated crystal oscillator (TCXO), and 3) "1 ppm to end of 20 year life" for a filter. For non-temperature controlled devices, the temperature at which the aging test is to be performed should also be specified. Guidance for specifying aging can be found in MIL-C-49468 [5] and MIL-O-55310 [1], the "mil-specs" for precision crystal units and crystal oscillators, respectively, and in a paper by Filler [6].

Aging Mechanisms.

The primary causes of crystal oscillator aging are mass transfer to or from the resonator's surfaces due to adsorption or desorption of contamination, stress relief in the mounting structure of the crystal, changes in the electrodes, package leaks, and changes in the quartz material.

Contamination Transfer Effects: Adsorption, Desorption, Oxidation and Permeation.

Because the frequency of a thickness shear crystal unit, such as an AT-cut or SC-cut unit, is inversely proportional to the thickness of the crystal plate, and because, for example, a typical 5-MHz 3rd overtone plate is on the order of 1 million atomic layers thick, the adsorption or desorption of contamination equivalent to the mass of one atomic layer of quartz changes the frequency by about 1 ppm. In general, if contamination equal in mass to 1 1/2 monolayers of quartz is adsorbed or desorbed from the surfaces, then the frequency change in parts per million is equal to the resonator's frequency in megahertz. In order to achieve low-aging, crystal units must be fabricated and hermetically sealed in ultraclean, ultrahigh vacuum environments, and into packages that are capable of maintaining the clean environment for long periods.

Fig. 3 shows some data on the oxidation of pure nickel in 76 torr of oxygen at 400 °C, 450 °C and 500 °C. Landsberg [7] reviewed and summarized a considerable amount of the type of data plotted in Fig. 3. Besides oxidation, data on adsorption and desorption of many types of gases on many types of solid surfaces were reviewed. For most of the reviewed work, the rates of adsorption, desorption and oxidation depended on the logarithm of the time. Some of the parameters of the log-time model depended strongly on the temperature. In studies of the low temperature (i.e., < 400°C) oxidation of metal films, logarithmic kinetics are usually observed [8]. A molecular model of these processes can produce the observed log-time dependence [7]. It was also pointed out that the rates of some particular systems of this type may not depend on log-time.

The adsorption and desorption of contamination is primarily a function of the nature of the contaminant, the nature of the adsorbing surface, and the temperature. Included in "the nature of the adsorbing surface" is the

crystallographic nature. For example, the adsorption properties of an AT-cut surface are likely to be different from that of an SC- or other cut.

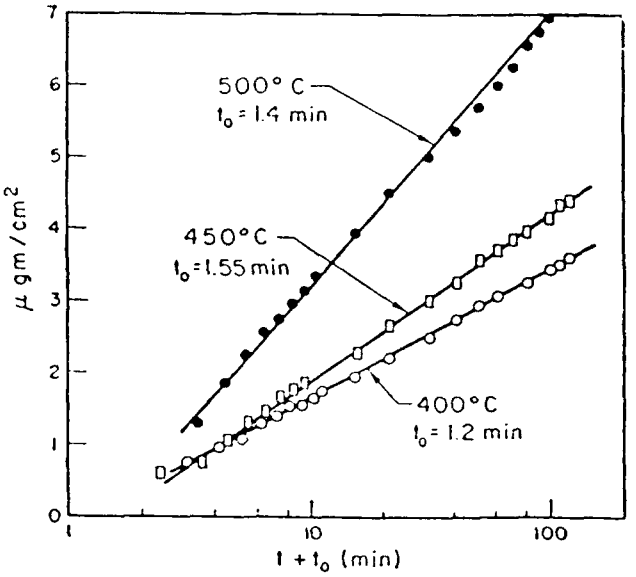


Figure 3 - The Oxidation of Nickel vs. Time.

Leaks into the enclosure, either due to a faulty hermetic seal or to permeation through the enclosure walls, and outgassing of adsorbed and dissolved gases can also result in aging. Even when the hermetic seal is perfect, gases can permeate through the enclosure walls [9,10]. Since most materials used in vacuum and resonator technology exhibit some degree of permeability to gases, maintaining low pressure levels inside the enclosure requires consideration of the permeation characteristics of the enclosure. Hydrogen and helium are notorious for their permeation through metals and glasses, respectively. Even if the permeating gases do not adsorb onto the resonator, if a significant pressure increase occurs, the frequency will change due to the changing hydrostatic pressure in the enclosure [11,12]. When a significant amount of gas leaks into a vacuum sealed resonator, aging can be caused by electrical and mechanical effects caused by changes in the package shape (sometimes called "oil canning," when the package changes with atmospheric pressure changes). Epoxy packages or seals that have been occasionally used in inexpensive resonators are especially susceptible to moisture permeation.

That high amounts of hydrogen permeation can produce aging has been demonstrated [13]. A 5-MHz fundamental mode resonator, which had copper electrodes, was hermetically sealed, in high vacuum, in a nickel HC-6 enclosure. After the aging of this resonator stabilized, it was immersed in one atmosphere of hydrogen and the aging measurements were continued. Three weeks after immersion, the aging had increased 30-fold, to about 3 X 10^{-9} per day. After the aging stabilized at this higher rate, the hydrogen was removed. Eight weeks later, the aging rate stabilized at a lower rate, but six months after removal of the hydrogen, the rate was still about five times higher than what it was prior to hydrogen exposure.

Electrode Effects.

Thin films of Au, Ag, Al and Cu are commonly used as electrodes on crystal resonators. The stresses in these films depend on the film material and thickness, and other factors. It has been shown [14] that the film stresses relax with a time constant of a few hours. (The temperature at which the relaxation was measured is not stated explicitly in this reference, but is implied to have been at room temperature. The film thicknesses were 100nm.)

The film stresses also depend strongly on the substrate material and crystallinity (amorphous, crystal surface orientation, etc.), on the substrate cleanliness, temperature, and chemical state, the gases present in the space around the substrate during the deposition process, the purity of the film material, the deposition rate, and on the deposition process (evaporation, sputtering, etc.) [15-20]. In some films, for example, the stress can be varied from tensile to compressive just by varying the background pressure in the vacuum system during deposition [21-23].

Numerous studies have shown that the properties of thin films change with time subsequent to deposition [24-33]. The stability of the film depends on the deposition conditions, especially the substrate temperature, and deposition rate (lower deposition rates under clean conditions usually result in fewer defects), and on the temperatures subsequent to deposition. For example, the resistivity changes in 700Å to 1650Å gold films deposited onto glass microscope slides were found to depend strongly on the substrate temperature [33]. When the films were deposited at 46°C, significant decreases (up to 8.5%) were observed during resistivity measurements, at room temperature, between the fifth hour and the first few weeks after deposition. When the substrate temperature was 112°C during deposition, which is near the recrystallization temperature of 124°C for gold films, the resistivity decreases were much smaller (0.8 to 1.5%). Films, in general, will tend to be more stable if the

substrate temperature during deposition is above the recrystallization temperature of the film material. Deposition at an elevated substrate temperature has a considerably greater stabilization effect than deposition without substrate heating followed by annealing at a later time. The reasons are that diffusion rates are higher at elevated temperatures, and surface diffusion is much faster than bulk diffusion, therefore, the surface atoms have more opportunity to anneal (i.e., move to low energy sites) before they become buried. Similarly, a low deposition rate (under clean conditions) results in a more stable film than a high rate. At very high rates, the surface atoms do not have a chance to anneal before they are buried. High substrate temperatures also contribute to minimizing adsorbed contaminants on the substrate and in the deposited film. If the deposition conditions are not clean, then a slow deposition rate may result in a less stable film due to incorporation of contaminants into the film.

The contribution of electrode stress relief to aging depends on resonator type. SC-cut [34] and electrodeless (e.g., BVA-type [35,36]) resonators are insensitive to such stress relief. For other types of resonators, electrode stress relief can be a contributor, especially to initial aging.

Other types of changes in the electrodes include diffusion effects and chemical reaction effects. For example, when an adhesion layer (such as Cr) is used under a weakly adhering film (such as Au), the two layers can gradually interdiffuse [37,38]. Chemical reactions can occur between the electrode and the quartz surface [16], and between the electrode and the gaseous contaminants in the resonator enclosure. Some metals react chemically with a quartz surface. When the heat of oxide formation is higher for the metal oxide than for SiO_2, the metal can reduce the SiO_2 to form the metal oxide and produce free silicon at the interface. The heat of oxide formation of α-quartz is -201.34 kcal/mol (= 8.73 eV), whereas Al_2O_3's is -399 kcal/mol and Cr_2O_3's is -270 kcal/mol. Therefore, Al and Cr, two commonly used materials in resonator fabrication, adhere strongly to quartz by forming metal oxides at the metal-quartz interfaces. The formation of Al_2O_3 and free Si at the SiO_2-Al interface has been demonstrated experimentally [39,40]. Reactions between oxide forming metals and the OH on quartz surfaces also occur.

The properties of thin films can change with time, e.g., the properties of an Al film and of the interface between Al and SiO_2 change with time [41,42]. The changes can be enhanced by temperature, temperature cycling, and the presence of an electric field. These changes can result in resonator aging.

Highly reactive metals such as Al or Cr are not the preferred electrode materials for low-aging quartz resonators because thin films of such metals produce aging due to: 1) changes at the metal-quartz interfaces, 2) gettering of oxygen and other residual gases inside the enclosure, 3) changes at high strain gradients that exist at electrode edges [43], and 4) changes in the high strains that result from temperature cycling due to the strong adhesion and thermal expansion coefficient differences between the metal film and the quartz. Liquid [44] and plasma [45] anodization have been used in attempts to minimize aging due to oxide growth on the electrodes.

Gold has been the preferred electrode material in the fabrication of low-aging resonators. The reasons for this are that: 1) gold is not highly reactive, it does not form an oxide under normal conditions (clean gold will getter organic contaminants from the air [46], however, such contaminants can be readily removed by UV/ozone cleaning [47]); 2) gold adheres weakly to quartz, the adhesion is strong enough for the electrodes not to be detached by a 36,000 g shock of 12 millisecond duration [48] but is weak enough not to support strain gradients that would show up on X-ray topographs (Cr, Al and Ni films produce strain in the quartz that can be readily seen in X-ray topographs) [43]; and 3) the stresses in pure gold films anneal rapidly [14]. These properties are probably the reason for the fact that the aging of the best long term aging AT-cut resonators with gold electrodes is about the same as that of the best SC-cut and BVA resonators, i.e., a few parts in 10^{12} per day. That the initial aging of the best SC-cut resonators is better than that of the best AT-cuts [49,50] is probably due to the superior insensitivity of the SC-cut to some important types of stresses, such as the stresses due to the electrodes.

It is also possible to make low-aging resonators with copper electrodes [51], however, in one unpublished study [52], although copper plated resonators exhibited low initial aging, the long term aging was poorer than that of similarly fabricated resonators with gold electrodes.

It is well known that, under the influences of high temperatures and high electric fields, electrode materials such as gold and silver will diffuse into the dislocations in quartz [53]. Although diffusion of gold and other electrodes into the quartz has been reported in high-temperature processed resonators [54,55], it is unlikely that such diffusion occurs at normal processing and operating temperatures without the presence of an electric

field. If such diffusion did occur, then the adhesion of, for example, gold on quartz would improve as the resonator ages. Such improved adhesion has not been observed. When five high precision glass enclosed resonators with gold electrodes, which had been fabricated 18 years earlier, were tested, it was found that the gold electrodes could be readily removed with the "Scotch-tape test." Even the weakly adhering "3M Post-it Self-Stick Removable Notes" readily removed these gold electrodes [56]. After removal of the electrodes, with either method, a narrow strip of very thin gold that outlines the perimeter of where the electrodes had been, remained (the strip can be scratched off with tweezers). This strip appears to be the "shadowing" that occurs when an evaporation mask is a finite distance from the quartz plate. The reason for the stronger adhesion of the strip is under investigation. A preliminary analysis of the quartz that had been under the center portion of the electrodes did not indicate the presence of any gold in the quartz. The analysis was performed using SIMS (secondary ion mass spectroscopy).

It has been shown that a DC voltage between the electrodes of a resonator can dramatically increase the initial aging, presumably because of electric field driven diffusion of impurities and electrodes [57]. Although oscillator designers often design circuits which (sometimes inadvertently) place a small DC voltage across the electrodes, one can readily minimize the DC voltage, without lowering the effective Q of the resonator, by placing a capacitor in series and a few megohm resistor in parallel with the resonator [57].

Surface catalytic reactions at silver electrode surfaces, accompanied by the emission of silver atoms, have also been reported to be an aging mechanism in silver-plated resonators [58].

Powerful tools to determine the possible aging contributions of stresses in films associated with particular film fabrication technologies are the form, values, and temperature dependence of the stress relaxation process. Since these parameters of the films are not too difficult to measure directly on actual resonator materials, it is surprising that very few results of this type have been reported for resonator fabrication technologies in current use. Powerful surface analytical tools for detecting diffusion and chemical reaction effects are also available.

Strain/Stress in the Resonator.

Radial forces applied at the perimeter of AT and SC plates shift the frequency [59]. Since the forces applied to the crystal wafer by the mounting clips and bonding materials are difficult to control and probably change (i.e., relax) with time, resonator aging will depend, to some extent, on the mounts' type, material, and location, on the crystal orientation, and on departures from the design in real resonators. When the bonding process is carried out at high temperatures, the structure is likely to be in equilibrium at a temperature higher than the normal operating temperatures of the resonator. In this case, mismatches in the thermal expansion coefficients of the various materials in the structure will cause stress-induced frequency shifts. When these stresses change with time, aging can result. Bonding materials, such as silver-filled epoxies and polyimides, change dimensions upon curing. This results in further stress changes. The clip forming and welding operations produce residual stresses which are also subject to stress relief.

X-ray topographs can be used to demonstrate the strains caused by a particular mounting clip structure [60,61]. If the strains change with time, aging can result. Radial and tangential thermal coefficients of linear expansion for AT-cut quartz wafers depend on direction. Thermal expansion coefficient differences between the crystal plate and the mounting structure (including the enclosure base) usually result not only in radial stresses on the crystal plate, but also tangential (i.e., torsional) stresses. Resonator structure designs must account for the differences in thermal expansion coefficients between the various parts of the resonator assembly to minimize temperature dependent stresses applied to the resonator.

Since stresses are inevitable, the materials used in the mounting structure of low-aging resonators should anneal either very rapidly, or very slowly, i.e., be either perfectly soft or perfectly elastic. Materials which anneal very rapidly are usually not a viable option because such materials generally result in unacceptable behavior under shock and vibration. Mounting structures that are nearly stress-free have been developed [35,62].

The slow and progressive deformation of a material under constant stress is called creep. Creep is observed in metals, glasses, polymers, and even in single crystals. Metals usually exhibit creep at a temperature greater than 0.4 T_m, where T_m is the metal's melting point in degrees Kelvin [63,64]. The rate of creep, especially in amorphous materials, and organic materials such as the epoxies commonly used for bonding crystals, is highly sensitive to temperature. Numerous alloys have been developed for high resistance to stress relief, for electrical connector and other spring applications. Mounting clips

made of such alloys can serve to minimize aging due to stress relief in the mounting structure, as can designs that use carefully oriented quartz for mounting [35], and for the enclosure [65].

The amount of aging produced by a given amount of stress relief is a function of the orientation of the mounting clips with respect to the crystallographic axes of the quartz plate, and the types of stresses [59]. For in-plane diametric forces, the force-frequency coefficient K_f vs. azimuth angle ψ has been found to have zeroes for all the commonly used cuts, such as the AT- and SC-cuts [59]. Therefore, one might conclude that aging due to stress relief in the mounting clips can be eliminated by mounting the crystals where $K_f = 0$. Unfortunately, it is difficult to completely eliminate aging due to stress relief in the mounting structure because: 1) the azimuthal angles where $K_f = 0$ are functions of temperature [66], so that the mounting point orientations would have to be different for resonators of different turnover temperatures, 2) the ψ where the effects of bonding stresses are zero is different from the ψ where the $K_f = 0$, at least for the AT-cut, the only cut for which bonding stress effects have been reported [67], and 3) the forces due to mounting clips are generally not purely in-plane diametric forces. This is especially true for three and four-point mounted resonators because, since the thermal expansion coefficient of quartz is highly anisotropic whereas that of the typical package base is isotropic, the forces due to thermal expansion coefficient mismatches will have tangential components. In two-point mounted resonators also, the base's thermal expansion applies torsional-type forces in addition to the in-plane diametric forces.

Even when the mounting stresses are made negligible, the bonding stresses alone can cause significant frequency shifts, which, upon annealing, can cause aging [67]. The temperature coefficient of frequency will also change with large changes in stress. In these cases, measuring temperature coefficients along with the aging can be used to determine whether or not stress relief is a significant aging mechanism.

Diffusion Effects.

Although it is likely that diffusion processes cause aging in resonators, very few authors have reported analyses of aging in terms of diffusion [4,68]. Solid-solid diffusion processes can occur between the electrodes and the quartz, within the quartz itself, and in the mounting and electrical attachments; gas-solid diffusion processes can be rate determining steps for some processes occurring at the resonator surface. Another conceivable aging mechanism is the diffusion of impurities in the quartz to dislocations and surfaces.

The rate of a simple diffusion-controlled process in an isotropic medium is given by [69]

$$\text{Rate} = -D * \text{concentration gradient}, \qquad (1)$$

where D, the diffusion constant, is

$$D = \frac{mh}{A(d_2 - d_1)t},$$

m is the mass of the diffusing material, h is the thickness of the slab in which the diffusion is occurring, A is the area participating in the diffusion, and d_2 and d_1 are the concentrations (mass per unit volume) of the diffusing material on the two sides of the slab [69,70].

Table I shows diffusion constants at room temperature for some metals commonly used as electrodes or attachments in resonators [70]:

Table I

Metal System	D (cm^2/sec)
Al into Cu	1.75×10^{-2}
Au into Cu	4 to 16×10^{-2}
Cu into Ag	5.95×10^{-5}
Ni into Cu	6.5×10^{-5}
Pd into Cu	1.6×10^{-4}

This short table shows that diffusion rates of potential interest to crystal suppliers vary by orders of magnitude depending on the metals involved in the diffusion.

The movements of point defects in the quartz lattice have energies and activation energies of 0.03 eV to 0.154 eV [71]. Although these activation energies are similar in magnitude to observed aging activation energies [72], the validity of a direct comparison has not yet been established.

Diffusion processes often have an approximately power dependence on time, with the power being about 0.5 [72]. For some diffusion configurations, the diffusion rate equation also contains an exponential time factor [72].

Diffusion rates are usually thermally activated, with an Arrhenius dependence on temperature [10,73]. The

activation energy for Cu into Al (33.9 kcal/mole) seems higher than most reported aging activation energies [72]. However, the activation energy for grain boundary diffusion of Ag into Ag is 21.5 kcal/mole, which is not much higher than the reported aging activation energies [72,73].

Changes in the Quartz.

Changes in the quartz due to stresses or other causes could lead to aging, although no reports of such changes (at normal temperatures and pressures) could be found in the literature. Perfect quartz would not be expected to change with time (by definition). The imperfections that are subject to change include surface and point defects, dislocations, impurities, inclusions, and twins. Surface defects, such as the microcracks produced by lapping, can change with time, however, by properly etching the surfaces subsequent to mechanical treatment [74], the possibility of changes can be greatly reduced.

It is unlikely that dislocation motion due to stresses is a factor in aging at typical operating temperatures, although dislocation motion can occur at high temperatures and pressures [75]. Even in sweeping experiments [76] which are usually conducted far above the normal operating temperatures of oscillators, no evidence of dislocation motion has been reported. The energy needed to anneal quartz damage due to neutron irradiation may be a clue to the energies needed to move dislocations. When quartz is irradiated with fast neutrons, displacement damage occurs. At high doses, the quartz gradually becomes disordered into an amorphous form. Annealing studies on neutron-damaged quartz indicate that the annealing temperature of quartz is above the inversion temperature [77,78]. The activation energy for structure annealing is 0.75 eV.

The outgassing of quartz is another possible aging mechanism, the magnitude of which is unknown. Although a large amount of information is available on the outgassing characteristics of vitreous SiO_2 [9,10], no reports on the outgassing characteristics of α-quartz were found. Presumably, since all materials outgas to some extent, α-quartz does too. Another unknown is the extent to which impurities in the quartz diffuse to dislocations and surfaces at normal operating temperatures. Quartz is known to contain impurities that are subject to diffusion [79]. The impurity that is present in highest concentrations is hydrogen. Natural and cultured quartz both contain hydrogen, in amounts ranging from 200 ppm to 2500 ppm relative to Si (3 ppm to 42 ppm by weight).

Circuit Aging and Other Electrical Aging.

The oscillator circuit includes electrical elements that are subject to change. These elements determine some important operating factors, such as resonator load reactance, dc bias, and drive level. Changes in these electrical elements and factors cause oscillator aging. Stray reactance changes due to movements of the electrical leads and the gradual deformation of circuit boards and enclosure walls also produce frequency changes. For example, in a 22-MHz fundamental mode AT-cut resonator, a lead length change of 0.005 mm results in a frequency change of 1×10^{-9} [6], and a load capacitance aging of 1 ppm per day causes oscillator aging of parts in 10^{10} per day. Capacitors can age due to the effects of humidity and temperature; e.g., humidity can change the dielectric constants and loss factors of capacitors and circuit boards. Inductors are notorious for their instabilities; e.g., the windings of inductors can stretch and move, especially during temperature changes. In one study [80], capacitors of various types aged 3 ppm/day to 70 ppm/day, and a 2 μH coil custom wound on a phenolic form (for high stability) aged 2 ppm/day. Amplifiers and varactors can also change with time. Drive level changes can result in frequency changes on the order of 10^{-7} per ma^2 [81]. Significant improvements in medium and long term stability can often be obtained by hermetically sealing the oscillator to minimize the frequency instabilities due to changes in humidity and atmospheric pressure [82,83].

As was previously noted in the "Electrode Effects" section, a DC voltage between the electrodes of a resonator can dramatically increase the initial aging of a resonator, presumably because of electric field driven diffusion of impurities and electrodes. Oscillator designers often design circuits which (sometimes inadvertently) place a small DC voltage across the electrodes. When high stability TCXOs from several manufacturers were examined, DC voltages ranging from a fraction of a volt to about 4 volts were found. The DC voltage effect can be easily minimized, without a significant lowering of the resonator's effective Q, by placing a capacitor in series and a few megohm resistor in parallel with the resonator [57].

DC voltages can also result from static charges generated during fabrication. Static charges can result in high (> 1 kV) voltages. Since the surface resistivity of clean quartz is high, the "static" charges can take a long time to decay. This can lead to (initial) aging, especially in doubly rotated crystals, such as SC-cuts, which can have

sensitivities of parts in 10^{-9} per volt. Well-known static ("ESD") control measures can be used to minimize this problem.

In OCXOs, aging of the temperature control circuitry can change the set point of the oven, resulting in aging. The amount of aging for a given set point change depends on the frequency vs. temperature characteristic of the resonator, and will, in general, be much smaller for a typical SC-cut OCXO than for an AT-cut one [84].

Temperature and Temperature-Cycling Dependence of Aging.

Although most of the known aging mechanisms are thermally activated, the expected strong temperature dependence of aging rates has not been observed in many medium stability and high stability resonators. For example, in one study of eight AT-cut resonator types, from several manufacturers, the resonators were aged at room temperature, 85°C and 120°C. No systematic aging rate variations with temperature were found [85]. The resonators did not always exhibit a higher aging rate at elevated temperatures; some even aged at a lower rate at the elevated temperatures. The resonators ranged from 14.4-MHz fundamental mode to 70-MHz third overtone.

In another study, groups of resonators were aged at 50°C, 60°C and 70°C. It was found that, although the aging improved for many resonators when the temperature was lower, the low-aging rate resonators did not change their aging characteristics notably when the temperature was lowered [86].

When groups of low and medium stability resonators were aged at various temperatures ranging from 25°C to 100°C, the aging rates were found to increase with increasing temperatures [87]. The increases varied with crystal type. A conclusion the authors draw concerning aging prediction is "that such prognosis will be sufficiently reliable only when the crystal unit will operate under more or less identical conditions during all its service life. If the ambient temperature or drive level changes, even the sign of frequency aging may change." In some earlier studies on precision 2.5-MHz 5th overtone AT-cut resonators [88], and 550-kHz wire mounted DT-cut resonators [89], an increase in aging temperature increased the aging rates.

The aging of groups of six high precision SC-cut resonators and two high precision AT-cut resonators were measured successively at the lower turnover point (LTP) and upper turnover point (UTP) temperatures. The UTP to LTP differences ranged from 19°C to 40°C for the SC-cuts, and it was about 100°C for the AT-cuts. The higher aging temperatures at the UTP introduced a small positive aging contribution to the SC-cuts' aging, i.e., the aging of both positive and negative aging resonators became more positive at the UTP. The results for the AT-cuts indicated no drastic improvements in aging rates at the LTP [57]. Aging at the SC-cuts' UTPs, which ranged up to 120°C, did not cause a significant aging rate degradation [57,90].

If aging were due to a single thermally activated process, then one ought to be able to observe phenomenally low-aging rates by cooling the resonators to cryogenic temperatures. Although the definitive aging experiment at low temperatures is yet to be performed, the limited data at such temperatures indicates no drastic improvements in aging rates [91-93]. When the aging of high precision SC-cut resonators was measured at (the lower turnover) temperatures in the vicinity of -10°C and compared with the aging of similarly fabricated resonators aged at turnover temperatures that were 90°C to 120°C higher, no significant improvement in aging was observed at the lower temperatures [57,94].

When an aging interruption is accompanied by a significant temperature excursion, the effects can range from drastic to small. For an example of drastic change, when the aging of an oscillator containing a high precision 5-MHz 5th overtone glass enclosed AT-cut resonator was interrupted and the oscillator was cooled to -40°C for nine days, the aging rate increased drastically subsequent to the resumption of aging. Months elapsed before the aging rate returned to its value prior to the interruption. Fig. 4 shows this example. This oscillator was aging at an approximately constant rate of -1.1×10^{-11} per day prior to interruption. After the aging resumed: 1) the aging rate became positive at first, 2) the rate reversed direction after about 150 days, and 3) the aging rate stabilized at -4.0×10^{-12} per day after about one year and it stayed at that rate for the subsequent four years.

Low temperature storage can produce drastic aging rate changes (and frequency offsets) in SC-cut oscillators, too [57,90]. In another aging experiment, groups of high stability OCXOs and TCXOs were on-off cycled and temperature cycled during the experiment. For most of the oscillators, the interruptions did not worsen the total frequency changes during the aging period [94]. Similarly, when high stability microcomputer compensated crystal oscillators were aged with repeated interruptions for temperature cycling, the aging rates were not made

worse by the cyclings [95].

It has been shown that stability under intermittent operation is dependent on the processes used in fabricating the resonator. For example, upon restarting an oscillator following an oven shutdown, those containing high temperature processed crystal units exhibit much smaller frequency offsets, and the aging rates return to their pre-shutdown values much faster, than do low temperature processed units [89,96].

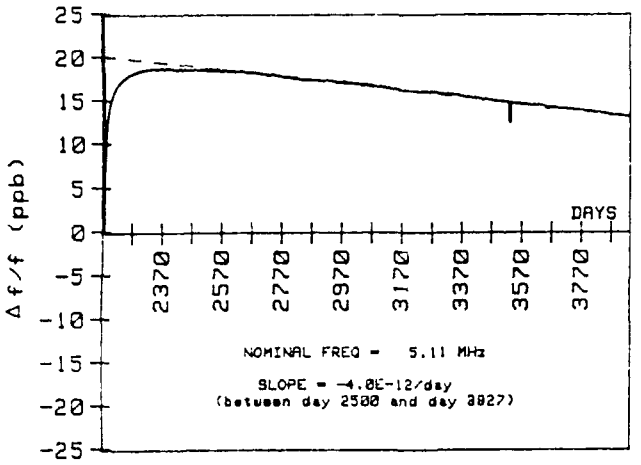


Figure 4 - The Effect of Aging Interruption.

Frequency and Overtone Dependence of Aging.

Since most of the known aging mechanisms are associated with the resonator's surfaces, it is not surprising that, in general, resonators of higher volume-to-surface ratio tend to exhibit a lower aging rate than resonators of lower volume-to-surface ratio. For a given fabrication process, the aging rate tends to scale with the volume-to-surface ratio of the resonator's active area, i.e., with the frequency of the plate. For example, the aging of 333-MHz fundamental mode SC-cut resonators [97] at 25°C (the test OCXOs were kept in a freezer to permit stabilization of the OCXO ovens at 25°C) was found to range from 1 to 3 x 10^{-8} per day after one year of continuous aging [98]. The aging of comparably fabricated 5-MHz fundamental mode resonators under the same aging conditions would typically be on the order of 333/5 times lower, i.e., on the order of a few parts in 10^{10} per day. The scaling with frequency appears to apply to SAW devices too; e.g., the aging rates of "good" 500-MHz SAW resonators [99] are typically on the order of 100 times higher than the aging rates of "good" 5-MHz bulk acoustic wave resonators.

The aging rates of 2.5-MHz and 5-MHz 5th overtone AT-cut resonators made according to designs and processes developed by Warner and coworkers in the late-1950's and 1960's [88,89,96,100-104] are still difficult to surpass. This is probably due, in part, to the large volume-to-surface ratio of these resonators. No larger volume-to-surface ratio resonators have been developed since. (In fact, 2.5-MHz 5th overtone resonators are no longer made regularly.)

In general, overtone resonator aging rates are lower than aging rates of comparably fabricated fundamental resonators of the same frequency. The reasons are, not only that overtone resonators have a larger volume-to-surface ratio, but also, that overtone resonators, having a smaller C_1, exhibit a lower aging rate due to oscillator circuit component aging and a lower sensitivity to changes at the resonators' edges.

Drive Level Dependence of Aging.

The experimental evidence concerning the effect of drive level is mixed. In one report on the aging of (low stability) 32.8-kHz flexural mode resonators [87], an increase to "inadmissibly high values of crystal unit drive level" from 10 μW to 100 μW, produced a significant increase in aging rates. (An increase in aging temperature was also found to increase the aging rates of these resonators.) The report, however, references an earlier study which showed "the absence of any amount of significant influence of drive conditions on the crystal unit aging...," at reasonable drive levels.

The aging rate of AT-cut resonators has been reported to be degraded by high drive levels [105]. Increasing the current through 2.5- and 5-MHz 5th overtone resonators ten fold from 75 μa resulted in an increase in monthly aging rate from 1 X 10^{-10} to 1.5 X 10^{-9}. The aging of BVA resonators has also been reported to be sensitive to drive level [106,107]. The drive level sensitivity has been used to produce oscillators that exhibit "zero aging" at a particular time. (It is highly unlikely, however, that such a balancing of aging mechanisms can last for long periods.)

Changing the resonator drive levels in discrete steps did not affect the aging rates of precision, high temperature processed SC-cut resonators, up to a 2.5-ma drive current (594 μW), the maximum that was tried in the study [94].

Since resonator frequency is a function of drive level [81], if the oscillator circuitry ages so as to gradually change the drive level, oscillator aging can result. At very high drive levels, the power dissipation in the resonator will

raise the temperature of the resonator's active area. Therefore, resonators which are adversely affected by increased aging temperatures would also be expected to be adversely affected by increased drive levels. From the limited data available, it appears that, at a constant drive level, the aging of low stability and AT-cut resonators is adversely affected by high drive levels, whereas that of precision SC-cuts is not affected. (It may be worthwhile to repeat the measurement of the drive level dependence of precision AT-cuts, in carefully designed oscillators, to eliminate the possibility that the drive level dependence of aging was circuit induced, e.g., via higher DC bias on the resonator at higher drive levels.)

Increasing the drive level increases the displacements, velocities and accelerations of particles at resonator surfaces. At high frequencies, especially, particle accelerations can be on the order of a million g's. High drive levels' ability to remove particulate contamination from surfaces is well known. Whether or not high drive levels can affect the desorption of adsorbed contaminants has been considered [108]. Since the increased kinetic energy of an adsorbed molecule due to a high drive level is very much less than 20 kcal/mol (which is the typical adsorption energy of concern), and is also much less than the thermal energies at normal operating temperatures, high drive levels probably do not directly affect adsorption-desorption phenomena.

The Effects of Aging Interruptions.

If during an interruption in aging the resonator or oscillator experiences a change in its environment, then upon resuming the aging, the aging rate can be significantly different from what it was prior to the interruption. The effects of thermal interruptions, drive level changes, and DC fields were discussed earlier. Other interruptions that can result in increased aging include mechanical and thermal shock, vibration, magnetic and electric fields, AC signals, and radiation [109,110]. Few studies have been reported on the effects of interruptions other than thermal. In general, the effects will depend on the nature of the interruption, and on the aging mechanisms that are disturbed by the interruption. For example, if during a shock some elastic limit in the resonator's mounting structure is exceeded, then the shock may change the stress-relief component of aging. Similarly, if a radiation pulse displaces impurities to higher energy sites in the lattice, then the subsequent aging may be affected by the migration of impurities to low energy sites.

Dependence of Aging on Material and Mode Type.

Since there are very few reports on common applications that use different materials or modes, the effects of material and mode type on aging are not easily separable. No reports on the effects of changing material for the same mode type were found in the literature. No reports on aging for bulk wave devices made from lithium niobate, berlinite, or other new piezoelectric materials were found. In this section, aging data on extensional mode lead zirconate titanate and lithium tantalate resonators, other modes using lead zirconate titanate, hydrophones made from poly (vinylidene fluoride), quartz flexural tuning forks, and quartz surface wave resonators are presented. Where possible, the material effects are separated from the device type effects.

Dependence of Aging on Resonator Material.

Extensional resonators made from lead zirconate titanate and similar materials age from 700 to 10000 ppm per decade in time [111]. Improved lead zirconate titanate type materials operating in all modes age about 1000 ppm to end-of-life of 10 years [112].

For piezoelectric ceramic materials the aging usually proceeds as log-time [113] and is therefore reported as a frequency change per decade of time. No recent reports with a detailed description of the time and temperature dependence of aging in materials like lead zirconate titanate could be found.

Extensional mode lithium tantalate resonators, with frequencies from 455-kHz to 2-MHz, were aged at room temperature for more than 900 days. Aging did not depend significantly on frequency. The frequency decreased about 100 ppm during the first year; after two years, the aging rate was 3×10^{-7} per month [114].

Hydrophones made from a plastic piezoelectric polymer, poly (vinylidene fluoride) age a few 10000 ppm per decade in time, with an upper temperature limitation of 80°C [115]. For the hydrophone aging, the mode of vibration used was not stated, but it was implied that this aging value was determined for the 31 and 33 modes.

Most reported aging studies have been for quartz and lead zirconate titanate devices.

Dependence of Aging on Mode Type.

Some reports on aging of low frequency and high

frequency quartz bulk wave devices, as well as the aging of surface acoustic wave (SAW) resonators and delay lines, are summarized by Gerber [4].

The aging of low frequency wire mounted quartz crystal resonators operating in different contour (length-width) modes was reported by Armstrong, et al. [89], and by Gerber and Sykes [116]. For these devices the aging proceeded as log-time.

Kanbayashi [117], Engdahl and Matthey [118], Yoda [119], and Forrer [120] have reported aging results for quartz flexural mode tuning forks at frequencies of about 32.768-kHz. Yoda reported aging results for 32.768-kHz flexural (XY, NT, tuning fork) mode resonators to be ± 5 ppm per year. Some examples of -0.2 to +0.3 ppm per year were shown [121]. Table II summarizes the aging of several types of low frequency quartz resonators.

The data in Table II suggest that flexural and width-shear resonators age less than extensional or face shear resonators. The data in Table II also suggest that lower frequency resonators age less than higher frequency resonators of the same type. From the data in Table II, it is not possible to assign the cause of the low-aging flexures and width-shears to mode type or frequency alone.

Gerber [4] also summarized some reports on the aging of quartz surface acoustic wave resonators and delay lines. For some 184- and 194-MHz SAW resonators there was little or no difference in aging at 50°C to 150°C, but increased aging was observed at 200°C to 250°C [122,123]. For some 160-MHz SAW resonators powered in an oscillator at 60°C the log-time aging rates approached 1×10^{-9} per day; -5 to $+5 \times 10^{-9}$ per day after 162 days. According to the general rule that the aging is proportional to the frequency, these SAW resonators had an aging equivalent to a bulk wave resonator of about 6×10^{-10} per day at 10-MHz. For these devices the best long term aging was -0.64 to -0.31 ppm per year [124-126]. For 160-MHz SAW resonators, aging data were best fitted with two simultaneous log-time aging mechanisms [125,126].

Some 300-MHz SAW resonators had aging rates of 1 to 2 ppm per 30 days at room temperature [127]. Some 1.4-GHz SAW delay lines aged several ppm in 52 weeks,

Table II
Aging of Low Frequency Quartz Crystal Resonator
(Probably at room temperature)

Mode Type (frequency)	Aging (ppm)	Time of Aging
Extensional (200-kHz)	8.0	100 days
Face Shear (200-kHz)	4.0	100 days
Extensional (100-kHz)	3.0	100 days
Width-Shear (990-kHz)	3.0	100 days
Flexure (8-kHz)	1.2	100 days
Flexure (445-kHz to 2-MHz)	-0.2 to +0.3	365 days
Width-Shear (550-kHz)	1.0	100 days
Width-Shear (230-kHz)	0.5	100 days
Flexural Tuning Fork (32.768-kHz)	-0.5 to +2.0	365 days

also probably at room temperature [128]; 200- to 400-MHz SAW delay lines in powered oscillators at about 30°Caged between -10 to +17 ppm in about 60 to 120 weeks [129]; 400-MHz SAW devices aged less than 1 ppm per year (-2 to +4 ppm per year) on a production basis [130,131]; 187-MHz to 425-MHz SAW resonator oscillators operating at room temperature aged between 0.1 and 0.5 ppm per year with aging times from 27 weeks to 80 weeks [65]; 425-MHz SAW resonators oscillators aged less than 0.1 ppm in 100 weeks at 60°C [99].

A survey of aging for SAW devices showed best aging rates of less than 0.1 ppm per year [131]. SAW aging continues to be reduced as the SAW fabrication technology adopts bulk wave processes of cleanliness, high temperature processing, and careful attention to the selection of materials, mountings, and packages.

Low-aging Oscillators: The State of the Art, Present and Future.

The lowest aging rates reported to date are a few parts in 10^{13} per day [90,105]. The date of reference number 105 is 1967. Unfortunately, the authors cite unpublished results obtained elsewhere. The aging rate of the parts in 10^{13} per day oscillator of reference number 90 had increased to 4 X 10^{-12} per day one year after publication of the reference, and remained at that rate for the next year (and is continuing at that rate as of mid-1991)[98].

Several authors have reported few parts in 10^{12} per day aging rates [36,88,90,94,96,100,132,133]. Such aging has been reported for only a few resonators, usually after extended stabilization periods. No manufacturer will guarantee such aging today even though the first report of parts in 10^{12} aging (1 X 10^{-10} per month) appeared in 1958 [100].

The slow progress in the best long term aging performance is puzzling because during the past 30 years, numerous advancements have taken place which ought to have yielded improvements. Among the advancements are: the SC-cut, better ultrahigh vacuum systems, better cleaning techniques, better understanding of stress effects, and better oscillator circuitry. These advancements seem to have resulted in significant improvements in initial aging, but not in the long term aging. As manufacturers adopted these advancements, the advancements have also resulted in improvements in the aging of resonators in high volume production.

In 1983-84, an informal survey of worldwide capabilities in making low-aging resonators was performed [57]. Ten organizations were identified which could make resonators with parts in 10^{11} per day aging after 30 days of continuous operation. Although the processes used to make these resonators varied widely, the end results with respect to long term aging did not (the stabilization periods, however, were much shorter for some than for others).

Parts in 10^{11} aging has been achieved with AT-, BT- and SC-cut resonators; with glass enclosed, metal enclosed, and ceramic flatpack enclosed resonators; with natural, cultured and swept cultured quartz; with lightly etched and deeply etched resonator plates; with mechanically polished and chemically polished plates; etc. A high temperature vacuum bake prior to sealing the resonators appeared to be the key step that was common to all the processes that produced low-aging. Although the bake temperatures, times and vacuum conditions varied widely, it is clear from the survey and other evidence, that vacuum baking before sealing is a necessary step in the production of low-aging resonators.

Aside from high temperature processing, the only parameter that was, in some respects, common to all the processes was the quartz material. No significant improvements in reducing quartz defects, such as dislocations and hydrogen content, have taken place during the past 30 years. In fact, the dislocation densities in cultured quartz have probably increased over the years. Since the quality of commercially available cultured quartz is adequate for nearly all high-volume applications, and since it is expensive for growers to "refresh" the seeds periodically by using natural quartz, the growers have had little incentive for refreshing seeds. If the seeds are not refreshed, cultured quartz grown on successive generations of cultured quartz seeds will contain increasing densities of dislocations. Although no data could be found on the defect densities in low-aging resonators from 30 years ago vs. today, it is conceivable that the aging of the lowest aging resonators is somehow related to quartz defects. The outgassing of quartz may also limit the lowest attainable aging.

Another parameter that is common to all resonators is background ionizing radiation due to cosmic rays and radioactive trace elements in the soil and building materials. The amount of background radiation depends on location. The average annual radiation dose from natural sources in the U.S.A. has been reported to be on the order of 0.1 rad [134].

The radiation sensitivities of resonators are a highly nonlinear function of dose at low doses. The low dose radiation effects are not well understood. The frequency changes per rad at low doses can be several orders of magnitude higher than at high doses [135]; sensitivity as high as 1×10^{-9} per rad has been reported upon initial irradiations [110]. A resonator with such sensitivity would exhibit apparent aging (i.e., drift, according to the new definitions [2]) of 1×10^{-10} per year or 3×10^{-13} per day due to the background radiation. Although other resonators have been found to have smaller sensitivities [109], and the radiation was deposited much faster in the experiments than the rate of deposition of background radiation, since the low dose effects are not well understood, and the rate dependencies at very low rates are unknown, it is conceivable that the effects of background radiation are not negligible in the lowest aging resonators. A better understanding of low dose effects (and the means to minimize these effects) may be a prerequisite to making substantial improvements in the technology of ultra low-aging resonators.

To achieve the lowest aging rates, the oscillator circuitry must also be carefully designed. Achieving lower than parts in 10^{12} per day oscillator stabilities is no easy task, however; the best circuits do not yet seem to limit the achievable long term aging of small C_1 resonators, such as 5-MHz 5th overtone SC-cuts. (Some of the lowest aging rates reported have been for 10-MHz 3rd overtone SC-cut oscillators [57,94].)

In spite of the lack of significant progress in improving the aging of the lowest aging oscillators, no evidence exists to indicate that the barriers to further improvements are insurmountable. The definitive experiments, in which all known aging mechanisms are minimized, are yet to be done.

Aging Acceleration Effects.

Multiple Aging Mechanism Pitfall.

The temperature dependence of aging rate depends on the processes used during a resonator's fabrication. Since the aging rates due to most known aging mechanisms have strong (e.g., exponential) dependencies on temperature, care must be taken in the use and interpretation of accelerated aging tests. Especially when two or more aging mechanisms are present, simple accelerated aging tests can lead to misleading or meaningless results [57,136].

For example, adsorption and desorption of contamination is believed to be a significant aging mechanism in many resonators. The half-life, $t_{1/2}$, of adsorbed molecules can be expressed as

$$t_{1/2} = t_0 e^{E_d/RT} \qquad (2)$$

where R is the gas constant, t_0 is about 10^{-13} seconds, E_d is the desorption energy, and T is the temperature in Kelvins [9]. Therefore, the half-life of a molecule at room temperature (25°C = 398°K), is about 1/2 minute when E_d = 20 kcal/mol; it is 30 years when E_d = 30 kcal/mol, and it is a billion years when E_d = 40 kcal/mol. Molecules that have E_d = 20 kcal/mol are desorbed relatively rapidly and are pumped away during processing in vacuum. A monolayer of molecules (with molecular weight comparable to that of quartz) with E_d = 30 kcal/mol contributes on the order of 10^{-9} per day to the aging of, e.g., 20-MHz fundamental mode AT-cut resonators. A monolayer of molecules with E_d = 40 kcal/mol contributes on the order of 10^{-17} per day, and can, therefore, be considered to be stable with time. Therefore, only a few molecular species are likely to contribute to aging (but these include some important ones, such as H_2O, CO_2, CO and CH_4).

In Eq. 2, E_d is sometimes found to increase with decreasing coverage, and after a monolayer is formed, it is generally also different for a second adsorbed layer. Although Eq. 2 gives the half-life of the desorption process, the details of the time dependence of desorption are not specifically included. This detail is supplied by an analysis such as the ones by Landsberg [7] and Glang, et al.[9].

If, for instance, a resonator's aging is determined by the desorption of two species of contaminants, one with E_d = 30 kcal/mol and the other with E_d = 40 kcal/mol, then accelerated aging at, for example, 150°C will provide meaningless results. At 150°C, the molecules with E_d = 30 kcal/mol, which are the molecules that determine the aging at room temperature, are desorbed within a fraction of a second. The subsequent aging is then determined by the E_d = 40 kcal/mol molecules, which at room temperature do not contribute measurably to the aging. The 150°C aging results, therefore, will not shed any light on the aging that can be expected at room temperature. Data at several temperatures is necessary to reveal the presence of multiple aging mechanisms and their effect on the aging at the normal operating temperatures.

Brief Review of the Theory of Rate Processes.

The following tutorial discussion of rate processes [137] is included in this review paper because understanding rate processes can be important in our attempts to understand aging and accelerated aging results.

The various aging mechanisms are specific cases of rate processes. Many studies of chemical rate processes have been reported and are summarized in textbooks on physical chemistry and chemical rate processes. Early studies of chemical reaction rates were on gas and solution systems.

For gas reactions, the reaction rate is often controlled by collisions between reacting gas atoms and molecules. Arrhenius proposed that the reaction rate could be expressed as:

$$Reaction\ Rate = K\, e^{-E_a/RT} \qquad (3)$$

where E_a was called the activation energy because it described the dependence of rate on the absolute temperature in degrees Kelvin (T). T is the resonator temperature, sometimes called the soak temperature. R was a constant that made the exponent argument unitless, as required. If R = 1 then E_a is in degrees Kelvin; R can be selected so that E_a is in electron volts (a common unit often used in studies of the aging of semiconductor devices), or calories or kilocalories per mole (which is a common unit used in chemical reaction rate studies). K includes factors specific to the system of interest, such as gas reactant types and pressures or dissolved reactant types and concentrations.

Boltzmann later developed a statistical thermodynamic theory of gas reactions that showed that E_a in the Arrhenius proposal ought to depend on temperature because it had a part associated with the distribution of system energy (entropy) as well as the part associated with changes in the average system energy (thermal energy). According to this theory E_a should really be F_a, the change in the free energy of the reaction process (rather than the change in the enthalpy). This more rigorous theory helped to make sense of a larger set of experimental results, including those of a few reactions which were actually slower at higher temperatures. These particular reactions had rates controlled by the distribution of thermal energy rather than by the average change in the thermal energy. Further studies of gas phase reaction rates have included quantum effects for simple systems.

Experimental studies have shown that reactions in solutions are more complicated than gas phase reactions. Experimental studies have also shown that reactions involving the interaction of a gas or dissolved reactant with a surface are generally more complicated than either gas or solution phase reactions. In spite of the additional complexity, the gas phase ideas of reaction rates controlled by an Arrhenius or Boltzmann rate law and by collisions between reacting atoms, molecules, free radicals, and surfaces, have been very useful in developing an understanding of general reaction rates.

This brief history of reaction rate science suggests that oversimplification can easily lead to an inadequate understanding of the processes being considered. For example, it was found experimentally that many gas and solution phase reactions have Arrhenius activation energies that produce a doubling of the rate for every ten degrees Celsius increase in reaction temperature. If no other evidence is available, this activation energy can be used to design experiments to efficiently estimate the actual activation energy for the system of interest. Credible long term estimates of crystal resonator and crystal oscillator aging can only be determined from these experimentally determined activation energies and not by the general rule of thumb.

For many gas and solution systems, the dependence of the concentration of a reactant or resultant species on time is exponential, giving a reaction rate that can approach zero, as a reacting species is completely used up in the reaction. In systems for which one of the reacting species has a very large excess concentration, so that the concentration does not change very much during the entire reaction process, changing the concentration of that species may not measurably change the reaction rate.

A Log-time Law of Chemisorption, Oxidation, and Stress Relief.

A review of chemisorption on metal surfaces and oxidation of metal surfaces [7], two processes which could easily affect the dependence of the frequency of a crystal resonator on time, showed how most of the available experimental data were consistent with a logarithmic rate law. The source of such a law was also proposed. In spite of some uncertainties, most of the available metal chemisorption and oxidation experimental data could be described by:

$$q = \left(\frac{1}{b}\right)\log\left[\frac{t+t_0}{t_0}\right] \qquad (4)$$

where q is the amount of adsorbed gas, t is the time, and b and t_0 are constants. In Eq. 4 t_0 is not the starting time for the adsorption, but is one of two parameters that characterize the aging process (t_0 will be used as the aging process starting time in a later section of this paper).

At least that part of the aging of crystal devices associated with chemisorption of contaminant gases on the crystal surfaces, or on the electrodes, or with the oxidation of the electrode material, or the quartz itself, might have a logarithmic rate law. Some quartz resonator designers have recognized that many resonators age a fixed amount per decade in time, which is a logarithmic rate dependence [4,6,68,71,89,96,104,138-142].

The atomistic/molecular model of the chemisorption process with a logarithmic rate law was expressed in terms of the specialized language of adsorption chemistry, but can be stated in much simpler and more general terms. The simpler and more general statements are that: 1) the adsorption does not depend on any concentrations (zero order), and 2) that the Arrhenius activation energy is a linear function of the amount of adsorption. Both of these statements can be expressed as:

$$\frac{dq}{dt} = C\, e^{-(E_{a0} + \langle E_{a1} q \rangle)/kT} \qquad (5)$$

where q is the amount of adsorbed material (proportional to the frequency shift in a resonator due to adsorbed material), t is the time, T is the soak temperature in degrees Kelvin, C does not depend on time or temperature, E_{a0} and E_{a1} are the constant and linear components of the activation energy, and k determines the units of E_{a0} and E_{a1}. For example, if E_{a0} is expressed in degrees Kelvin and E_{a1} is expressed in degrees Kelvin per second, then k = 1 and t is in seconds. Only a few studies on the use of Eq. 4 in the analysis of aging of resonators using quartz or other piezoelectric materials have been reported [4,6,68,71,89,96,104,138-142].

Before examining several implications of Eq. 5, we should note that thermal stress relaxation in metals such as zinc [143] has also been reported to follow a logarithmic rate law [7], as does the oxidation of many metals [8]. Consequently, even if the logarithmic rate law were established as valid for crystal aging, other evidence would be needed to separate the effects of the chemisorption, oxidation, and thermal stress relaxation parts of the aging in the resonator [72]. The authors do not know about any other proposed mechanism for the aging of crystal resonators for which a basis as good as that of the logarithmic time mechanism has been developed.

Implication of the Log-time Aging Law.

An important question is: under what conditions can Eq. 5 be used to understand and control the aging of resonators? A careful examination of the consequences of Eq. 5 is one place to start developing an answer to this question.

Integrating Eq. 5 from t_0 to t and $q(t_0)$ to q(t) gives:

$$q(t) = q(t_0) + \frac{KT}{E_{a1}} \ln[1 + \langle B(t-t_0) \rangle]$$

$$\text{where } B = \frac{E_{a1} C}{kT\, e^{\left[\frac{E_{a0} + [E_{a1} q(t_0)]}{kT}\right]}} \qquad (6)$$

Eq. 6 describes the dependence of q on time and soak temperature. In Eq. 6, t_0 is the starting time for the process, $q(t_0)$ is the amount of adsorbed material at $t=t_0$, and the other quantities were defined earlier.

The t_0 defined here is not the same quantity as the t_0 defined in Eq. 4; t_0 in Eq. 4 is equal to (1/B) in Eq. 6. In Eq. 4 the starting t and starting q were both 0. In Eq. 6, $q(t_0)$ may also depend on temperature. This reversible temperature dependence is not time dependent and is the temperature coefficient of q at $t=t_0$.1

Eq. 6 can be expressed in the following simpler form by defining combinations of variables as:

$$q(t) - q(t_0) = D \ln [1 + \langle B(t - t_0) \rangle]$$

$$\text{where } D = \frac{kT}{E_{a1}} \qquad (7)$$

Fig. 5 shows a schematic plot of q versus log-time according to Eq. 7.

Eq. 6 and Fig. 5 show that at times short compared to 1/B, $q(t)-q(t_0)$ is approximately zero. At times long compared to 1/B, $q(t)-q(t_0)$ is approximately $D\ln[B(t-t_0)]$ or $D\ln(B)$

+ $D\ln(t-t_0)$. The two parameters of interest in this aging model are D, the per decade change in q for long times, and (1/B), the characteristic time at which the log-time behavior begins to appear. For the simple model described in Eq. 6,

$$BD = C\, e^{[-v_0]}, \tag{8}$$

$$\text{where } v_0 = E_{a0} + \frac{[E_{a1}\, q(t_0)]}{kT}.$$

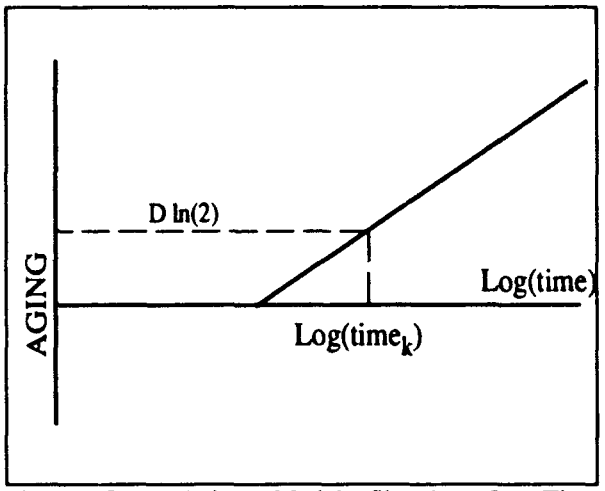


Figure 5 - Aging Model Showing Log-Time Dependence.

Therefore, BD is activated in an Arrhenius sense, if C, E_{a0}, and E_{a1} are independent of soak temperature.

It can be inferred from experimental results on high quality 8-MHz fundamental mode quartz resonators that B is a strongly decreasing function of temperature and that D is a strongly increasing function of temperature [138]. The authors are not aware of any reports on the dependence of BD on temperature.

Aging with Two Simultaneous Log-time Mechanisms.

There are no factors in Eq. 6 that could reverse the direction of the resonator aging. Since several authors have reported measurements that show a change in the direction of aging at some particular time [6,57,87,94,139,144], either a more complicated aging model must be found or there must be more than one simple log-time aging mechanism active in such resonators. For example, Fig. 6 shows schematically how two simple logarithmic aging mechanisms acting at the same time in a particular resonator can produce aging that changes direction at a particular time.

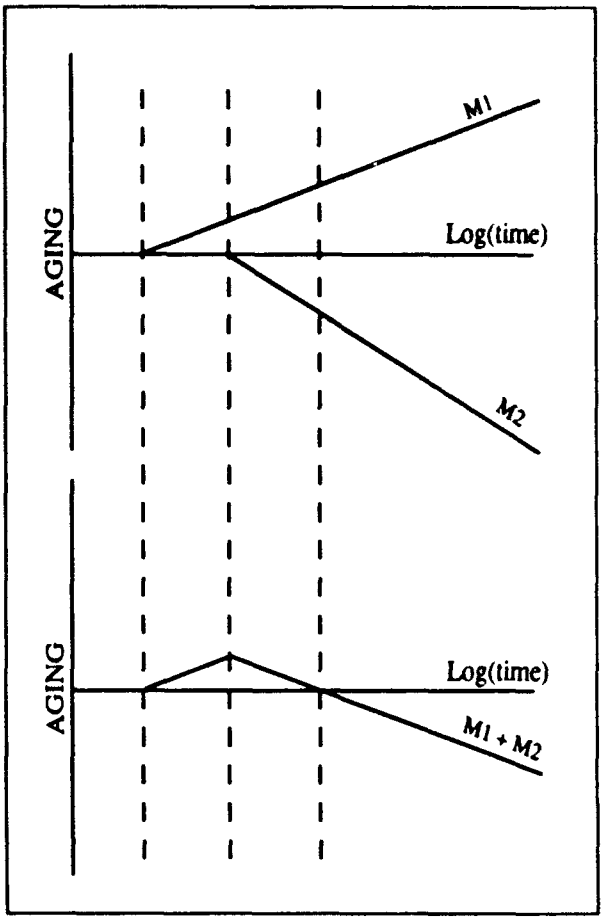


Figure 6 - Aging Model With Two Simple Log-time Mechanisms.

For this case the early logarithmic slope and time of the maximum positive change are mostly determined by the mechanism with the larger B. The time to zero aging has a component from both mechanisms. The long term logarithmic time slope is the sum of the log-time slopes of the two mechanisms. If the two long term slopes could be made equal (compensation), the aging would rise to a fixed value and then stop.

For some cases it may turn out that there is only a limited amount of possible aging in the particular device for one of the mechanisms. This case is not consistent with Eq. 4, and has not been reported in analyses of crystal aging.

In addition to some partly exponential time dependencies in diffusion, exponential time dependencies can arise from first order rate processes associated with adsorption-

desorption [9,68,72]. These kinds of processes can have an exponential time dependence if the activation energy does not depend on the amount of adsorption/desorption. A time dependence more complicated than exponential or logarithmic can arise if the activation energy also depends on the amount of adsorption/desorption.

In one report [68], an exponential aging model and simultaneous power law (diffusion) model fitted some isothermal data better than two different simultaneous log-time models. The significance of this result for other crystal fabrication technologies or for the majority of the filters made with the same technology is not known.

In another report [72], aging data were fitted better by log-time or by square root time (diffusion) models than by exponential time models, depending on the particular resonator being studied. In this work the log-time model was the best overall model.

Two experimental accelerated aging techniques, isothermal aging and thermal step stress aging, have been used to study the aging of crystal resonators (and the aging of other devices as well). These two techniques are discussed in the next two sections.

Isothermal Aging.

In this experimental technique each device (resonator, oscillator, etc.) is stored at only one of a set of temperatures and the frequency is measured at several times of particular interest [138,145]. Nearly all reported accelerated aging experiments have been of this type. The measurement times are selected to optimize the cost and resulting use of the data. Measurement time choices depend on many factors, such as the need for fixture calibration, the number of samples, and the type of device. For determining the parameters of a log-time model for a filter application, with large numbers of samples and the need for frequent fixture calibrations, a logarithmic spacing of measurement times is particularly efficient. For an oscillator aging study, closer spaced measurements may often be desirable, e.g., to characterize frequency jumps.

Fig. 7 shows a schematic plot of some typical isothermal aging data for a negative aging device at temperatures of T_1 and T_2 (where $T_2 > T_1$). The important information here is the general shape of the curves, the times at which the rapid changes begin, and the slopes of the rapidly changing parts.

A measurement temperature, which may be different from the soak temperature, for practical reasons, is selected to make the frequency measurement as accurate as possible. This measurement temperature is often a turnover temperature for high stability resonators, or can be room

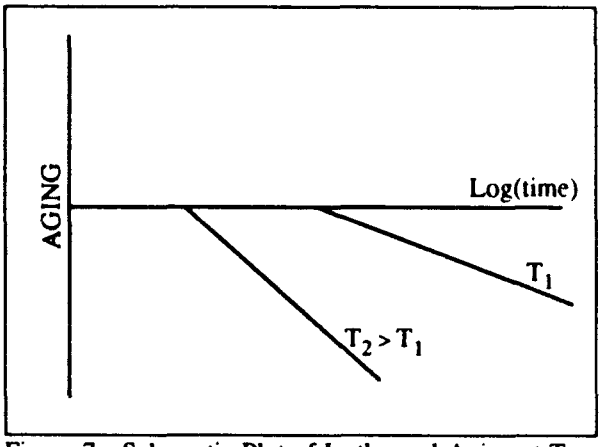


Figure 7 - Schematic Plot of Isothermal Aging at Two Temperatures.

temperature for other types of resonators. The resonator parameters may be measured directly or the resonator frequency may be measured in a test oscillator. Both kinds of measurements may be made in a temperature-controlled chamber.

For the most accurate measurements, the frequency is corrected to a fixed temperature to reduce the effects of small unavoidable temperature variations. It is important to show that the aging caused by the thermal shocks connected with the change from the soak temperature to the measurement temperature can be neglected or removed from the data.

The time dependent data is fitted (by least squares, for example) to a particular aging law to derive statistical estimates and associated confidence limits for the values of the aging law parameters or of the aging at a particular time of interest. For the log-time law the parameters are B and D; particular times of interest could be 1 year, 5 years, 20 years, etc.

Each member of a group of resonators is assigned to a different soak temperature, preferably by statistical experiment design rules. For each resonator soak temperature experiment the data are fitted to the selected aging law as above. The aging law parameters or frequency shifts calculated at a particular time of interest and their confidence limits can then be fitted to a selected law of the temperature dependence. An Arrhenius

temperature dependence law often turns out to be a useful choice.

The temperature dependence fit gives estimates and confidence limits of device aging or aging law parameters at any temperature of interest. For this temperature dependence fit, there are three sources of variability: 1) the choice of the wrong aging model, 2) the variability of the parameter estimates from each device, and 3) the error of each of the estimates.

The results of the two fittings, i.e., over time and over temperature, can only have meaning within the applicability of the assumed aging and temperature dependence laws. Since the results are highly statistical in nature, this technique can only characterize a particular fabrication technology, with the statistical meaning of the final predictions dependent on the validity of the choice of aging model, the number of samples chosen, and the number and ranges of times and temperatures selected for the design of the experiment. Within this context, statistical aging predictions can be made for any particular member of the characterized device group.

Thermal Step Stress (Differential Thermal Analysis).

In this experimental technique each device (resonator, oscillator, etc.) is stored for a fixed time interval at a series of increasing temperatures; the frequency is measured after each temperature soak [138,145]. Fig. 8 shows a schematic plot of typical thermal step stress data for a negatively aging device. The important information here is the general shape of the curve, the temperature at which rapid change begins, and the slope of the rapidly changing part. A measurement temperature is selected to make the frequency measurement as accurate as possible. This measurement temperature is often a turnover temperature or room temperature. For the most accurate measurements, the frequency is corrected to a fixed temperature to reduce the effect of small unavoidable temperature variations on the results. It is important to show that the aging caused by the thermal shocks connected with the changes from the soak temperatures to the measurement temperature can be neglected or removed from the data. This data can be plotted versus temperature to provide a curve, the shape of which is a signature of the distribution of aging mechanisms for that particular device.

For an assumed aging model, this data can also be used to determine estimates and confidence limits for the aging law parameters. For the log-time law, Eq. 5 gives an

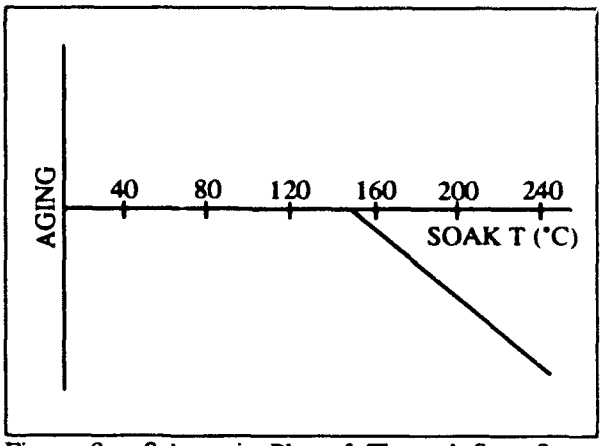


Figure 8 - Schematic Plot of Thermal Step Stress Aging

expression for the slope of frequency versus time as a function of temperature. After each temperature step, the change in frequency during that step is the slope for that temperature, as

$$\ln \frac{df(t)}{dt} = \ln C - \frac{E_{a0} + E_{a1} f(t)}{RT}. \qquad (9)$$

For example, Eq. 9 shows that the logarithm of the frequency change per unit time during each temperature step should be proportional to the reciprocal of the absolute temperature of the step; the proportionality constant is

$$\frac{E_{a0} + E_{a1} f(t)}{R},$$

where f(t) can be taken as the average frequency during the temperature step. The intercept of the fitted data with the T = infinity axis is the logarithm of C.

After the parameters of the aging law have been determined, Eq. 9 can be used to calculate the dependence of frequency on time at any temperature of interest. These calculations can be compared with isothermal results on similar devices.

The advantage of the thermal step stress technique over the isothermal technique is that the aging parameters for a particular device can be derived in a short time, e.g., two weeks for a 16 hour time step. When the thermal step stress results compare well with the isothermal results, the thermal step stress technique can quickly

provide useful estimates of aging parameters.

If the high temperature parts of the thermal step stress results show that extraneous high temperature aging mechanisms are present, it may be necessary to use a longer time step at lower temperatures to obtain the desired data. These extraneous high temperature mechanisms may include melting processes or phase changes for some of the materials in the device being tested. The thermal step stress results provide a useful way to quickly reveal the presence of these mechanisms so that their contribution to the device long term aging can be considered and resolved.

Other Aging Characterization Techniques.

1. Soaking the completed resonators at a specified temperature for a specified period of time sufficient to produce a small but measurable aging of the resonator. Resonators with frequencies that change too much or too little, or that change in a different direction during this temperature soak, differ from other resonators in the same production lot and should, therefore, be candidates for rejection and failure mode analysis to determine the possible causes for the differences.

2. Installing the completed resonators in test ovens at the system use temperature and carefully measuring frequency changes per unit time until the change per unit time reduces to a prescribed value. If the prescribed frequency change per unit time is not reached in a prescribed time, the resonator is a candidate for rejection. Resonators that achieve the desired frequency change per unit time too early probably should also be rejected, but this is not often done.

It is very important that all thermally accelerated aging procedures be based on a specified aging model. The procedures should be performed to provide data from which parameters of the assumed aging model can be determined. Determination of the appropriate aging models requires intuition and data other than aging data, as well as the resonator aging data. Inappropriate aging models can sometimes be identified with resonator aging data alone. Some very-low-aging resonators appear to exhibit a linear frequency change with time after some stabilization period. It may be that data on resonators with log-time aging and an extended stabilization period may appear to be linear over a particular measurement time frame, such as 5 years.

There may also be apparent resonator aging mechanisms, i.e., drift, which exhibit linear time dependence. For example, resonator frequency changes due to continuous low level radiation from space or background sources might be linear.

Crystal resonators may have many kinds of aging mechanisms. Since the various aging mechanisms can cause both positive and negative aging, compensation among the various mechanisms is possible. This compensation may occur during the early history of the resonator and then become ineffective later; or, occasionally, it may occur years after the aging starts. For this reason great care must be taken to characterize the aging mechanisms associated with a particular resonator fabrication technology over a sufficiently wide range of temperatures and time before aging predictions can be validated or qualified. After the fabrication technology is qualified to produce low-aging resonators for the required times and temperatures of the application, subsequent production must be monitored to assure that the manufacturing process remains in qualification.

Conclusions.

1. Even though several possible aging mechanisms are well understood, the aging of resonators is still not well understood.

2. Aging performance, including variations with temperature and drive level, depends on the resonator design and fabrication technology.

3. Many processing deviations can degrade aging performance.

4. High temperature processing seems necessary (but not sufficient) for the production of low-aging resonators.

5. The SC-cut, and modern ultrahigh vacuum and high temperature fabrication techniques have resulted in resonators which achieve low-aging in a shorter period of time than the best resonators made a generation ago, however, the aging of the best modern resonators after extended periods is no better today than what was reported for the best resonators in the 1960s.

6. Accelerated aging studies are useful for process control. Using accelerated aging data for long term aging predictions is possible, but considerable work and expense are required to reduce the risk of error to acceptable confidence levels.

7. The best reproducible long term aging rates seem to be a few times 10^{-11} per day. Occasional resonators exhibit aging rates of a few times 10^{-12} per day after extended periods.

8. Environmental changes can produce frequency changes that appear to be aging. This apparent aging is now called "drift."

Although the state-of-the-art in the long term aging of low-aging resonators has been on a plateau for more than a generation, there is no reason to believe that the factors responsible for limiting the achievable long term aging are insurmountable. The definitive experiments, in which all known aging mechanisms are minimized, are yet to be performed. It is the authors' hope that the review in this paper will assist future researchers in the design of experiments that result in significant improvements in aging.

References.

1) "MIL-O-55310, Military Specification, Oscillators, Crystal, General Specification for," the latest revision is available from Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

2) CCIR Recommendation No. 686, "Glossary," CCIR 17th Plenary Assembly, Vol. VII, "Standard Frequency and Time Signals (Study Group 7)," (1990). Consultative Committee on International Radio (CCIR); copies available from: International Telecommunications Union, General Secretariat - Sales Section, Place des Nations, CH-1211 Geneva, Switzerland.

3) "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology," IEEE Std 1139-1988, The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017.

4) E. A. Gerber, "Long Term Stability and Aging of Resonators," in Precision Frequency Control, Volume 1, E. A. Gerber and A. Ballato (eds.), Chapt. 6, pp. 271-284, Academic Press, New York, 1985.

5) "MIL-C-49468, Military Specification, Crystal Units, Quartz, General Specification for," the latest revision is available from Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia. PA 19111-5094.

6) R. L. Filler, "Aging Specification, Measurement, and Analysis," Proceedings of the 7th Quartz Devices Conference & Exhibition, pp. 93-104, 1985; The Electronic Industries Assoc., 1722 Eye St., NW, Washington, DC 20006.

7) P. T. Landsberg, "On the Logarithmic Rate Law in Chemisorption and Oxidation," The J. of Chem. Physics, Vol. 23, pp. 1079-1087, 1955.

8) I. M. Ritchie, "The Oxidation of Metal Films," Thin Solid Films, Vol. 34, pp. 83-86, 1976.

9) R. Glang, et al., "High Vacuum Technology," in Handbook of Thin Film Technology, L. I. Maissel and R. Glang, editors, McGraw Hill, pp. 2-39 to 2-59, 1970.

10) W. G. Perkins, "Permeation and Outgassing of Vacuum Materials," J. Vac. Sci. Technol., Vol. 10, pp. 543-556, 1973.

11) C. D. Stockbridge, "Effect of Hydrostatic Pressure on Rotated Y-Cut Quartz Crystal Resonators," Vacuum Microbalance Techniques, Vol. 5, Plenum Press, New York, 1966, pp. 179-191.

12) C. D. Stockbridge, "Effects of Gas Pressure on Quartz Crystal Microbalances," Vacuum Microbalance Techniques, Vol. 5, Plenum Press, New York, 1966, pp. 147-178.

13) J. R. Vig and E. Hafner, "Packaging Precision Quartz Crystal Resonators," R & D Tech. Report ECOM-4134, July 1973, AD 76?215.

14) R. Abermann, "Measurements of the Intrinsic Stress in Thin Metal Films," Vacuum, Vol. 41, pp. 1279-1290, 1990.

15) R. W. Hoffman, "Stresses in Thin Films: The Relevance of Grain Boundaries and Impurities," Thin Solid Films, Vol. 34, pp. 185-190, 1976.

16) D. M. Mattox, "Thin Film Metallization of Oxides in Microelectronics," Thin Solid Films, Vol. 18, pp. 173-186, 1973.

17) M. Adamov, et al., "Electrical and Structural Properties of Thin Gold Films Obtained by Vacuum Evaporation and Sputtering," Thin Solid Films, Vol. 24, pp. 89-100, 1974.

18) J. G. Vaughan, et al., "Effect of Partial Pressure and Temperature on Properties of Vacuum Deposited Chromium-Gold Thin Films," J. Vac. Sci. Technol., Vol. 20, No. 3, pp. 383-387, March 1982.

19) B. Kebabi, et al., "Stress and Microstructure Relationships in Gold Thin Films," Vacuum, Vol. 41, Nos. 4-6, pp. 1353-1355, 1990.

20) R. Abermann and R. Koch, "Internal Stress of Thin Silver and Gold Films and its Dependence on Gas Absorption," Thin Solid Films, Vol. 62, pp. 195-208, 1979.

21) L. I. Maissel and M. H. Francombe, An Introduction to Thin Films, pp. 200-210, Gordon & Breach Science Publishers, 1973.

22) E. Murr and M. C. Inman, "Effects of Vacuum Environment on the Sub-structure of Evaporated F.C.C. Metal Films," Phil. Mag., Vol. 14, pp. 135-153, 1966.

23) E. Klokholm and J. F. Freedman, "Magnetostress Effects in Evaporated Ni Films," J. Appl. Phys., Vol. 38, pp. 1354-1356, 1967.

24) F. R. Botzen, et al., "Creep of Thin Metallic Films," Vacuum, Vol. 41, pp. 1287-1290, 1990.

25) J. M. Heras and E. E. Mola, "Resistivity Study of the Annealing Kinetics of Thin Evaporated Gold Films," Thin Solid Films, Vol. 35, pp. 75-82, 1976.

26) A. P. Dorey and J. Knight, "The Variation of Resistance of Gold Films with Time and Annealing Procedure," Thin Solid Films, Vol. 4, pp. 445-451, 1969.

27) R. J. Miller and A. Gangulee, "Electromigration in Gold and Copper Thin Film Conductors," Thin Solid Films, Vol. 69, pp. 379-386, 1980.

28) J. D. Wilcock, et al., "The Internal Stress in Evaporated Silver and Gold Films," Thin Solid Films, Vol. 3, pp. 13-34, 1969.

29) C. A. Neugebauer, "Tensile Properties of Thin, Evaporated Gold Films," J. of Appl. Physics, Vol. 31, No. 6, pp. 1096-1101, 1960.

30) L. Eckertova, Physics of Thin Films, Plenum Press, New York and London, 1977, pp. 154-158.

31) H. Hieber, "Aging Properties of Gold Layers with Different Adhesion Layers," Thin Solid Films, Vol. 37, pp. 335-343, 1976.

32) M. Laugier, "Adhesion and Internal Stress in Thin Films of Aluminum," Thin Solid Films, Vol. 79, pp. 15-19, 1981.

33) J. R. Lloyd and S. Nakahara, "Low Temperature Void Growth and Resistivity Decay in Thin Evaporated Gold Films," Thin Solid Films, Vol. 45, pp. 411-420, 1977.

34) E. P. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proc. 29th Ann. Symp. on Frequency Control, pp. 1-4, 1975.

35) R. J. Besson, "A New Electrodeless Resonator Design," Proc. 31st Ann. Symp. on Frequency Control, pp. 147-152, 1977.

36) J. R. Norton, "BVA-Type Quartz Oscillator for Spacecraft," Proc. 45th Ann. Symp. on Frequency Control, 1991.

37) G. Majni, et al., "Interdiffusion of Thin Cr and Au Films Deposited on Silicon," Thin Solid Films, Vol. 38, pp. 15-19, 1976.

38) P. H. Holloway, "Gold/Chromium Metallizations for Electronic Devices," Solid State Technology, pp. 109-115, February 1980.

39) Y. E. Strausser, et al., "Observations of Al_2O_3 and Free Silicon at the Interface Between Aluminum Films and SiO_2," Thin Solid Films, Vol. 52, pp. 203-214, 1978.

40) R. J. Blattner and A. J. Braundmeier, Jr., "Solid Phase Reduction of SiO_2 in the Presence of an Al Layer," J. Vac. Sci. Technol., Vol. 20, No. 3, pp. 320-323, March 1982.

41) E. Philofsky and E. L. Hall, "A Review of the Limitations of Aluminum Thin Films on Semiconductor Devices," IEEE Trans. on Parts, Hybrids, and Packaging, Vol. PHP-11, No. 4, pp. 281-290, December 1975.

42) J. R. Black, "The Reaction of Al with Vitreous Silica," 15th Ann. Proc. Reliability Physics 1977, pp. 257-260, 1977, IEEE Cat. No. 77 CH1195-7PHY.

43) K. Haruta and W. J. Spencer, "Strain in Thin Metal Films on Quartz," Journal of Applied Physics, Vol. 37, No. 6, pp. 2232-2233, May 1966.

44) V. E. Bottom, "A Novel Method of Adjusting the Frequency of Aluminum Plated Quartz Crystal Resonators," Proc. 30th Ann. Symp. on Frequency Control, pp. 249-253, 1976.

45) C. W. Shanley and L. N. Dworsky, "DC Plasma Anodization of Quartz Resonators," Proc. 36th Ann. Symp. on Frequency Control, pp. 108-114, 1982.

46) J. L. Jellison, "The Role of Surface Contaminants in the Solid-State Welding of Metals," in Treatise on Clean Surface Technology, Vol. 1, ed. by K. L. Mittal, Plenum Press, pp. 205-234, 1987.

47) J. R. Vig, "UV/Ozone Cleaning of Surfaces," in Treatise on Clean Surface Technology, Vol. 1, ed. by K. L. Mittal, Plenum Press, pp. 1-26, 1987.

48) R. L. Filler, et al., "Ceramic Flatpack Enclosed AT and SC-cut Resonators," Proc. 1980 IEEE Ultrasonics Symp., pp. 819-824, 1980.

49) Private communication, M. Bloch (to J. Vig), Frequency Electronics, Inc., May 1991.

50) Private communication, C. Jensik (to J. Vig), Piezo Crystal Co., May 1991.

51) E. Hafner and R. S. Blewer, "Low-Aging Quartz Crystal Units," Proceedings of the IEEE, Vol. 56, No. 3, pp. 366-368, March 1968.

52) Private communications, C. A. Adams and J. A. Kusters, Hewlett-Packard Co. (to J. Vig), May 1991.

53) J. G. Gualtieri, "The Influence of Temperature and Electric Field on the Etch-Channel Density in Swept-Cultured Quartz," Proc. 39th Ann. Symp. on Frequency Control, pp. 247-254, 1985.

54) R. B. Belser and W. H. Hicklin, "Comparison of Aging Performance of 5-MHz Resonators Plated with Various Electrode Metals," Proc. 23rd Ann. Symp. on Frequency Control, pp. 132-142, 1969.

55) J. C. Brice, "Crystals for Quartz Resonators," Reviews of Modern Physics, Vol. 57, No. 1, pp. 105-138, January 1985.

56) M. Hendrickson and J. R. Vig, unpublished, 1991.

57) R. L. Filler, et al., "Aging Studies on Quartz

Crystal Resonators and Oscillators," Proc. 38th Ann. Symp. on Frequency Control, pp. 225-231, 1984.

58) P. I. Aseev, et al., "Influence of Surface Catalytic Reactions on Crystal Unit Aging," Proc. 45th Ann. Symp. on Frequency Control, elsewhere in these proceedings, 1991.

59) A. Ballato, et al., "The Force-Frequency Effect in Doubly Rotated Quartz Resonators," Proc. 31st Ann. Symp. on Frequency Control, pp. 8-16, 1977.

60) R. A. Young and C. E. Wagner, "X-ray Source-Image Distortion Technique for the Study of Crystal Distortion and Vibration," Journal of Applied Physics, Vol. 37, No. 11, pp. 4070-4076, October 1966.

61) R. A. Young, et al., "Special X-Ray Studies of Quartz Frequency Control Units," Proc. 19th Ann. Symp. on Frequency Control, pp. 23-41, 1965.

62) D. L. Hammond, "Precision Quartz Resonators," Proc. 15th Ann. Symp. on Frequency Control, pp. 125-138, 1961.

63) Z. D. Jastrzebski, The Nature and Properties of Engineering Materials, 2nd ed., pp. 253-260, 1976.

64) B. Chalmers, Physical Metallurgy, pp. 139-191 and 328-332, John Wiley & Sons., 1959.

65) T. E. Parker, "A New All Quartz Package for SAW Devices," Proc. 39th Ann. Symp. on Frequency Control, pp. 519-525, 1985.

66) E. P. EerNisse, "Temperature Dependence of the Force-Frequency Effect for the AT-, FC-, SC-, and Rotated X-cuts," Proc. 34th Ann. Symp. on Frequency Control, pp. 426-430, 1980.

67) R. L. Filler and J. R. Vig, "The Effect of Bonding on the Frequency vs. Temperature Characteristics of AT-cut Resonators," Proc. 30th Ann. Symp. on Frequency Control, pp. 264-268, 1976.

68) G. L. Dybwad, "Aging Analysis of Quartz Crystal Units with Ti Pd Au Electrodes," 31st ASFC, pp. 144-146B, 1977.

69) J. Crank, The Mathematics of Diffusion, Clarendon Press, Oxford, 1956/1975.

70) American Institute of Physics Handbook, D. E. Gray (ed.), Second Edition, Office of Science Information, National Science Foundation, McGraw Hill Company, New York, 1957/1963, pp. 2-77 and 2-78.

71) L. E. Halliburton and J. J. Martin, "Properties of Piezoelectric Materials," in Precision Frequency Control, Volume 1, E. A. Gerber and A. Ballato (eds.), Chapter 1, pp. 100-106, Academic Press, New York, 1985.

72) M. R. Miljkovic, et al., "Aging Prediction of Quartz Crystal Units," Proc. 42nd Annual Symp. on Frequency Control, pp. 404-411, 1988.

73) L. H. Van Vlack, Elements of Material Science, Second edition, Addison-Wesley Publishing Co., Reading, Massachusetts, pp. 100-106, 1959/1964.

74) J. R. Vig, et al., "Chemically Polished Quartz," Proc. 31st Ann. Symp. on Frequency Control, pp. 131-143, 1977.

75) R. A. Murray, "Movement of Dislocations in Quartz," Proc. 45th Ann. Symp. on Frequency Control, elsewhere in these Proc., 1991.

76) D. W. Hart, et al., "A Study of the Electrodiffusion Process in Quartz," Proc. 44th Ann. Symp. on Frequency Control, pp. 222-227, 1990.

77) F. B. Johnson and R. S. Pease, "The Pile Irradiation of Quartz Crystal Oscillators," Phil. Mag., Vol. 45, pp. 651-654, 1954.

78) W. Primak, "Fast-Neutron-Induced Changes in Quartz and Vitreous Silica," Phys. Rev., Vol. 110, pp. 1240-1254, 1958.

79) R. A. Laudise and R. L. Barns, "Perfection of Quartz and its Connection to Crystal Growth," IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, Vol. 35, pp. 277-287, 1988.

80) H. W. Jackson, "Tactical Miniature Crystal Oscillator," Proc. 34th Ann. Symp. on Frequency Control, pp.449-456, May 1980.

81) R. L. Filler, "The Amplitude-Frequency Effect in SC-cut Resonators," Proc. 39th Ann. Symp. on Frequency Control, pp. 311-316, 1985.

82) F. L. Walls, "The Influence of Pressure and Humidity on the Medium and Long Term Frequency Stability of Quartz Oscillators," Proc. 42nd Ann. Symp. on Frequency Control, pp. 279-283, 1988.

83) F. L. Walls, "Environmental Sensitivities of Quartz Crystal Oscillators," Proc. 22nd Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, NASA Conference Publ. 3116, pp. 465-486, 1990.

84) 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 Frequency Control, pp. 180-188, 1978.

85) K. E. Grata, "Long Term Crystal Stability Study," Proc. 5th Quartz Crystal Conference, pp. 214-221, 1983; The Electronic Industries Assoc., 1722 Eye St., NW, Washington, DC 20006.

86) O. N. Efremov, et al., "Aging of Precision Quartz Resonators," Izmeritel'naya Tekhnika, No. 3, pp. 74-75, March 1975.

87) M. I. Jaroslavsky and V. D. Lavrentsov, "Long Term Frequency Variations of Quartz Crystal Units Under Different Environmental Conditions," Ferroelectrics, Vol. 43, pp. 51-56, 1982.

88) A. W. Warner, "Design and Performance of Ultraprecise 2.5-mc Quartz Crystal Units," The Bell

System Technical Journal, pp. 1193-1217, September 1960.

89) J. H. Armstrong, et al., "Aging Characteristics of Quartz Crystal Resonators," Proc. 20th Ann. Symp. on Frequency Control, pp. 192-207, 1966.

90) J. Messina, et al., "Results of Long Term Testing of Tactical Miniature Crystal Oscillators," Proc. 43rd Ann. Symp. on Frequency Control, pp. 47-50, 1989.

91) P. A. Simpson and A. H. Morgan, "Quartz Crystals at Low Temperatures," Proc. 13th Ann. Symp. on Frequency Control, pp. 207-212, 1959.

92) P. A. Simpson and A. H. Morgan, "Investigation of Stability of Quartz Resonators at Low Temperatures," NBS Report 6080, 1959, AD 230714.

93) G. Robichon, et al., "Frequency Stability of Quartz Crystal at Very Low Temperatures: Preliminary Results," Proc. 38th Ann. Symp. on Frequency Control, pp. 201-205, 1984.

94) R. L. Filler, et al., "The Aging of Resonators and Oscillators Under Various Test Conditions," Proc. 41st Ann. Symp. on Frequency Control, pp. 444-451, 1987.

95) R. L. Filler, et al., "Frequency-Temperature and Aging Performance of Microcomputer Compensated Crystal Oscillators," Proc. 43rd Ann. Symp. on Frequency Control, pp. 27-33, 1989.

96) R. J. Byrne and J. L. Hokanson, "Effect of High-Temperature Processing on the Aging Behavior of Precision 5-MHz Quartz Crystal Units," IEEE Transactions on Instrumentation and Measurement, Vol. IM-17, No. 1, pp. 76-79, March 1968.

97) R. C. Smythe and R. B. Angove, "Chemically Milled UHF SC-Cut Resonators," Proc. 42nd Ann. Symp. on Frequency Control, pp. 73-77, 1988.

98) R. L. Filler and R. C. Lindenmuth, U. S. Army Electronics Technology and Devices Laboratory, 1991, unpublished.

99) T. E. Parker and G. K. Montress, "Precision Surface-Acoustic-Wave (SAW) Oscillator," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 35, No. 3, pp. 342-364, May 1988.

100) A. W. Warner, "Ultra-Precise Quartz Crystal Frequency Standards," IRE Trans. on Instrumentation, Vol. I-7, pp. 185-188, December 1958.

101) W. J. Spencer and W. L. Smith, "Precision Quartz Crystal Controlled Oscillator for Severe Environmental Conditions," Proc. 16th Ann. Symp. on Frequency Control, pp. 405-420, 1962.

102) R. A. Sykes, et al., "Performance of Precision Quartz-Crystal Controlled Frequency Generators," IRE Trans. on Instrumentation, Vol. I-11, pp. 243-247, December 1962.

103) W. J. Spencer and R. L. Reynolds, "Characteristics of Frequency Control Devices for Satellite Environments," in Symposium on Cleaning and Materials Processing for Electronics and Space Apparatus, ASTM Spec. Publ., Vol. 342, pp. 155-163, 1962.

104) A. W. Warner, et al., "Fundamental Studies of Aging in Quartz Resonators," IEEE Transactions on Sonics and Ultrasonics, Vol. SU-12, No. 2, pp. 52-59, June 1965.

105) E. A. Gerber and R. A. Sykes, "Quartz Frequency Standards," Proceedings of the IEEE, Vol. 55, No. 6, pp. 783-791, June 1967.

106) R. J. Besson and U. R. Peier, "Further Advances on B. V. A. Quartz Resonators," Proc. 34th Ann. Symp. on Frequency Control, pp. 175-182, May 1980.

107) R. J. Besson, et al., "Performance of New Oscillators Designed for 'Electrodeless' Crystals," Proc. 34th Ann. Symp. on Frequency Control, pp. 457-462, May 1980.

108) J. R. Vig, "The Effects of Surface Contamination on the Noise and Drive-Level Sensitivity of Piezoelectric Resonators," U. S. Army Laboratory Command R&D Tech. Report SCLET-TR-87-5, May 1987, AD A181299.

109) J. J. Suter and R. H. Maurer, "Low and Medium Dose Radiation Sensitivity of Quartz Crystal Resonators with Different Al-Impurity Content," Proc. 40th Ann. Symp. on Frequency Control, pp.134-139, 1986.

110) L. J. Palkuti and Q. T. Truong, "An X-ray Irradiation System for Total-Dose Testing of Quartz Resonators," Proc. 38th Ann. Symp. on Frequency Control, pp. 55-62, 1984.

111) D. A. Berlincourt, et al., "Piezoelectric and Piezomagnetic Materials and Their Function in Transducers," in Physical Acoustics, W. P. Mason, editor, Ch. 3, Volume 1 - Part A. Academic Press, pp. 201 and 207, 1964.

112) S. Fujishima, "Research and Development of Piezoelectric Devices in Japan," Proc. Ultrasonics Symp., pp. 128, September 1979.

113) "IRE Standards on Piezoelectric Crystals - Measurements of Piezoelectic Ceramics," Proceedings of the IRE, Vol. 49, No. 7, pp. 1161-1169, 1961.

114) R. Lefevre, et al., "Laser Processed Miniature $LiTaO_3$ Resonators and Monolithic Filters," Proc. 39th Ann. Symp. on Frequency Control, pp. 333-337, 1985.

115) J. M. Power, et al., "Long Range Hydrophones," Applications of Ferroelectric Polymers, T. T. Wang, et al., editors, Chapter 6, pp. 128, 1988.

116) E. A. Gerber and R. A. Sykes, "State-of-the-Art Quartz Crystal Units and Oscillators," Proc. IEEE 54, pp. 103-116, 1966.

117) S. Kanbayashi, et al., "Analysis of Tuning Fork Crystal Units and Application into Electronic Wrist

Watches," Proc. 30th Ann. Symp. on Frequency Control, pp. 167-174, 1976.

118) J. Engdahl and H. Matthey, "32-kHz Quartz Unit for High Precision Wrist Watch," Proc. 29th Ann. Symp. on Frequency Control, pp. 187-194, 1975.

119) H. Yoda, et al., "Low Power Oscillation for Electric Wrist Watch," Proc. 26th Ann. Symp. on Frequency Control, pp. 140-147, 1972.

120) M. P. Forrer, "A Flexure-Mode Quartz for an Electronic Wrist Watch," Proc. 23rd Ann. Symp. on Frequency Control, pp. 157-162, 1969.

121) H. Yoda and N. Hoie, "Technical Aspects of Wrist Watches," Proc. 28th Ann. Symp. on Frequency Control, pp. 57-66, 1974.

122) D. T. Bell, Jr., "Aging Processes in SAW Resonators," Proc. 1977 Ultrasonics Symp., pp. 851-856, 1977.

123) W. R. Shreve, "Aging in Quartz SAW Resonators," Proc. 1977 Ultrasonics Symp., pp. 857-861, 1977.

124) C. A. Adams and J. A. Kusters, "Improved Long Term Aging in Deeply Etched SAW Resonators," Proc. 32nd Ann. Symp. on Frequency Control, pp. 74-76, 1978.

125) W. R. Shreve, et al., "Fabrication of SAW Resonators for Improved Long Term Aging," Proc. 1978 Ultrasonics Symp., pp. 573-579, 1978.

126) W. R. Shreve, "Active Aging of SAW Resonators," Proc. 1980 Ultrasonics Symp., pp. 188-192, 1980.

127) J. I. Latham and D. Saunders, "Aging and Mounting Developments for SAW Resonators," Proc. 1978 Ultrasonics Symp., pp. 513-517, 1978.

128) M. Gilden, et al., "Long Term Aging and Mechanical Stability of 1.4 GHz SAW Oscillators," Proc. 1980 Ultrasonics Symp., pp. 184-187, 1980.

129) T. E. Parker, "Analysis of Aging Data on SAW Oscillators," Proc. 34th Ann. Symp. on Frequency Control, pp. 292-301, 1980.

130) T. E. Parker, "Development of Precision SAW Oscillators for Military Applications," Proc. 36th Ann. Symp. on Frequency Control, pp. 453-458, 1982.

131) T. E. Parker, "Precision Surface Acoustic Wave (SAW) Oscillators," Proc. 1982 Ultrasonics Symp., pp. 268-274, 1982.

132) E. Simon, "Long Term Frequency Stability of Crystal Oscillators," Technical Report ECOM-2787, December 1966, AD 647742.

133) M. B. Bloch and J. L. Denman, "Further Development on Precision Q [illegible] Resonators," Proc. 28th Ann. Symp. on Frequency Control, pp. 73-84, 1974.

134) Nuclear Energy, pamphlet no. EDM-1016, U.S. Department of Energy, May 1976.

135) T. M. Flanagan, et al., "Evaluation of Mechanisms for Low-Dose Frequency Shifts in Crystal Oscillators," 40th Ann. Symp. on Frequency Control, pp. 127-133, 1986.

136) J. R. Vig, "Resonator Aging," 1977 Ultrasonics Symposium Proceedings, IEEE Cat. No. 77CH1264-1SU, pp. 848-849, 1977.

137) S. Glasstone, Textbook of Physical Chemistry, Second Edition, Seventh Printing, D. Van Nostrand Co., New York, 1940/1951.

138) S. H. Olster, et al., "A6 Monolithic Crystal Filter Design for Manufacture and Design Quality," Proc. 29th Ann. Symp. on Frequency Control, pp. 105-112, 1975.

139) J. R. Gehrke and R. Klawitter, "Experimental Results on Aging of AT-cut Strip Resonators," Proc. 42nd Ann. Symp. on Frequency Control, pp. 412-418, 1988.

140) A. A. Feinberg, "Parametric Failure Rate Model for Quartz Crystal Device Aging with Application to Surface Acoustic Wave Filters," Proc. 41st Ann. Symp. on Frequency Control, pp. 360-364, 1987.

141) D. E. Beetley, "Nine-Year Aging Behavior of the Ceramic Flatpack Resonator," Proc. 3rd European Time and Frequency Forum, pp. 139-142, 1989.

142) D. E. Beetley, "Nine-Year Aging Behavior of the Ceramic Flatpack Resonator," U. S. Dept. of Energy contract report, GE Aerospace Tech. Info. Series, GEPP-TIS-1168, UC-706, March 6, 1990. Copies available from: GEND Tech. Info. Center, P.O. Box 2908, Largo, FL 34649-2908, or from NTIS.

143) A. H. Cottrell and V. Aytekin, "The Flow of Zinc Under Constant Stress," J. Inst. Metals, Vol. 77, pp. 389-422, 1950.

144) D. A. Emmons, "Tactical Quartz Oscillator Development," RADC-TR-90-232 Final Technical Report, pp. 7-30, September 1990.

145) T. R. Meeker and W. R. Grise, "Packaging and Reliability of SAW Filters," Proc. 1983 Ultrasonics Symposium, pp. 117-124, 1983.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INFLUENCE OF SURFACE CATALYTIC REACTIONS ON CRYSTAL UNIT AGEING

P.I.Aseev, V.B.Grouzinenko, A.S.Rodichev

Scientific-Research Institute "Phonon"
Moscow 105023, Krasnobogatyrskaya 44, U.S.S.R.

Abstract

The process of heterogen-catalytic reactions of deep hydrocarbons oxidation at the plated electrode surface over a wide temperature range (20°C to 500°C) is described. It is shown, that the existence of Na and Ca type contaminants accelerates and those of S and Cr slower the reaction rate. Direct relation between the positive component of quartz crystal unit ageing and the value of catalytic activity of electrode. Silver emission from the electrode surface accompanying catalytic selver oxidation both during the manufacture and the storage of quartz crystal units was detected.

Introduction

Ageing continues to remain one of the most urgent problems of further crystal unit improvement. In order to resolve it, the amelioration of synthetic quartz crystal growing process is necessary, ensuring its high homogeneity and improvement of the material structure; the search for new double-rotated crystal cuts; development of the present state-of-the-art techniques of surface treatment and techniques of removing surface contaminants; use of ion-beam techniques for treatment of crystal unit details; development of new design solutions etc. This paper is an attempt to consider the poorly investigated mechanism of crystal unit frequency variations depending on metallized piezoelectric elements, namely, the interaction of electrode material with the environment.

Experimental

The processes taking place within the resonator volume can be classified into quick and slow ones. The activation energy of the quick processes corresponds to the energy of disperse interaction of residual gases with the surface of the order of 10^{-2} eV. Over this energy range a dynamic equilibrium is quickly established between the surface regions, and these processes do not influence the ageing. The activation energy of slow processes is equal to or is higher than the energy of hydrogen bonds. Thus, sorbtion and desorbtion - are the essence of the initial and the final stages of slow surface and presurface processes going on with different characteristic times up to $10^{7} - 10^{8}$ s, which corresponds to characteristic times of catalytic reactions under low pressures.

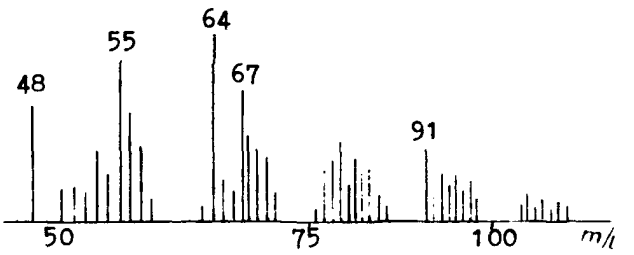


Fig. 1: Part of mass-spectres over the range of 45-1000 a.m.u. at the isometric portion 400°C of piezoelectric elements for crystal units after hermetical sealing

Fig. 1 shows a part of a mass-spectre of piezoelectric elements for quartz crystal units after their hermetical sealing in an enclosure. The peaks 55, 67, 78 and 91 a.m.u., serve as an

CH2965-2/91/0000-102 $1.00 © 1991 IEEE

indication of the presence of different hydrocarbon compounds on the electrode surface.

Analysis of a mass-spectre of negative secondary ions coming from the surface of the crystal unit electrode (Fig.2) showed the existence of an intense peak equal to 16 a.m.u., which is connected with the presence of oxygen in silver electrode. It does not contradict to the data given in a number of papers dealing with investigation of forms of oxygen adsorbed at the surface of silver single crystals at room temperatures [1, 2]. As distinct from other gas impurities oxygen strongly adsorbs and dissolves in silver electrode, oxygen dissolubility being increased with the increase of the temperature of thermal treatment, which agrees well with the results described in paper 1 .

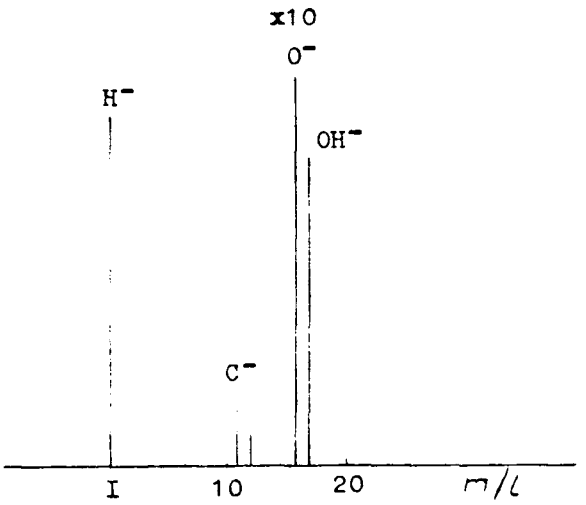


Fig. 2: Part of mass-spectre of negative secondary ions going from the silver electrode surface of a quartz crystal unit over the range of 1-20 a.m.u.

The intense negative peak $\bar{O}$ of oxygen obtained by SIMS method in the freshly evaporated silver electrodes can be compared to the results published in the paper [1] , where by using the RSCA method dissociatively adsorbed oxygen was found, i.e. oxygen existing in an atomic form, the molecular form of oxygen being completely desorbed already at the temperature of 100°C. Oxygen in the atomic form [3] takes part in a deep destructive oxidation of hydrocarbon to take the form of CO_2 and H_2O. Therefore, the intense peaks of 44 and 18 a.m.u., found by thermodesorbtion mass-spectrometry while studing electrodes of quartz crystal units can be explained by deep oxidation of hydrocarbons at the electrode surface. This is the reason for enhancement of CO_2 and H_2O peaks intensity depending on the interoperational storage time and on the presence of tuning spots, because the hydrocarbon sorbtion takes place without saturation [4] and is determined by oxigen concentration value at the electrode surface. Therefore, the yield of the reaction increases in this case.

Thus, during thermal treatment under vacuum or in the air, the surface of electrodes is released from absorbed hydrocarbons owing to the catalytic oxidation of hydrocarbons up to CO_2 and H_2O.

In the process of catalytic oxidation of hydrocarbons a silver atoms emission from electrode surface was observed [5] . The theoretical value of desorbtion energy of silver atoms is approximately 0.5 eV [6] . The energy generated during hydrocarbons oxidation with the formation of hydroxyl bond only, equal to 5.1 eV under condition that the hydrocarbon bond is completely destroyed. Under assumption that the hydrocarbon bond is not destroyed, the energy difference $E_{C-H} - E_{O-H}$ equals to 0.46 eV. This is rather close to the energy value for the silver atom desorbtion. It follows from this that catalytic reactions are observed at the electrode surface, accompanied by silver atoms emission.

Resonator ageing

Experimental results showed that the values of catalytic activity of crystal unit electrodes manufactured in a single technological cycle differ, nevertheless. This differentation is observed not only between electrodes of different resonators, but also at the electrode surface of a single resonator. The catalithic activity of electrodes was determined as follows. The crystal unit was inserted in a vacuum chamber, where hydrocarbon similar forms of radicals of C_3H_8 type were formed. An initiator of broken bonds was widely used, which constituted plasmatron of homopolar

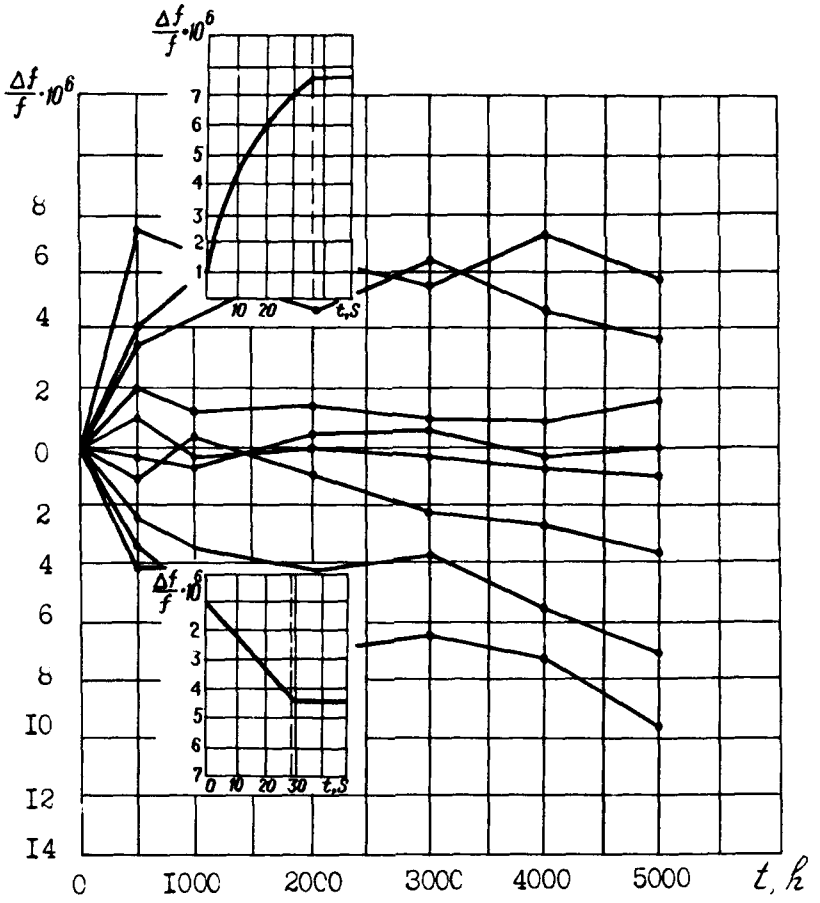


Fig. 3: Typical curves of crystal unit ageing at the storage temperature of 100°C

type, localizing gas discharge under the pressure of order 10^{-2} Pa within a small chamber volume; the crystal unit without enclosure, connected to the oscillator being placed at the opposite part of the chamber. Thus, the correlation was realized between catalytic reactions taking place at the electrode surface and frequency deviations without amendment of the mechanism of the main ageing phenomenon because during hydrocarbon absorption radical similar forms appear on metals [7], during catalytic oxidation of which CO_2 and H_2O are formed. Figure 3 shows ageing curves for a crystal unit group with silver electrodes. At t°=1000 h the crystal units chosen for investigation having positive and negative frequency shifts were inserted in the chamber for the analysis of catalytic activity of electrodes with their enclosures removed. The crystal units with positive frequency deviations had electrodes in the catalysis area, and with the negative ones - beyond its limits. In the test chamber the following processes occured. In the first case the hydrocarbons of radical similar forms reacted actively with the electrode causing, in particular, silver atoms emission from the surface. The frequency was elevated during this process. In the second case the hydrocarbons being sorbed at the surface, were polymerized and the frequency of the crystal unit was lowered.

Conclusion

From the above-mentioned analysis it follows that on the manner of crystal unit ageing is influenced by the mass-transfer process from the surface of a piezoelectric element, which in its turn, is determined by the catalytic activity of electrodes.

References

1 Zhdan P.A., Lastushkina G.Ya., Khasin A.V. Oxygen adsorbtion on silver. Izvestiya AN USSR. Phys. series. 1979, v.43, No.9, pp.1802-1808. (In Russian).

2 Engelhardt H., Menzel D. Adsorbtion of oxygen on silver single crystal surfaces. Surface Science. 1976, v.57, pp.591-618.

3 Kilty R., Sachtler W. The mechanism of the selective oxidation of ethylene to ethylene oxide. Cat. Rev. 1974, v.10(1), pp.1-16.

4 Margonis L.Ya. Hydrocarbon oxidation on the surface of heterogene catalysers. Moscow, Chem., 1977, p.328. (In Russian).

5 Gutman E.I., Myasnikov I.A., Rodichev A.S., Aseev P.I. Cold emission of silver atoms from silver film evaporated on quartz surface. J.Phys.Chem., 1985, v.59. (In Russian).

6 Zavyalov S.A., Myasnikov I.A., Gutman E.I. Silver atoms emission from the solid-state surface. J.Phys.Chem., 1980, v.54, issue 2, pp.516-517. (In Russian).

7 Voge H., Adams C. Catalytic oxidation of olefins. Adv. on Catalysis. 1967, v.17, pp.151-218.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

BULK WAVE QUARTZ RESONATORS FABRICATED BY A HOLLOW CATHODE RF PLASMA ETCHING TECHNIQUE

STEFFEN SCHREITER
Technical University of Chemnitz, PSF 964, O-9010 Chemnitz,Germany

DIMITAR Y. YANKOV
Institute of Solid State Physics, Bulgarian Academy of Sciences, 72 Trakia Blvd., 1784 Sofia, Bulgaria

Abstract

Vibration modes in thin quartz membranes are considered. Assuming vibration modes to be highly trapped in the vicinity of the nearly spherical central region of the membranes, to the first approach the boundary condition at the periphery region are ignored.

A hollow cathode maskless plasma etching method for fabrication of thin quartz membranes is presented. A special geometric arrangement of electrodes and substrata allows the complete plasma structure (plasma sheath, bulk plasma) to be transferred to the substrata area during the etching process. The process has successfully been applied for preparing thin quartz membranes with plane-convex and plane-parallel shape and thicknesses of less than 5 µm.

The membranes are used for realization of bulk acoustic wave resonators at fundamental frequencies over 60 MHz and their electrical characteristics are presented.

1. Introduction

The fabrication of high frequency quartz resonators is limited by the level of technological equipment for preparation of thin quartz plates. By using modern mechanical methods for lapping and polishing of quartz wafers a thickness of 80 µm can be obtained, which corresponds to about 20 MHz fundamental resonance frequency. Lately this method has been considered to be on the cutting edge of its efficiency.

To overcome these problems plasma etching methods have been used to fabricate thin quartz plates. Etch gases are C-F - containing systems like CF_4, C_2F_6 or CHF_3 with different additives in the most cases. Common plasma processing where the substrata are freely supported on a wide rf-electrode (e.g. in [1]) allows only a full area thinning of the quartz plates. Very thin plates (several µm) can be etched in this way. However, these plates are unhandily for further treatment (evaporation of electrodes, positioning in the holders). A lithographic structuring of the quartz blanks fails since the mask material (photoresist, Al, Al_2O_3) cannot withstand the aggressive etch process (etch rates reach values up to 1 µm/min). To avoid these disadvantages a hollow cathode etching method is developed excluding lithographic pattering. Structuring is carried out by the confinement of the plasma onto the desired regions of the quartz blank using hollow cathodes with corresponding diameters. Power concentration and hollow cathode effect provide for high etch rates (several µm/min, trapping of electrons and therefore higher production of etch active species [8]). The quartz membranes etched with the hollow cathode technique show a high mechanical stability, the shape of the membranes can be simply varied over a wide range (plano-concave, plano-parallel, plano-convex) by changing processing pressure. For the fabrication and the study of high resonators we used etched membranes of the plano-convex type (Figs.

CH2965-2/91/0000-106 $1.00 © 1991 IEEE

1, 3d).

The vibration modes in plano-convex quartz resonators were analyzed and solutions for coupled thickness shear and thickness twist vibrations were obtained [2, 3]. In these analyses the change in the thickness, which results from the contouring, is small, compared to the thickness of the plate. Since the modes is highly trapped in the plate center, the edge was ignored, which simplified the analysis. The energy trapping effect is due to both contouring and electroded area impact. It is known that a given mode of vibration is characterized with a certain mechanically active region and the dimension of this region is determined by an equivalent resonant radius [4, 5]. In the case considered here if the dimension of the mechanically active region is smaller than the internal membrane diameter the influence of the peripheral region can be ignored too. A relationship similar to this for a flat plate is used to calculate the vibration mode frequencies of the membranes.

2. Vibration modes in thin membranes.

A schematic diagram of a membrane resonator is shown on Fig. 1.

As it is shown in [2] the fundamental vibrations in AT-cut quartz resonators are the coupled thickness shear (TS) and thickness twist (TT) ones. The modes of the coupled TS-TT vibrations in circular contoured plates, referred to the $X_1X_2X_3$ axes are described by

$$\hat{u}_1(x_1,x_2,x_3,t) =$$

$$= u(x_1,x_3)\sin(kx_2)\exp(i\omega t), \qquad (1)$$

where $k = n\pi/2h$ is the thickness wave number, $n = 1,3,5\ldots$ is the number of the harmonic overtone, $2h$ is the thickness of the crystal in the X_2 direction, and ω is the eigenfrequency of vibration. If u is a constant, (1) describes a plane shear standing wave, reflected from the two surfaces and represents the modes of an infinite (in the X_1 - X_3 plane) flat crystal. In the case of a finite crystal u is a function of x_1 and x_2. The various values of $u(x_1,x_3)$ possible for each n determine the anharmonic overtone modes, each characterized with a resonant frequency.

The inhomogeneous differential equation, governing coupled TS-TT vibrations is [2]:

$$M_r\frac{\partial^2 u_1}{\partial x_1^2} + c_{55}\frac{\partial^2 u_1}{\partial x_3^2} - \frac{n^2\pi^2}{4h^2}c_{66}u_1 - \rho\frac{\partial^2 u}{\partial t^2} =$$

$$= \rho\omega^2\frac{e_{26}Vx_2}{c_{66}2h}e^{i\omega t}, \qquad (2)$$

here ω is the frequency, c_{ijkl}, e_{ijk} are the elastic and piezoelectric constants, ρ is the density of quartz and M_r is a constant, which depends on the elastic constants of quartz and the harmonic number n [1, 2]. This equation can be generalized to describe a contoured plate simply by allowing h to be a slowly varying function of x_1 and x_3, i.e. $h = h(x_1,x_3)$ and consequently $k = k(x_1,x_3) = \pi/h(x_1,x_3)$. In our case the thickness of the plate is nearly constant in the central region (where $(x_1^2 + x_3^2)^{1/2} < d/2$) and is a quickly-varying function of x_1 and x_3 in the periphery region (where $(x_1^2 + x_3^2)^{1/2} > d/2$). Since the vibration modes are assumed to be highly trapped in the vicinity of the contoured plate center, to

the first approach the boundary condition at the periphery region can be ignored.

The resonant frequencies of a flat plate are given by:

$$f_n = n \frac{N}{2h_0}, \qquad (3)$$

where $N = \frac{1}{2}\sqrt{\hat{c}_{66}/\rho}$ is the frequency factor for an infinite flat plate.

The resonant frequencies of contoured piezoelement can be expressed by a relationship, similar to (3) as follows:

$$f_{nmp} = n \frac{N}{2h(\xi_{nmp})}, \qquad (4)$$

where $2h(\xi_{nmp})$ is the equivalent resonance thickness for any $(nmp)^{th}$ vibration mode. The radial coordinate ξ_{nmp}, defined as an equivalent resonance radius, on which the equivalent resonance thickness depends [3], is given by:

$$\xi_{nmp} = \frac{1}{n}\left(\frac{(2h_0)^3 d^2}{n^2 \cdot (R - \sqrt{R^2 - d^2/4})} \right)^{1/4} \cdot$$

$$\cdot \left(\sqrt{\hat{M}_{..}/\hat{c}_{66}}\,(2m+1) + \sqrt{\hat{c}_{55}/\hat{c}_{66}}\,(2p+1) \right)^{1/2},$$

and we can obtain for $2h(\xi_{nmp})$:

$$2h(\xi_{nmp}) = 2h_0 \cdot$$

$$\cdot \left[1 + \left(\frac{n}{2}\right)^2 \frac{(R - \sqrt{R^2 - d^2/4})}{2h_0 \cdot d^2/4}\, \xi_{nmp}^2 \right]^{-1/2} \cdot$$

(5)

A given mode of vibration is characterized with a certain mechanically active region of the piezoelement beyond which $u_{nmp} = 0$. The dimension of this region is [3]:

$$d^a = \min(2 \cdot 1.06\, \xi_{100},\ 2 \cdot 0.94\, \xi_{100}).$$

Here the central region of the membrane is almost spherical and in order to use (4) is sufficient that $d > d^a$.

3. Hollow cathode plasma etching method for preparation of thin quartz membranes.

The etching process was done in cylindrical glass vessel with a volume of 30 l. The electrode system is located in the center of the vessel. The ground electrode surround the rf-electrode as cylindrical shield. With some cm thick PTFE the electrodes are electrically separated each from another. The top of the rf-electrode contains the hollow cathode.

One hollow cathode arrangement is given in the scheme of Fig. 2. The real hollow cathodes are cylindrical holes in a graphite plate. Graphite was chosen as hollow cathode material since no redeposition effects occur in the $C\text{-}CF_4\text{-}SiO_2$ - system that may cause a considerable roughness on the etched surface. The dimensions of the hollow cathodes that have been used in the experiments are 5 mm in height and 5 mm in diameter. The edge (1.5 mm) of the quartz blank is covered by the graphite plate to confine the plasma onto the central part of the blank. In this way the edge of the sample remains at the initial thickness during the etch process and the etched membrane of several µm offers a sufficient mechanical stability. The quartz blank for itself rests on the rf-electrode (copper) which has a diameter of only 5 mm for power concentration. In this case isolation of the rf-electrode against unrequired power dissipation is made from

ceramics because of its heat resistance.

The rf-power (27.12 MHz) is applied to the reactor asymmetrically. The plasma facing area of the ground electrode is assumed to be large enough to allow a considerable self-bias-voltage with respect to the plasma potential to develop. To get a high self-bias-voltage a high electrode surface area ratio of mass electrode to rf-electrode is necessary [6]. The self-bias-voltage accounts for the acceleration of the positive ions that impinge on the surface of the quartz to be etched and is therefore responsible to the rate in the ion induced etch process of SiO_2 [7].

The AT-cut single crystal quartz blanks used in the experiments are about 100 µm thick and have diameter of 8 mm. The samples are polished on both sides.

Etch gas is CF_4 in all cases in a closed system. After 5 min processing the gas has been renewed. During a processing time of t<10 min no remarkable decrease of the etch rate has been observed.

The discharge cannot be concentrated completely into the hollow cathodes, a certain amount of power is dissipated in the common glow discharge existing in front of the graphite plate (Fig. 2). Thus, the rf-power we indicate in the paper is the total power input as a gross value.

Etch depth and profiles have been measured by means of optical microscopy focusing.

4. Experiments and results.

Without the graphite plate in Fig. 2 (non hollow cathode mode) one kind of profile shape can be etched (Fig. 3a). The profile is determined by the rf-electrode diameter. As lower the pressure as flatter the central etched region of the sample is. Convex profiles have not been yielded at all.

Figs. 3b-d demonstrate the development of profile shapes in the hollow cathode mode with increasing pressure. The complete overlapping of the glow lights (part of plasma sheath) is followed by a strong concave etch profile (Fig. 3b). A structure like that in Fig. 3c is obtained at a pressure, where the glow light regions in the plasma begin to divide. The diameter of the very flat area in the center of the structure (+/- 1 µm waviness) is of about one half of the diameter of the hollow cathode. Further increasing of the pressure values (p > 450 Pa) results in more and more convex etch structures (Fig. 3d). The physical background of the described etch profiles in dependence of operation pressure is the distribution of the particles in the plasma that take part in the etching process. An example of the radial density distribution of charge carriers in a hollow cathode discharge is given in Fig. 4 [8]. In other papers [9,10] similar radial density distributions of excited species in hollow cathode discharges are given. For that, most of the etch active particles (ions as well as radicals) have their origin in the glow light region of the plasma sheath that leads to the highest etch rates in the immediate neighborhood of the plasma sheath along the hollow cathode walls. The increase and decrease of the plasma sheath with altering pressure causes an according change in the etch profiles.

Silver electrodes with a diameter of 2 mm and a thickness of 800 Å are deposited on the central part of the membranes by vacuum evaporation. These resonance systems are fixed in holders for

M2P-standard miniature resonators.

Fig. 5 is the frequency spectrum of a resonator from the type in Fig. 3d.

Table 1 gives the theoretical predictions calculated by (4) and the experimentally observed resonance vibrations in the vicinity of the main resonant mode for the same membrane resonator.

Table 1. Frequency spectrum (measured and calculated in n = 1) for a membrane resonator with $2h_0$ = 21 µm, R = 430 mm, d = 3 mm, d^2 = 0.64 mm

n	m	p	f_{nmp}, [MHz]		ξ, [mm]
			Measured	Calculated	
1	0	0	62.96	62.98	0.302
1	0	2	63.52	63.54	0.502
1	2	0	63.78	63.68	0.543
1	0	4	-	64.09	0.643
1	2	2	64.26	64.23	0.675
1	4	0	64.36	64.38	0.706
1	2	4	64.78	64.78	0.785
1	4	2	64.92	64.92	0.812
1	4	4	65.50	65.46	0.906

In previous publications, where the vibration modes of circular contoured plates are investigated, only solutions of (2) with m, p = 0, 2, 4... are assumed to exist because of symmetry in X_1 and X_3 direction [2]. In most cases however, the real samples are not exactly symmetrical (e.g. for technological imperfections) and vibration modes with odd values of m and p appear in the frequency spectra. The frequencies of these modes can be calculated by (4) and can be observed in Fig. 5 (some of the modes interact and cannot be identified).

5. Conclusions

Hollow cathode rf plasma etching turns out to be an appropriate method for the fabrication of thin quartz membranes. The membranes are mechanically stabilized by a surrounding edge and can be etched in different profile shapes in dependence on the operation pressure. With CF_4 as feed gas etch rates of several µm/min are possible. The quartz membranes with plano-convex central profiles are the basic part of high frequency bulk wave resonators.

The frequencies of AT-cut membrane resonators can be predicted with sufficient accuracy by assuming that the energy of the vibrations is trapped in the nearly spherical central region of the membranes. A good agreement between theoretical and experimental values was observed. The results obtained show the complexity of the problems related to the design and fabrication of membrane resonators with good electrical and mechanical characteristics.

6. Acknowledgements

The authors would like to gratefully acknowledge Dr. H.-U. Poll from the Technical University Cemnitz, Germany and Dr. L. Spassov from the Institute of Solid State Physics of the Bulgarian Academy of Sciences, Bulgaria for their support in carrying out the experiments. One of the authors (D. Yankov) wishes to thank Dr. I. Avramov from the Institute of Solid State Physics of the Bulgarian Academy of Sciences, Bulgaria for the fruitful discussions.

7. References

1. G. C. Schwartz, L. B. Rothman, T J Schopen, "Competitive Mechanisms in Reactive Ion Etching in a CF_4-Plasma", J. Electrochem. Soc., Vol. 126, No. 3, p. 464. 1979
2. H. F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of

Coupled Thickness Shear and Thickness Twist", J. Acoust. Soc. Am., Vol. 59, No. 4, p. 879, 1976

3. H. F. Tiersten, R. C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtones of Coupled Thickness Shear and Thickness Twist", J. Acoust. Soc. Am., Vol. 65, No. 6, p. 1455, 1979
4. S. H. Slavov, "Equivalent Resonance Radius of Contoured AT-Cut Quartz Resonators", Appl. Phys. A43, No. 2, p. 111, 1987
5. S. H. Slavov, "Modes of Vibration, Motion Inductance, and Resonance Interval of Circular, Convex AT-Cut Bevelled Design Trapped Energy Quartz Resonators", Appl. Phys. A40, No. 1, p. 59, 1986
6. R. A. Gottscho et. al., "The Effect of Electrode Area ratio on Low-Frequency Glow Discharges", J. Appl. Phys., Vol. 66, No. 2, p. 492, 1989
7. H. F. Winters, J. W. Coburn, "Plasma Assisted Etching Mechanisms: The Implication of Reaction Probability and Halogen Coverage", J. Vac. Sci. Technol. B, Vol. 3, No. 5, p. 1376, 1985
8. F. Howorka, M. Pahl, "Experimentelle Bestimmung innerer und aeusserer Parameter des NGL-Plasmas einer zylindrischen Hohlkathode in Argon", Z. Naturforsch., No. 27a, p. 1425, 1972
9. H. W. Lehmann, R. Widmer, "Metal - Vapor Production by Sputtering in a Hollow Cathode Discharge: Theory and Experiment", J. Appl. Phys., Vol. 50, No. 9, p. 5694, 1979
10. E. M. van Veldhuizen, F. J. de Hoog, "Analysis of a Cu-Ne Hollow Cathode Glow Discharge at Intermediate Currents", J. Phys. D: Appl. Phys., No. 7, p. 953, 1984

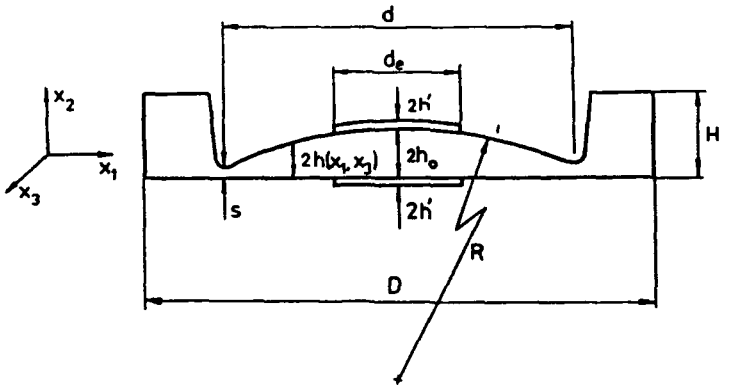


Fig. 1. Schematic diagram of a membrane resonator

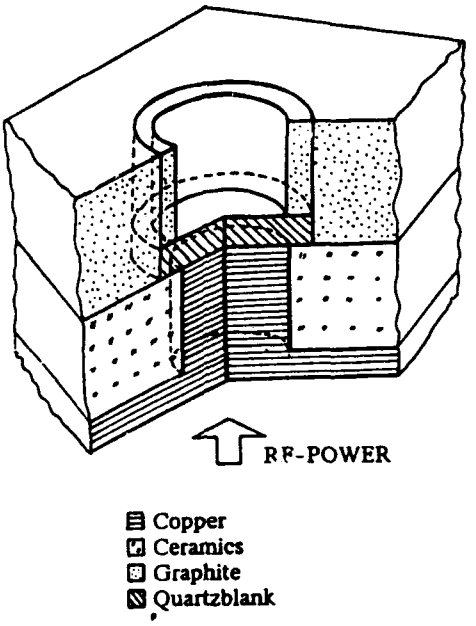
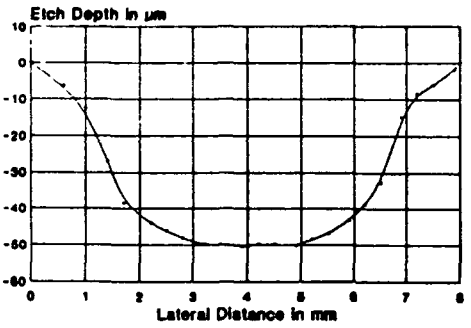
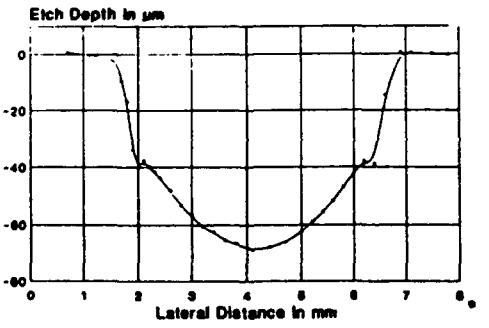


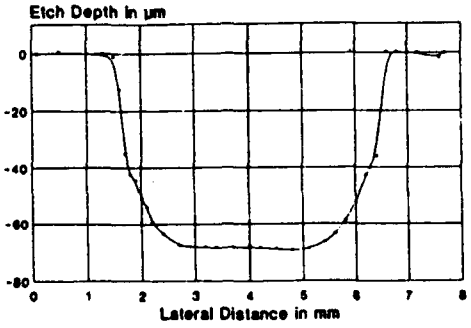
Fig. 2. Hollow cathode arrangement



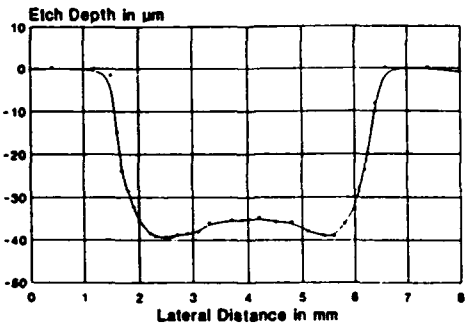
a) non hollow cathode mode,
p=600 Pa, P=200 W, Etch rate = 2.3 µm/min



b) hollow cathode mode,
p=133 Pa, P=230 W, Etch rate = 2.4 µm/min



c) hollow cathode mode,
p=332 Pa, P=200 W, Etch rate = 3.5 µm/min



d) hollow cathode mode,
p=612 Pa, P=200 W, Etch rate = 2.7 µm/min

Fig. 3 Etch profiles of quartz membranes

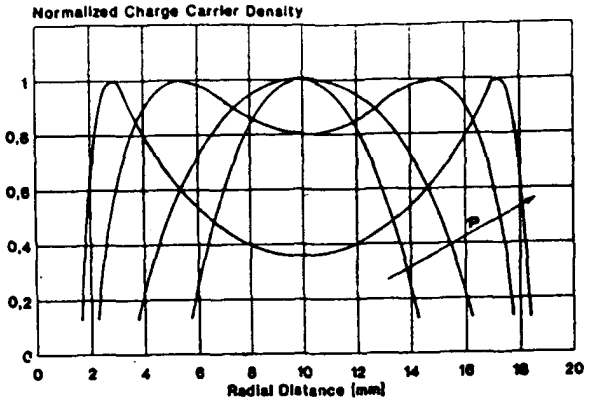


Fig. 4. Radial charge carrier density distribution (p=13, 26, 52 and 93 Pa) in Argon [8]

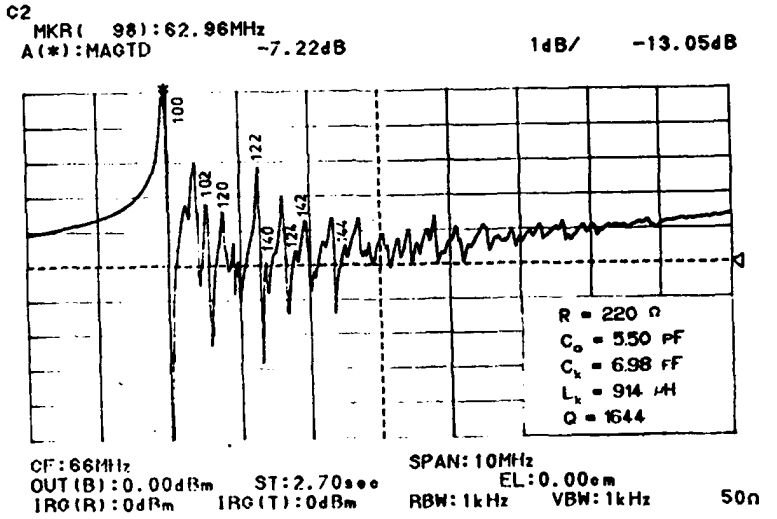


Fig. 5. Frequency spectrum of a resonator from the kind of Fig. 3d

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AUTOMATED X-RAY SORTING MACHINE FOR ROUND QUARTZ BLANKS

Hans Bradaczek

Freie Universität Berlin, Institut für Kristallographie

ABSTRACT

An X-ray machine is presented, which allows the sorting of quartz blanks at the beginning of the manufacturing process. The cycle time is about 5 sec's with an X-ray exposure time of 2 sec's. The standard deviation of this instrument is about 6'' of arc. Blanks can be sorted into 25 boxes, each of typically 15'' of arc. The range of the cutting angle is about $\pm$ 5° around the AT-cut. A PC-AT 386-25 computer controls measurements. Movements are controlled by a CNC device. High sophisticated statistic programs allow a permanent observation of the sorting process.

INTRODUCTION

Quartz crystals are mainly used as stabilizing elements in oscillator circuits. To reduce drift of resonant frequency with varying temperature the blanks are cut from the raw Quartz block with a distinct cutting angle between their surface and a selected lattice plane as e.g. the widely used AT-cut. To minimize temperature dependence this angle has to be kept in close tolerance - almost down to a few seconds of arc.

One of the useful methods for accurate measurement of the cutting angle applied for decades of Quartz technique is X-ray diffraction. Two major disadvantages of this method - long exposure times and complicated adjustment - made it unsuited for high production output of modern oscillator manufacturing. Thus it had been restricted to some basic investigations or to spot-checks in the manufacturing process, although a continuous check of the raw blanks could increase yields drastically.

Our newly developed AT-cut sorting machine allows for cutting angle determination with standard deviations of less than 6'' of arc at cycle times of about 5 sec (including 2 sec X-ray exposure time).

METHODS

The diffraction of X-rays by a certain lattice plane follows Bragg's law

$$\sin \Theta_B = n\lambda/2d$$

Θ_B = Bragg angle, λ = wavelength of X-rays, n = order of reflections, d = lattice plane distance.

The Bragg angle is about 27 degrees for a quartz lattice at the $02\bar{2}2$ lattice plane and Cu radiation.

In a setup with X-ray source and detector fixed in a suited "reflecting" position (Fig. 1), rotating the blank around an axis vertical to its surface will show reflections at two different angular positions (Fig. 2).

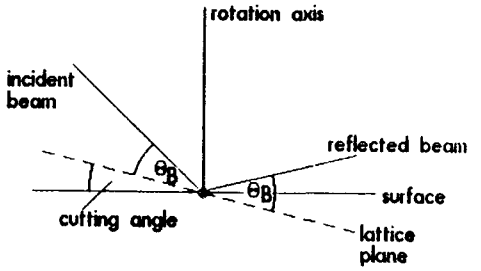


Fig. 1: Diagram of X-ray reflection

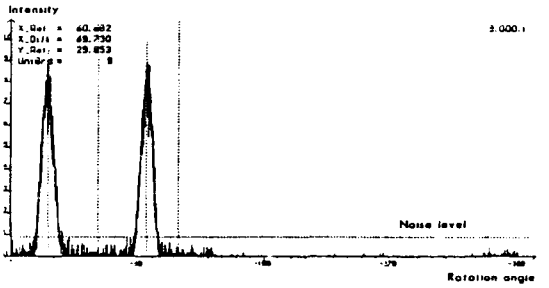


Fig. 2: X-ray intensity as a function of the crystal rotation

The cutting angle is related to the angular distance of these two reflections and can be calculated from it. For an AT-cut and the $02\bar{2}2$ lattice plane the transfer ratio is about 1 : 60, which means: to get a cutting angle resolution of about 6'' of arc the angular distance of the reflections must be measured with an accuracy of about 0.1°.

APPARATUS*)

The main parts of the instrument (X-ray source,

*) Manufactured by EFG GmbH, Berlin

CH2965-2/91/0000-114 $1.00 © 1991 IEEE

detector and rotating blank holder) are mounted on a granit table for stability. The blank holder rotates around a vertical axis at a speed of 0.5/ sec.

X-ray tube, blank and detector are positioned following Bragg's law (Fig. 3).



Fig. 3: Rotating blank holder, feeding arm (right), distribution arm (left), X-ray tube collimator (left), detector collimator (right).

The position of the X-ray tube with a 0.5 mm collimator determines the direction of the incident beam; the accuracy of its angle adjustment has to be within 3'' of arc. The position can be changed in a range of about 5^{o} to cover applications for other cutting angles.

During the exposure the blank is mounted on a 3 pin support and fixed to it by underpressure. The support needs to have a high precision, because tenth of a micron dislocation leads to significant mismeasurements of the cutting angle.

Even small deviations between the vertical of the blank and the rotation axis (eccentricity) result in an noticeable increase of the errors in measurement. So mechanical and computer corrections are necessary. Handling of the blanks starts with picking up the blanks from a tube magazine which contains up to 1000 blanks to be sorted. The feeding arm transports the blanks into the measuring position. The measuring time of 2 sec's is devided in 4000 intervalls, in which the X-ray data are collected and stored to an external buffer, from which the data are read from the control computer in a single block transfer. After the 2 sec's exposure the distribution arm moves the blanks to one of the 25 boxes (Fig. 4).

For special investigations single selected blanks can be measured too.

Data evaluation is done by a PC-AT 386-25 while a separate CNC unit is used for process control.

The whole machine is covered by a plexiglas box for X-ray and dust protection as well as for temperature control (stability about $1^{o}C$) and is mounted on a rack containing the X-ray generator and the CNC control device (Fig. 5).

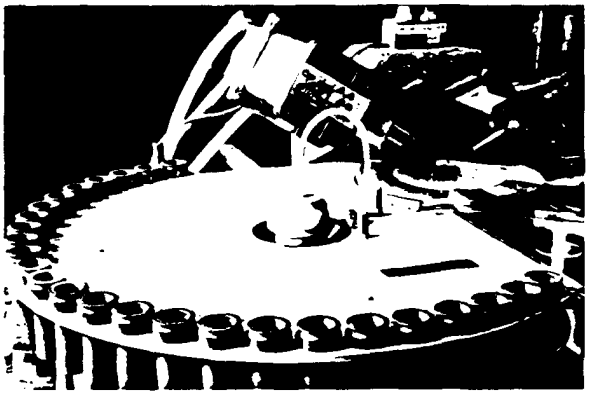


Fig. 4: X-ray tube and detector, rotating blank holder, sorting magazine for 25 boxes, distribution arm.

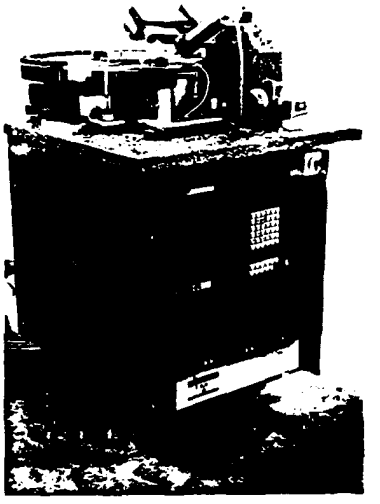


Fig. 5: Complete sorting machine without plexiglas cover.

RESULTS

To check the stability and resolution of the apparatus the same blank was measured in a fixed position on the 3 pin support 1000 times (Fig. 6).

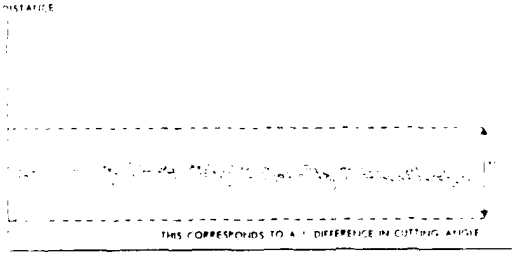


Fig. 6: Peak distance values from 1000 measurements of the same blank.

The reproducibility of the measurement is given by the standard deviation of 6'' of arc (Fig. 7).

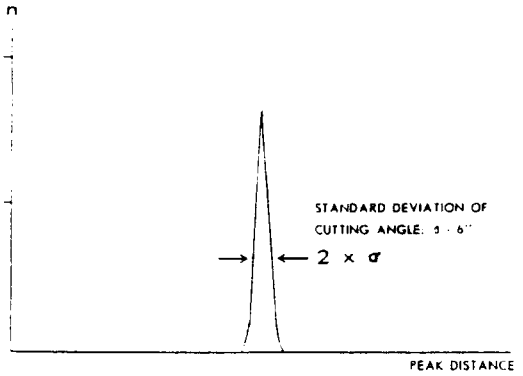


Fig. 7: Peak distance distribution for 1000 measurements of the same blank.

Additional sources of error are: eccentricity, non-parallelism of blanks, surface roughness, dust e.a.

Fig. 8 shows the results of sorting 600 blanks into 25 boxes.

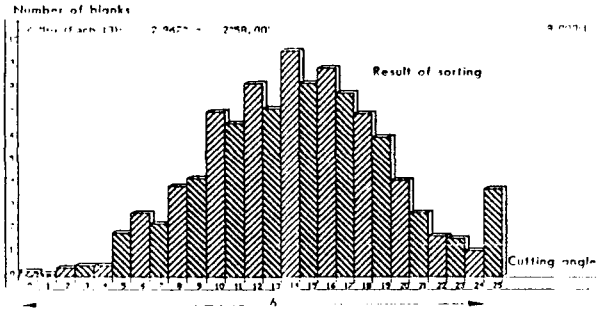


Fig. 8: Sorting result of 600 blanks (sorting time less than 1 h)

Resorting of the blanks from one box leads to the result shown in Fig. 9.

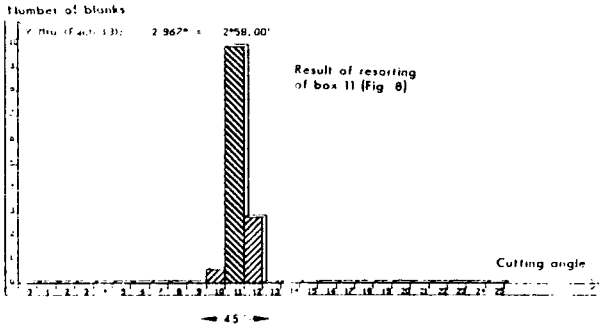


Fig. 9: Resorting of box 11 from Fig. 8.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A VERY LOW PROFILE, SHOCK-RESISTANT, SURFACE-MOUNTABLE QUARTZ CRYSTAL UNIT[1]

R. C. SMYTHE, Piezo Technology Inc.
Box 547859, Orlando, FL 32854

J. GREENSTEIN, Coleman Research Corp.
5950 Lakehurst Dr., Orlando, FL 32819

J. LIPELES, International Dynamics Corp.
316 Spring Run Circle, Longwood, FL 32779

Abstract

A very low profile, surface-mountable quartz crystal unit has been developed capable of withstanding shock levels of 100,000 G's or more. The new resonator is equally well-suited for very high shock applications, such as smart projectiles and air-dropped surveillance devices, and surface-mount applications where an extremely low profile and small footprint are required. An AT-cut, 33 MHz, 3rd overtone unit has overall dimensions of .31 inch (7.9 mm) diameter and .05 inch (1.3 mm) maximum height. The unit consists of a top cover, a resonator plate, and a bottom cover. All three pieces are made of identically-oriented, polished and deep-etched, swept, premium-Q quartz. A finite-element model of the three-piece sandwich, constructed using both orthotropic and general-purpose anisotropic elements, allowed the unit to be designed to the high G level. To obtain the necessary material properties, measurements were made of the breaking strength of samples of the quartz used for a range of surface treatments. In addition, bond shear strength and bond tensile strength were measured.

Introduction

Future development of smart weapons and munitions requires the availability of acceleration-hardened electronics. In many electronic systems, crystal resonators, whether used in oscillators or in filters, are the most fragile components.

Conventional crystal resonators can withstand shock levels from a few tens of G's to perhaps a few thousand G's. A major development in the acceleration-hardening of quartz resonators was the introduction of deep etching [1-3], which greatly increases the breaking strength of AT-cut quartz. Illustrating this, deep-etched quartz resonators which can withstand as much as 33,000 G's were developed some years ago [4].

One factor limiting the robustness of conventional crystal units is that the crystal plate is supported by two to four small metal ribbons or wires which are bonded to the edges of the plate. For the crystal units reported here, the resonator plate is sandwiched between two, much thicker, cover plates, making an extremely stiff three-piece assembly. Our design objective was a resonator which would withstand 100,000 G's. Finite-element stress analysis indicates that it should withstand in excess of 500,000 G's if rigidly supported.

Description

The crystal unit developed is an AT cut operating at a third overtone frequency of 33 MHz. Typical electrical parameters are Q = 150,000, C_1 = 0.56 fF, R_1 = 55 ohms, and C_o = 1.5 pF. Overall dimension are 0.310 inch (7.9 mm) diameter, 0.05 inch (1.27 mm) maximum height. Figure 1 shows a cutaway view. It consists of a resonator plate, approximately 0.006 inch (0.15 mm) thick, and top and bottom cover plates, 0.019 inch (0.48 mm) thick. The inner face of each cover has a shallow, circular recess approximately 0.001 inch (0.025 mm) deep. The three pieces are bonded together using 0.001 inch (0.025 mm) thick, annular epoxy pre-forms. The completed unit is intended to be mounted by bonding one face to a suitably stiff substrate. Electrical connections are made

[1] This work was sponsored by SDIO and monitored by USAF Armament Lab, Contr. No. FO8635-90-C-0101.

CH2965-2/91/0000-117 $1.00 © 1991 IEEE

to two metallized terminal regions, not visible in figure 1, on the rim of the bottom cover plate.

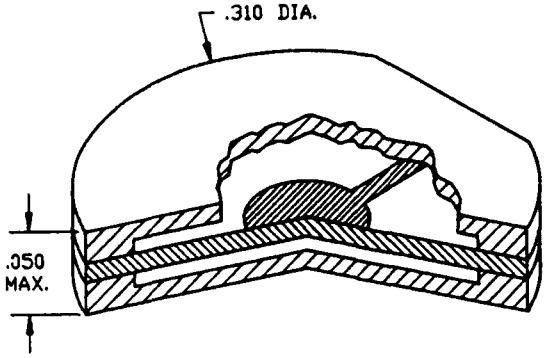


Fig. 1. Cutaway view of the low-profile, surface-mount resonator. Not visible are metallized connection pads on rim of bottom cover.

The resonator plate and cover plates are made from swept, premium-Q quartz, and are polished and then chemically etched for maximum strength. As nearly as possible, the crystallographic orientations of the three pieces are identical. The resonator electrodes are vacuum-evaporated gold, with a chrome adhesion layer.

Materials Testing

Measurements of the breaking stress of crystalline quartz have been reported by Bechmann [5], Vig, et al [1], and Chao and Parker [6]. The measurements by Vig, et al were for thin AT-cut plates and indicated that deep etching ("chemical polishing") greatly increased the breaking stress. As a first task in the structural design of the crystal unit, measurements were made of the breaking stress of resonator and cover plates for a range of surface treatments, to supplement the available data. A finite-element model, constructed using both orthotropic and general-purpose anisotropic elements, allowed the measured displacement vs force data and breaking force data to be converted to equivalent stiffness and breaking strength. The model also allowed the average breaking stress to be calculated from Vig's data.

Testing was carried out by the Materials Testing Laboratory of the University of Florida College of Engineering, Gainesville and is reported in detail elsewhere [7]. Each specimen was simply supported by a steel ring and centrally loaded by means of a steel ball. The specimens were loaded to failure with an MTS hydraulic test machine, while recording both load and deflection. Tests were made of unswept, electronic grade quartz and swept, premium-Q quartz, both from the same supplier. Samples were mechanically polished prior to etch. The amount of etch ranged from none to 33.2 micrometers (16.6 micrometers per side, or $\Delta f = 2Of_1f_2$.) Etching was in a production system using saturated ammonium bifluoride at a temperature of 50° ± 3° C, with vertical agitation. Typically, 10 to 12 samples were measured for each combination of material and surface treatment. Weibull curves were fitted to the data to estimate the cumulative probability of failure.

Figure 2 shows Weibull plots for mechanically polished, swept, premium-Q resonator plates. The plates were 7.87 mm (0.310 inch) in diameter with a thickness prior to etch of 166 ± 1 micrometers. The upper curve is for unetched plates; the lower, for plates etched 16.6 micrometers. It will be seen that etching has increased both the mean breaking strength and the variability of the breaking strength. This is typical of all results obtained. This increasing variability is believed to result from the increasing density of etch pits and, in some cases, the deepening of scratches, with increased etch. For the plates with no etch, the 50% failure level is 1.35 x 10^5 lb/in^2, while 16.6 micrometers etch gave a 50% failure level of 2.07 x 10^5 lb/in^2, which is very nearly the mean breaking strength of 2.12 x 10^5 lb/in^2.

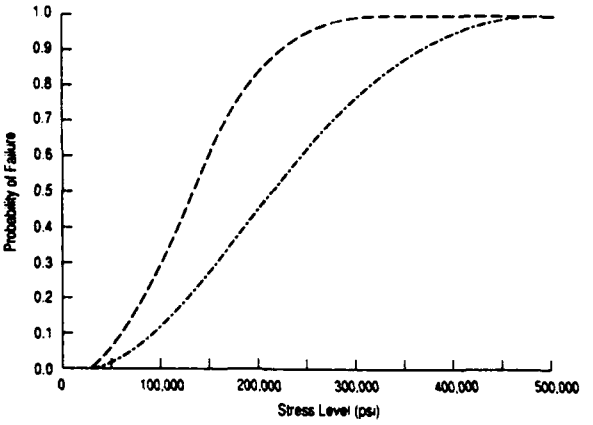


Fig. 2. Weibull plot of cumulative probability of failure as a function of tensile stress for polished resonator plates having no etch (left-hand curve) and 16.6 micrometers chemical etch (right-hand curve).

These data may be compared with the average breaking strength calculated from measurements by Vig, et al [1]. Their tests also used a steel ball to apply an axial force to an AT-cut plate simply supported near its outer edge. They tested twelve 20 MHz AT-cut plates which had been finish-lapped in 3 micrometer abrasive and then etched $15 \cdot f_1 f_2$ (25 micrometers), and found an average breaking load of 2.2 kg. Using the same method of analysis as for the measurements made during the present program, we calculate an average breaking stress of 2.28×10^5 lb/in^2, which is remarkably close to our mean breaking strength for 16.6 micrometers etch.

A Weibull plot of the breaking strength of premium-Q, swept quartz covers, etched 33.2 micrometers and with a thickness before etch of 500 ± 3 micrometers, is shown in figure 3. The 50% failure level is 1.06×10^5 lb/in2, about half the level of the etched resonator plates. The data seem to indicate that the failure stress for properly etched thin plates is greater than for thicker plates. The reason for this is not understood.

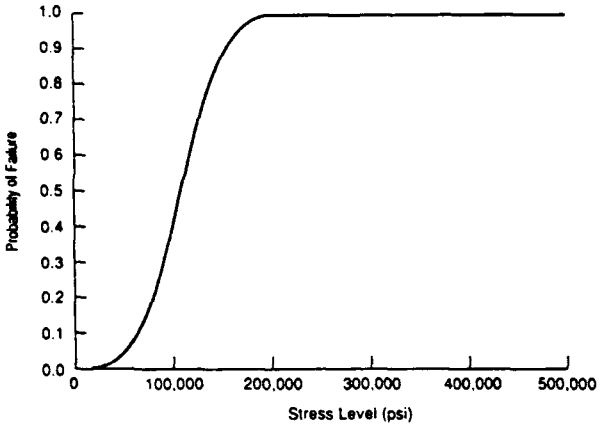


Fig. 3. Weibull plot of cumulative probability of failure as a function of tensile stress for polished cover plates having 33.2 micrometers chemical etch.

In addition, tests were made of tensile and shear strength of two candidate bond materials, a low-outgassing liquid epoxy and an epoxy pre-form -- four test groups in all. Observed failure loads in each group varied widely from sample to sample, due primarily to inadequate control of the bonding process. This was in part related to the design of the samples, and in part to the fixturing used. For the tensile test samples the average strength of the four best pre-form bonds was 2628 lb/in^2, while the best liquid epoxy bond failed at 2472 lb/in^2. In the shear tests of the pre-form bonds, results were inconclusive due to failure of auxiliary bonds between the quartz test plates and aluminum support blocks; the best observed bond strength was 1057 lb/in^2. For the liquid epoxy samples, the best observed bond shear strength was 634 lb/in^2.

While further measurements of bond strength are needed, some tentative conclusions can be drawn. First, for the procedures used, the pre-forms were markedly superior to the liquid epoxy. This may mean that the pre-forms are easier to use, rather than that they have greater inherent bond strength. Second, control of the bonding process is crucial. It is believed that this will be easier to achieve for the actual resonator bonds than it was for the test samples. A useful feature of the resonator configuration is that the bonds can readily be visually inspected. To obtain further assurance of bond strength, a simple shear test fixture was devised. Using this fixture, as part of the crystal unit fabrication process, each resonator bond was subjected to a static shear load of 20 lb, which is equivalent to a shear stress of 510 lb/in^2. There were no failures, even for bonds which exhibited significant voids based upon optical microscope examination. Since the previous tests indicated that bond tensile strength is greater than bond shear strength, shear testing should be a simple and useful means of assuring adequate bond strength.

Structural Analysis

Using NASTRAN, a finite element stress analysis of the complete resonator structure has been performed for static acceleration loading normal to the major faces. Orthotropic plate elements, the most general available with NASTRAN, were used. Taking advantage of symmetry, a one-quarter sector model was used to reduce the number of elements required. Because there is significant deformation at high acceleration loads, the stiffness formulation of the model must be able to change as a function of the applied load. To provide stability to the solution algorithm, the load was progressively incremented up to the maximum.

The original goal was to design a crystal unit which could survive at least 100,000 G's. Because the stress levels at 100,000 G's were quite modest, the analysis of the crystal assembly was continued for accelerations up to 1,000,000 G. For very large accelerations normal to the major faces, the resonator plate may deform sufficiently to contact one cover,

thereby limiting further deformation of the resonator plate and transferring a portion of the acceleration load to the more rigid cover. (At this point, of course, the resonator will be non-operational.) The acceleration level at which this load sharing initially occurs depends critically upon the spacing between the resonator and the inner face of the cover. This spacing is the sum of the bond thickness and the cover recess depth. For the dimensions used, load sharing commences slightly above 600,000 G. Table 1 gives the resonator peak effective stresses (Von Mises stresses) in the resonator plate and cover plates as a function of acceleration. For these calculations, it is assumed that the bottom cover is simply supported around its outer edge. These calculations show that, with proof testing the quartz elements can survive acceleration levels of 1,000,000 G's.

Table 1. Peak Effective Stresses as a Function of Normal Acceleration (lb/in^2)

Acceleration (G)	Resonator	Top Cover	Bottom Cover
100,000	10230	7285	7379
600,000	59800	42431	45586
800,000	73137	58443	61332
1,000,000	82447	75568	77294

Bond stresses were also calculated, Table 2, both for edge support and uniform support at the bottom face. These calculations show that the bond stress is critically dependent upon the mounting conditions, with the two calculations representing worst and best cases. These stresses are essentially linearly proportional to acceleration; therefore, clearly, the strength of the crystal unit will be determined by the bond strength, not the strength of the quartz elements, and by the rigidity of the support.

Table 2. Peak Tensile Bond Stress (lb/in^2) 100,000 G's Normal Acceleration

	Upper Bond	Lower Bond
Edge Support	1916	2140
Uniform Support	350	460

Electrical Design and Performance

In order to ensure high Q, Tiersten's three-dimensional trapping analysis [8] was used to design so that the third overtone mode amplitude is essentially zero at the inner circumference of the bond region. At the same time, electrode parameters were selected so that the fundamental mode Q is low. The purpose of the latter was to simplify the design of the oscillator sustaining stage in the intended application by reducing the need for L-C frequency selectivity.

Measurement of a number of units verifies that both objectives have been met. Table 3 shows the measured fundamental mode, third overtone mode, and fifth overtone mode parameters of a representative unit. All measurements were made using a computer-controlled Hewlett-Packard 4191A r-f impedance analyzer and PTI software [9]. For eight units measured, the fundamental mode Q ranged from 3000 to 18000, and the resistance from 161 ohms to 847 ohms, while the third overtone resistance was less than 50 ohms for 7 out of 8 units. Although the unit was designed for third-overtone operation, fifth overtone performance is excellent. For the same eight units, the fifth overtone Q ranged from 133,000 to 194,000, and the resistance from 78 ohms to 118 ohms.

Since not all expected applications could be specifically addressed, one very important requirement, resonators for microprocessor clocks and other digital timing applications was selected. A generic microprocessor clock oscillator application was targeted, with a nominal frequency of 33 MHz. Ten units were delivered, having the electrical equivalent circuit parameters given in Table 4. Q ranged from 131,000 to 228,000, which is more than adequate, while motional resistance ranged from 38 to 65 ohms. Motional capacitance and shunt capacitance values were quite uniform, which is also desirable.

Figure 4 shows a representative plot of frequency and resistance versus temperature. The data show good agreement with the expected f-T behavior and an absence of significant coupled modes over the -40° to +100° C temperature range. Figure 5 shows aging measurements of four resonators. The data show considerable scatter, which may be due in large part to the test method and test fixture. The units were stored, non-operating at 85° C. Frequency measurements were made at room temperature, which varied over about a 4° C range, with the units being inserted one by one into the measurement fixture. Due to the fixture,

Table 3. Equivalent Circuit Parameters as a Function of Overtone

OT	Frequency (kHz)	Motional Resistance R_1 (ohms)	Motional Capacitance C_1 (fF)	Shunt Capacitance C_o (pF)	$Q/10^3$
Fund	11033.72	444.3	4.87	1.48	7
3rd	32987.48	40.0	0.55	1.40	218
5th	54954.02	85.4	0.18	1.40	187

Table 4. Equivalent Circuit Parameters, Final Units

Unit No.	Motional Resistance R_1 (ohms)	Motional *Capacitance* C_1 (fF)	Shunt *Capacitance* C_o (pF)	$Q/10^3$
U-1	39.1	0.57	1.45	215
U-2	43.4	0.57	1.44	195
U-3	38.4	0.55	1.42	228
U-4	55.5	0.56	1.45	155
U-5	56.4	0.57	1.44	150
U-6	55.3	0.55	1.42	157
U-7	63.5	0.55	1.38	139
U-8	65.0	0.57	1.46	131
U-9	45.5	0.55	1.40	192
U-10	52.1	0.57	1.48	162

repeatability of the frequency measurement was no better than ±1 ppm.

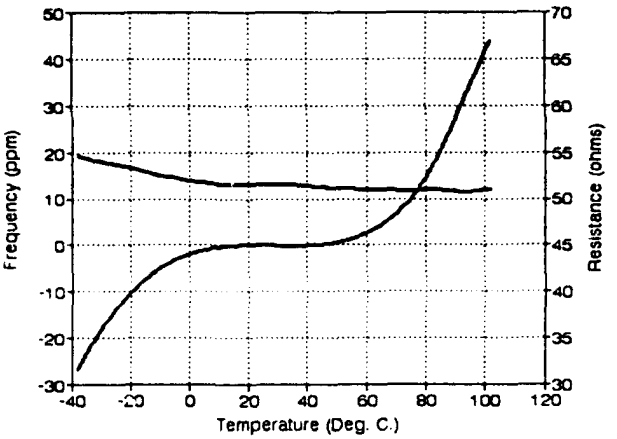


Fig. 4. Frequency and resistance versus temperature for a representative crystal unit; the curves represent approximately 400 data points.

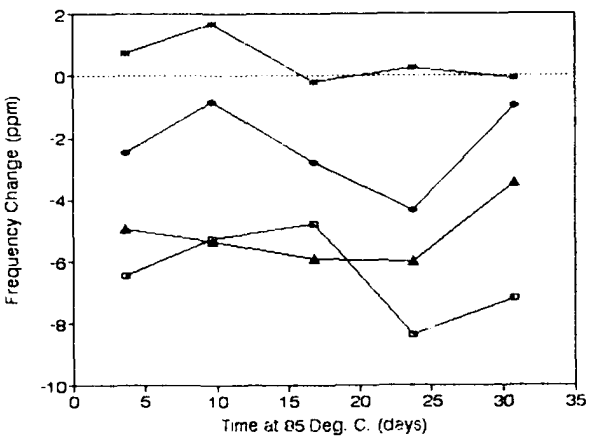


Fig. 5. Frequency aging of four crystal units, aged at 85° C, non-operating, and measured at room temperature.

Conclusion

The work performed has demonstrated the feasibility of the high-shock resonator configuration from the standpoint of electrical performance and manufacturability. Although the development was focussed on a particular frequency, the configuration has potential applicability over almost the entire frequency range of AT-cut crystals -- very conservatively, 10 to 100 MHz. Also, other thickness-shear cuts can be fabricated in this form, in particular the SC-cut.

Although the configuration has not yet been tested to shock levels as high as the initial target of 100,000 G's, calculations of the acceleration-induced stress, together with measurements of the breaking strength of the resonator and cover plates show that they are capable of surviving much higher levels. The most critical elements of the configuration are the bonds. For acceleration normal to the major faces of the unit, the bond stress has been shown to vary by a factor of nearly 5, depending upon the manner in which the unit is supported. The strength of the device, as designed, is therefore between 100,000 and 600,000 G's.

A most important feature of the new crystal unit configuration is that, in addition to being acceleration-hardened, it is an extremely low-profile, surface-mountable device. Its present height, .05 inch (1.3 mm) maximum, may be less than that of any other available crystal unit.

References

[1] J.R. Vig, J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," Proc. 31st Annual Symposium on Frequency Control, pp. 131-143, 1977.

[2] J.R. Vig, R.J. Brandmayr, and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd Annual Symposium on Frequency Control, pp. 351-358, 1979.

[3] J.R. Vig and R.J. Brandmayr, U. S. Pat. 4,274,907, June, 1981.

[4] R.L. Filler, L.J. Keres, T.M. Snowden, and J.R. Vig, "Ceramic Flatpack Enclosed AT- and SC-Cut Resonators," Proc. IEEE Ultrasonics Symp., pp. 819-824, 1980.

[5] R. Bechmann and P.L. Parsons, "Breaking Strength of Various Piezoelectric Crystals and Maximum Safe Current," Rpt. No. 10 in Piezoelectricity, P.O. Research Station, London, pp. 321-328, 1957.

[6] R.L. Chao and T.E. Parker, "Tensile Fracture Strength of ST Cut Quartz," Proc. 37th Annual Symposium on Frequency Control, pp. 116-124, 1983.

[7] J. Greenstein, "Hardened Electronics for EML Projectiles, Phase II SBIR, Final Technical Report," Contr. No. F08635-90-C-0101, 1991.

[8] H.F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," J. Acoust. Soc. Am., v. 59, pp. 879-888, 1976.

[9] R.C. Smythe, "An Automated Resonator Measurement System Using a Reflection Coefficient Bridge." Proc. 35th Annual Symposium on Frequency Control, pp. 73-77, 1981.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SUPPRESSION OF THE UNWANTED RESPONSE OF AN SC-CUT CRYSTAL UNIT

M. KOYAMA, S. MIZUNO, M. OKAZAKI

Nihon Dempa Kogyo Co., Ltd.
1275-2, Kamihirose, Sayama,
Saitama 350-13, JAPAN

SUMMARY

We have conducted a researches on the radius of curvature of plano-convex crystal wafers and beveling operation for the purpose of suppressing the unwanted response of the 3rd overtone 5 MHz SC-cut crystal units to be used in an OCXO.

Our study revealed that the series resistance of the unwanted response became 5 times the series resistance of the main response of an SC-cut crystal unit under the conditions that the radius of curvature on plano-convex is 140 mmR and the beveling is conducted asymmetrically.

An SC-cut crystal unit whose unwanted responses are suppressed have superior frequency temperature characteristics, free from activity dips. Measurement of the equivalent circuit parameter revealed that the Q value 2.14×10^6, the capacitance ratio 1.78×10^4, the equivalent resistance 80 ohm and the motional inductance is 5.5 H.

The OCXO using an SC-cut crystal unit whose unwanted responses are suppressed, has the phase noise of -105 dBc/Hz in terms of 1 Hz off-set frequency from carrier. thermal hysteresis and orientation effect of frequency are in oredr of 10^{-10}, while the aging rate is 2×10^{-9}.

INTRODUCTION

Use of an oscillator generating highly stable frequency is requested strongly to realize more effective use of limited radio frequencies . The OCXO (Oven Controlled Crystal Oscillator) is an efective means, for time being, to meet such requirement as a time and frequency substandard source. With the trend toward digitalization of communication systems, further improvement for phase noise characteristics has been required.

Use of OCXOs has an advantage in improvement of phase noise characteristics. The current requirement with respect to the phase noise reaches to about -80 dBc/Hz at 1 Hz off-set frequency from the carrier, desirably to -90 dBc/Hz to -105 dBc/Hz as shown in Figure 1.

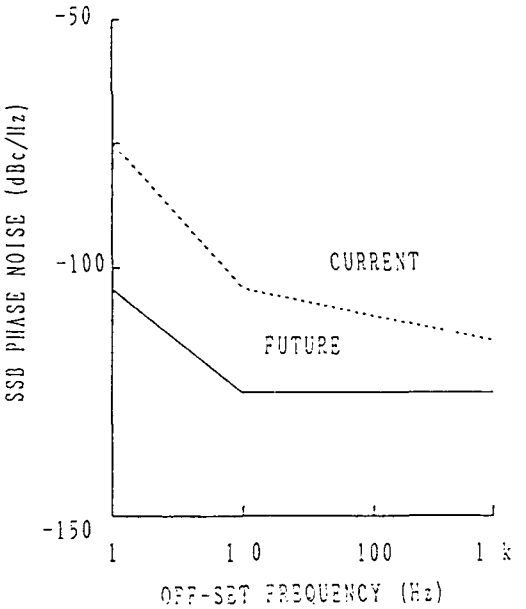


Figure 1 Phase Noise Requirement for OCXO

The crystal units built in OCXOs being used to be AT-cut crystal units, SC-cut crystal units can be performed more superior to AT-cut crystal units, so that many researches have been conducted on SC-cut crystal units exten-

CH2965-2/91/0000-124 $1.00 © 1991 IEEE

sively. However, an SC-cut crystal unit having strong unwanted responses as shown in Figure 2, these unwanted responses have to be suppressed effectiveilly in case of using it in OCXOs. Many an efforts have been madeto try to suppress the unwanted responses on SC-cut crystal units.

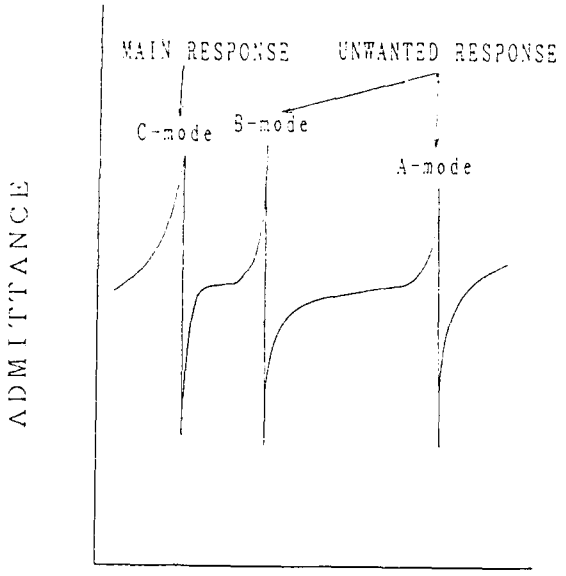


Figure 2 Unwanted response of SC-cut Crystal Resonator

In this paper , the suppression technique of the unwanted responses is discussed for the purpose of practcal use and then other items such as phase noise, orientation effect of frequency and thermal hysteresis are touched.

SYNTHETIC QUARTZ CRYSTALS

Due to the fact that OCXOs using SC-cut crystal unit require the stability of less than 1×10^{-9}, synthetic quartz crystals should meet severe requirements in terms of purity and Q value. In other words, we used the synthetic quartz crystals which have less than 12 impurities pieces per cm^3 and whose Q value is more than 2.4×10^6 when calculated from infrared absorption rate. For these synthetic quartz crystals, sweeping operation has not been performed.

RESONATOR FABRICATION PARAMETERS

Design parameters for the prototype SC-cut crystal unit were held constant, as follows: the frequency was 5 MHz on the 3rd overtone mode, the holder was HC-36/U on generally used for OCXOs. For the other specifications, refer to Table 1.

TABLE 1 : SPECIMEN OF SC-CUT CRYSTAL UNIT

FREQUENCY	5MHz
MODE OF VIBRATION	3RD OVERTONE
HOLDER	HC-36/U
DIAMETER OF WEFER	φ14.0mm
ELECTRODE	Cr+Au
DIAMETER OF ELECTRODE	φ7.0mm

SUPPRESSION OF UNWANTED RESPONSE

In general,SC-cut crystal unit have 3 types of thickness mode near the main response.[1],[2] These modes are classified into C-mode (thickness - shear mode), B-mode (thickness-twist mode) and A-mode (extensional mode). The C-mode corresponds to the main response to be obtained. On the other hand, the A-mode and the B-mode correspond to the unwanted responses to be suppressed. In not a few cases, the equivalent resistance for the B-mode is smaller than that for the C-mode. In many cases, connecting such a crystal unit to an oscillation circuit and operating it causes the oscillation in B-mode. That is why the crucial step for putting SC-cut crystal units in practical use is the suppression of the equivalent resistance for the B-mode not that for the C-mode.

According to R.L. Filler, et.al.[3] processing a crystal wafer in the form of plano-convex allows the B-mode to be suppressed. For this method, the suppression of B-mode equivalent resistance (Rb) is not sufficient as the C-mode equivalent resistance (Rc) is about twice as much as Rc.Then, we measured the frequency for each unwanted response while changing the radius of curvature on plano-convex step by step. Based on these measurement results, 8 kinds of radiuses of curvature were selected over the range of 50 mmR to 900 mmR for making resonators. For each resonator, Rc and Rb were measured by means of a network analyzer for measuring the Rb suppression effect. In addition, the suppression effect of the beveling upon Rb was measured.

EFFECT OF CURVATURE

Figure 3 shows the measurement results: Figure 3 indicates the relation between the radius of curvature on plano-convex and the ratio of Rb to Rc.

On the other hand, the dimensional tolerance of the processing jig for plano-convex processing was ± 15 persent. As the ratio of Rb to Rc became greatest when the radius of curvature on plano-convex fell between 140 mmR to 200 mmR, this value was specified as 140 mmR. But, in this case, our measurement revealed that Rc is no more than about twice as much as Rb and the B-mode suppression effect was insufficient.

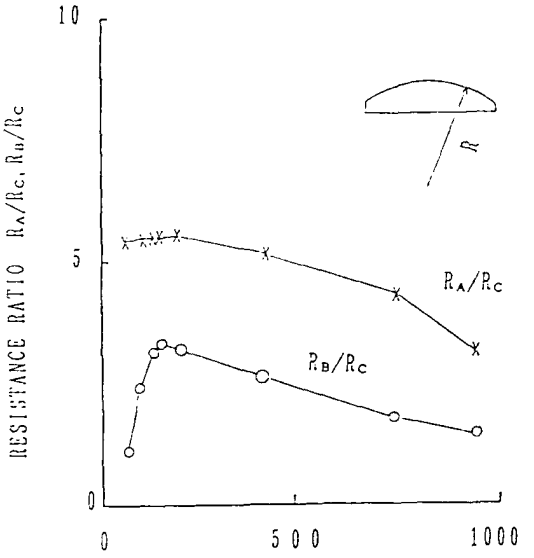


Figure 3 Effect of Radius of Curverture on Equivalent Resistance Ratio

EFFECT OF BEVELING

We conducted the beveling operation on both sides of a plano-convex-shaped crystal wafer to facilitate its supporting prior to making a crystal. In order to minimize the number of steps for the beveling, the internal beveling diameters of the crystal wafer were asymmetrically specified for both surfaces. Then, 7 kinds of beveling internal-diameter ratios for both surfaces were selected over the range of 0.8 to 1.4. Figure 4 shows the results for the crystal.

Figure 4 shows the equivalent resistance for each of the C-mode and the B-mode measured for each crystal and the relation between D2/D1 (D1 : Internal diameter of the beveling surface 1, D2: Internal diameter of the beveling surface 2) and Rb/Rc . Rb/Rc becomes 3 to 5 when the ratio of the plano-convex -side beveling internal diameter to the flat-side beveling internal diameter is 1.4.

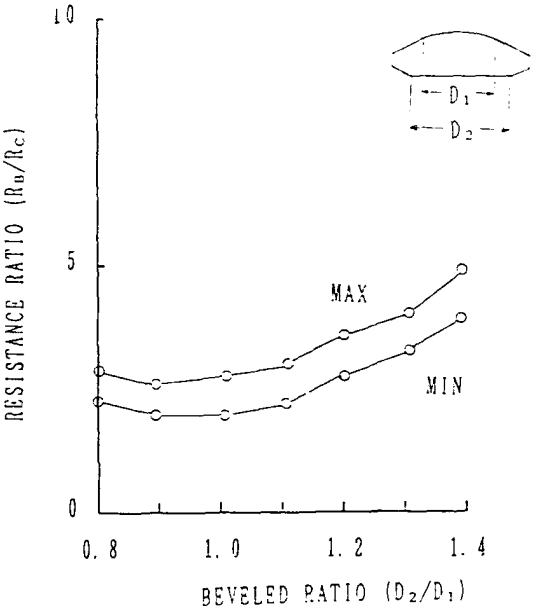


Figure 4 Effect of Beveling on Equivalent Resistance Ratio

OVERTONE EQUIVALENT RESISTANCE

Based on the above-mentioned result, we measured the equivalent resistance for each of the fundamental mode (R1), the 3rd overtone mode (R3) and the 5th overtone mode (R5). Figure 5 shows the measurement results: R1 was about 8 times as great as R3 and was about twice as great as R5. On the other hand, R5 became 4 times as great as R3. As R3 was smaller than R1 and R5, it was easy to forecast that the SC-cut crystal units developed oscillate with ease in a "coilless oscillation circuit."

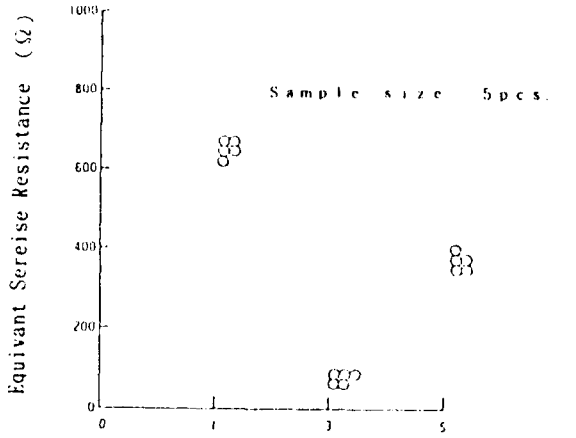


Figure 5 Overtone Equivalent Resistance

FREQUENCY TEMPERATURE CHARACTERISTICS

We measured the temperature characteristics of the SC-cut crystal unit.

For the measurement of the frequency characteristics, we employed the S-parameter method, using a network analyzer. We set the drive level at 10 ±1 μW.

Under such conditions,we made measurement of the frequency versus temperature characteristics over the temperature range of 30 'C to 105 'C at 0.1 'C step. Measurement result was as shown by Figure 6.

As the result, the frequency of the crystal proved to be stabilized at about 75'C. In addition, activity dips was not detected and frequency was shifted smoothly.

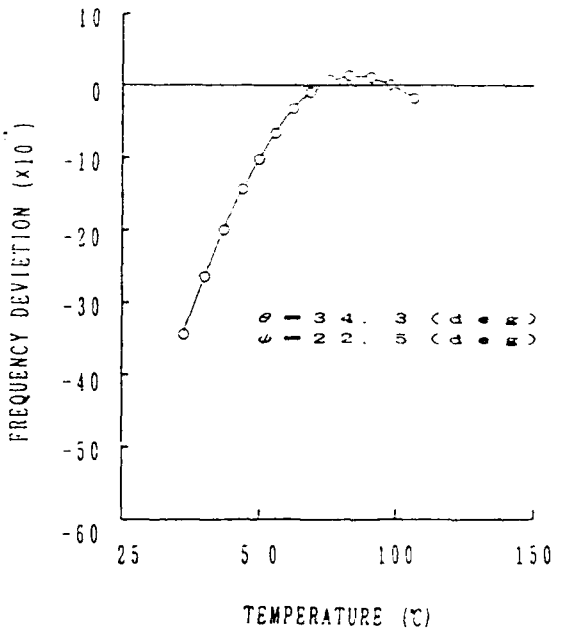


Figure 6 Frequency Temperature Characteristics

ELECTRIC EQUIVALENT PARAMETERS

The elctric equivalent parameters were measured by means of an automatic network analyzer, with the temperature kept at 75.00 'C ± 0.01 'C. Table 2 shows the measurement results. For reference, the measured values for an AT-cut crystal unit are listed here. The analysis of Table 2 reveals that SC-cut crystal unit are suitable for OCXOs due to the fact that the typical Q value is 2.14×10^6 and the typical capacitance ratio is 1.78×10^4. According to this figure,the Q value and the capacitance ratio of an SC-cut crystal unit are greater than those of an AT-cut crystal unit.

TABLE 2 Typical electric equivalent parameters

ITEM	SC-cut crystal unit
Q value (×10⁶)	216
Series resistance Rr (Ω)	80.6
Motional capacitance C1 (fF)	0.186
Motional inductance L1 (H)	5.50
Parallel capacitance Co (pF)	3.2
Capacitance ratio γ (×10⁴)	1.78

PHASE NOISE

Figure 7 shows the phase noise for an AT-cut crystal unit and an SC-cut crystal unit.

For an SC-cut crystal unit, the phase noise level was -105 dBc/Hz at the off-set frequency of 1Hz. On the other hand, for an AT-cut crystal unit, the phase noise level was -85dBc/Hz. As the difference between these two values is 20 dBc/Hz, the phase noise performance of the SC-cut crystal unit was 100 times better than that of the AT-cut crystal unit.

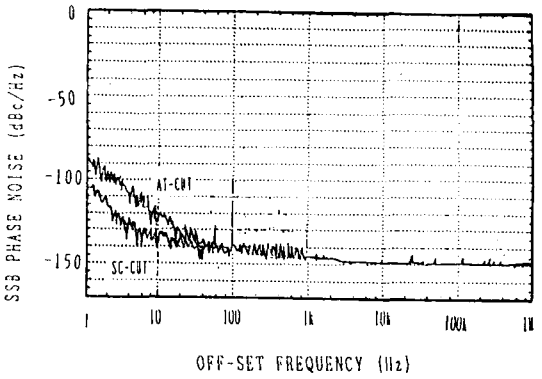


Figure 7 Measured Phase Noise of OCXO Using AT-cut and SC-cut

EFFECT OF ORIENTATION ON FREQUENCY

If the frequency stability was less than 1 x 10^{-8}, the influence exerted by the gravity of the earth was not negligible and the direction in which an OCXO was put causes its frequency to be changed. Table 3 showed the frequency change in the case of applying gravity in the Y axis and the Z axis, with the frequency in the case of applying gravity in the X axis as the standard. As shown by Table 3, the frequency stability of as SC-cut crystal unit was expressed in terms of 10^{-10} while that of an AT-cut crystal resonator was expressed in terms of 10^{-9}. Such a difference showed that the frequency variation due to gravity applied to an SC-cut crystal unit was smaller and the SC-cut crystal unit is superior to the AT-cut crystal unit in terms of frequency stability.

TABLE 3 Effect of orientation on frequency drift by X axis

AXIS	Frequency drift ($\times 10^{-9}$)	
	SC-cut crystal unit	AT-cut crystal unit
X	0.0	0.0
Y	0.5	-1.8
Z	0.2	2.0

THERMAL HYSTERESIS

For an OCXO, the crystal unit used was heated in an oven in order to enhance the frequency stability. H.J. Foster [5] and Mr. Okazaki, [6] et al., pointed out the problem lied in how the repetition of operation and nonoperation of an OCXO exerts influence upon the reproducibility of frequency. Figure 8 shows the measured data on frequency hysteresis. The SC-cut crystal unit have the frequency hysteresis expressed in terms of 10^{-10}, showing superior performance.

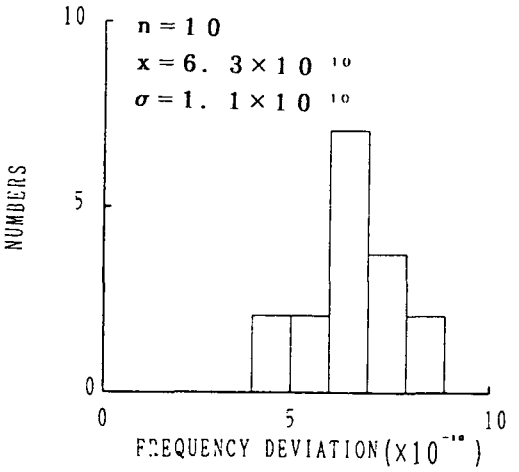


Figure 8 Distribution of Thermal Histeresis on Frequency

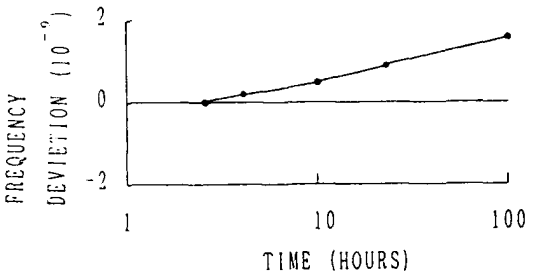


Figure 9 Frequency Aging Characteristics

AGING CHARACTERISTICS

Figure 9 shows the aging characteristics. We measured the frequency aging of OCXO using the crystal unit, as a frequency shift from cadinal point of 5 hours to 100 hours. Our measurement reveals that the aging ratio was 2 x 10^{-9}.

CONCLUSION

The optimum radius of the curvature and the asymmetric beveling technique can allow for us realize the more effective suppresion means of B-mode and A-mode of SC-cut crystal units than we have.

We can also evaluate the excellent performance on an OCXO by building in it the SC-cut crystal unit of ever research.

The SC-cut crystal units can be expected to be used in more applications.

ACKNOWLEDGEMENT

We would like to express an acknowledgement for Mr. Chiba and Mr. Suganuma, technical staffs of Nihon Dempa Kogyo.

REFERENCE

[1]E.P. Eernisse, "Quartz Resonator Frequency Shifts Arising From Electrode Stress", Proc, 29th AFCS, 1975, pp.1-4

[2]A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators", Phisical Acoustics : Principles and Methods , Vol. 13, W.P. Mason and R.N. Thurston, Eds., New York : Academic Press, 1977, Chapter 5, pp.115-181

[3]R.L. Fillter, J.R. Vig, "Fundamental Mode SC-cut Resonators", Proc, 34th, AFCS, 1980, pp.187-192

[4]Y. Oomura, Y. Watanabe, "Precise Determination of VHF Quartz Crystal Resonator Equivalent Parameters, Proc, 44th, AFCS, 1990, pp.363-366

[5]H.J. Forster, "Thermal hysteresis of AT+SC-cut quartz crystal resonators ; automated measurement method and results ", Proc. 36th, AFCS, 1982, pp.140-147

[6]M. Okazaki, M. Koyama, "Minimization of Short-Term aging of TCXO", Proceeding of 9th Quartz Devices Conference , 1987, pp.48-58

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

TT-Cut Torsional Quartz Crystal Resonator

Hirofumi KAWASHIMA and Mitsuhiro NAKAZATO

Seiko Electronic Components Ltd.
Hiraicho Tochigi-shi 328, Japan

Abstract

This paper describes a new cut torsional quartz crystal resonator which is designated "TT-Cut". The object of this paper is to propose a new cut torsional quartz crystal resonator and to clarify its frequency characteristics, frequency temperature behavior and electrical equivalent circuit parameters.

First, in analysis procedure, equation of motion is derived from an energy method in this paper. next, by solving the equation of motion under the boundary conditions of "free-free bar" or "clamped-free bar", the frequency equations are derived as a function of thickness z_0, width x_0 and length y_0.

Furthermore, from the frequency equation a relationship of frequency constant($f \cdot y_0$) versus thickness-to width ratio Rzx($=z_0/x_0$) and a relationship between thickness to width ratio Rzx and cut angles(ϕ, θ) where the first order temperature coefficient α reaches zero are theoretically derived. As a result, numerous relationships where α reaches zero are found to exist between thickness-to-width ratio Rzx and cut angles(ϕ, θ), especially, the second order temperature coefficient β has a small value of $-1.16\times10^{-8}/℃^2$ whose absolute value is approximately one third of the well-known flexural mode quartz crystal resonator. The value of β is then compared with the measured data of $-1.29\times10^{-8}/℃^2$, so that both results are found to agree sufficiently well.

Finally, series resistance R_1 and a quality factor Q of a tuning fork-type resonator are examined, consequently, it is shown that a tuning fork-type torsional quartz crystal resonator is obtained with a small R_1 of 2.2 to 14.4kΩ and a large Q value of 276,000 to 378,000 in frequency range of 385 to 444kHz.

§1 Introduction

A tuning fork-type flexural quartz crystal resonator with a zero temperature coefficient is much used as a resonator for a wristwatch, because of noteworthy features of easy miniaturization, low frequency, high shock resistance and so on. However, frequency temperature behavior of the tuning fork-type flexural quartz crystal resonator which is used widely, have approximately a parabolic curve with a temperature coefficient of about $-3.5\times10^{-8}/℃^2$; because its absolute value is comparatively large, it was impossible to obtain a quartz crystal resonator with a small frequency change over a wide temperature range.

Furthermore, though GT cut quartz crystal resonators excellent in frequency temperature behavior were reported [1]-[6], as the resonator for [1]-[4] has large consumption power, because of high frequency, and as the other is inferior in miniaturization and shock resistance, it is very difficult to utilize the GT cut quartz crystal resonators in a wristwatch.

In addition, recently, it was reported [7]-[9] that a coupling between a flexural mode vibration and a torsional mode vibration of a tuning fork-type gives an improvement in frequency temperature behavior of the flexural mode vibration. However, because of a coupling resonator, it is very difficult to obtain resonators with a constant coupling between both vibrations in mass production, they were hardly put to practical use, because the subject remains such that frequency temperature behavior is not obtained stably. A single mode quartz crystal resonator of low frequency with a value of the second order temperature coefficient in an absolute value smaller than that of the tuning fork-type quartz crystal resonator is, therefore, desired earnestly and expected.

A torsional mode quartz crystal resonator exists as a vibration mode where by comparatively low frequency can be realized. However, few papers on frequency temperature behavior were reported by Hermann[10], Vasin[11] and so on. Further, they designed a torsional quartz crystal resonator with length oriented along the electrical axis(x-axis) and examined its frequency temperature behavior with a rotation around the x-axis, but, the resonator is not turned to practical use yet. In addition, frequency temperature behavior of a torsional quartz crystal resonator with length oriented along the optical axis(z-axis) and the mechanical axis(y-axis) is not examined at all.

The object of this paper is to propose a new-cut torsional quartz crystal resonator with an absolute value of the second order temperature coefficient smaller than that of the tuning fork-type flexural quartz crystal resonator, which is designated "TT-Cut", and to clarify its frequency characteristics, frequency temperature behavior and electrical characteristics. First, in analysis procedure, equation of motion is derived from an energy method in this paper, by solving the equation of motion under the boundary conditions of "free-free bar" or "clamped-free bar", the frequency equations are derived. Furthermore, from the equation, a

CH2965-2/91/0000-130 $1.00 © 1991 IEEE

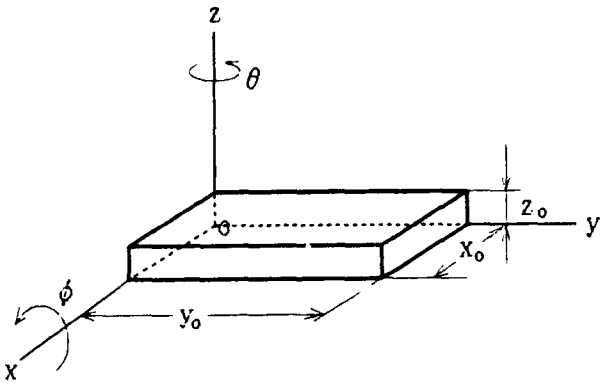


Fig. 1 Torsional quartz crystal bar and its coordinate system.

relationship between thickness-to-width ratio Rzx and cut angles(ϕ, θ) where the first order temperature coefficient α reaches zero is shown. Especially, cuts whereat α reaches zero at a positive cut angle of ϕ are designated "TT(X_1)-Cut" and cuts whereat α reaches zero at a negative cut angle of ϕ are designated "TT(X_2)-Cut". In addition, an absolute value of the second order temperature coefficient becomes approximately one third of the tuning fork-type flexural quartz crystal resonator by combination of ϕ and θ of TT(X_1)-Cut and TT(X_2)-Cut. Thus, a tuning fork-type torsional quartz crystal resonator excellent in frequency temperature behavior is theoretically found to be obtained. The analysis results are then compared with the measured values, consequently, both results are also found to agree comparatively well. Let us describe the detail below.

§2 Analysis Procedure

2.1 Equation of motion

Figure 1 shows a torsional quartz crystal bar and its coordinate system. This resonator is expressed by IRE expression of zywt(ϕ/θ). In addition, the resonator has a dimension of thickness z_0, width x_0 and length y_0. A torsional mode quartz crystal resonator, as is well known, is a resonator having torsional moment about the y-axis.

Now, let us take torsional displacement angle ϕ [8], torsional rigidity Ct, density of quartz ρ, polar radius of gyration κ and sectional area A for a resonator. When length y_0 is extremely larger than width x_0 and thickness z_0 ($y_0 \gg x_0, z_0$), the term of lateral motion which is expressed by torsion-function can be ignored. Therefore, when time t is taken, kinetic energy K and potential energyU are, respectively, given as

$$K=\frac{1}{2}\rho A\kappa^2\int_0^{y_0}\left(\frac{\partial\phi}{\partial t}\right)^2dy \qquad (1)$$

$$U=\frac{1}{2}Ct\int_0^{y_0}\left(\frac{\partial\phi}{\partial y}\right)^2dy \qquad (2)$$

From Hamilton's principle, since the following relation must be satisfied:

$$\delta\int_{t_0}^{t_1}(K-U)dt=0 \qquad (3)$$

The equation of motion and the boundary conditions can be derived. Consequently,

(Equation of motion)

$$\rho I_0\frac{\partial^2\phi}{\partial t^2}-Ct\frac{\partial^2\phi}{\partial y^2}=0 \qquad (4)$$

where

$I_0=A\kappa^2$:Polar moment of inertia

$$Ct\,[8]=\frac{x_0z_0^3}{3s'_{66}}\left[1-\frac{z_0}{x_0}\sqrt{\frac{s'_{44}}{s'_{66}}}\,\frac{192}{\pi^5}\sum_{j=0}^{\infty}\frac{1}{(2j+1)^5}\cdot\tanh\left\{\frac{(2j+1)\pi}{2}\,\frac{x_0}{z_0}\sqrt{\frac{s'_{66}}{s'_{44}}}\right\}\right]$$

s'_{66}, s'_{44}:Elastic compliances

(Boundary conditions)

○ $y=0,\ y_0;\ \dfrac{\partial\phi}{\partial y}=0$ for free-free bar (5)

○ $y=0\ ;\ \phi=0$
$y=y_0\ ;\ \dfrac{\partial\phi}{\partial y}=0$ } for clamped-free bar (6)

Accordingly, the equation of motion of equation (4) is easily solved by employing the boundary conditions of equations (5) and (6).

2.2 Frequency equation

The following exhibits a solution of the equation of motion and derive the frequency equation. As there are the two boundary conditions of equations (5) and (6), the frequency equation is first derived from the boundary condition of equation (5) and next, from that of equation (6). Finally, the obtained equations are collectively described. Now, when taking angular frequency ω, a torsional displacement angle ϕ can be expressed by the form of

$$\phi=(A\sin a\,y+B\cos b\,y)\cos\omega t \qquad (7)$$

From the boundary condition of equation (5), $A=0$ and $b=m\pi/y_0$ ($m=1,2,3,\cdots$) is also obtained. Therefore, equation (7) is newly expressed by $\phi=B\cos b\,y\cos\omega t$, the frequency equation is obtained under the boundary condition of "free-free bar", substituting the ϕ into equation (4). On the other hand, from the boundary condition of equation (6), $B=0$ and $a=n\pi/2y_0$ ($n=1,3,5,\cdots$) is also obtained. In this case, equation (7) is newly expressed by $\phi=A\sin a\,y\cos\omega t$, the frequency equation is, as well, obtained under the condition of "clamped-free bar", substituting the ϕ into

equation (4). The results with respect to the resonant frequency f are collectively described as follows:

(Free-free bar)

$$f = \frac{m}{y_0}\sqrt{\frac{1}{\rho s'_{66}}}\frac{(\frac{z_0}{x_0})}{\sqrt{1+(\frac{z_0}{x_0})^2}}\sqrt{1-h} \qquad (8)$$

$(m=1, 2, 3, \cdots)$

(Clamped-free bar)

$$f = \frac{n}{2y_0}\sqrt{\frac{1}{\rho s'_{66}}}\frac{(\frac{z_0}{x_0})}{\sqrt{1+(\frac{z_0}{x_0})^2}}\sqrt{1-h} \qquad (9)$$

$(n=1, 3, 5, \cdots)$

where $$h = \frac{z_0}{x_0}\sqrt{\frac{s'_{44}}{s'_{66}}}\frac{192}{\pi^5}\sum_{j=0}^{\infty}\frac{1}{(2j+1)^5}\cdot \tanh\left\{\frac{(2j+1)\pi}{2}\frac{x_0}{z_0}\sqrt{\frac{s'_{66}}{s'_{44}}}\right\}$$

The each resonant frequency f is given as a function of length y_0, thickness-to-width ratio Rzx(=thickness/width) and elastic compliances. In particular, when Rzx is extremely small, the resonant frequency f is approximately in proportion to the thickess-to-width ratio Rzx, so that f lowers without limit, according as Rzx is taken to be small infinitely. In addition, as is easily understood from comparison with equations (7) and (8), a resonator with the boundary condition of "clamped-free bar" is suitable so as to obtain that of lower frequency at the same dimension. Accordingly, the tuning fork-type quartz crystal resonator which can be considered as two cantilevers having the boundary condition of "clamped-free bar", is examined in this paper and its various characteristics are estimated.

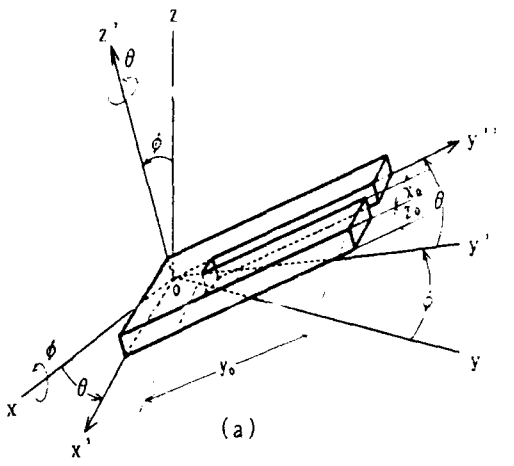
Next, resonator's shape and size to be treated in this paper, and its excitation electrode construction are described.

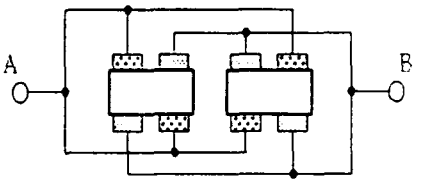
§3 Resonator's Shape and Excitation Electrode

Figure 2 shows a shape of a tuning fork-type torsional quartz crystal resonator (a) and the sectional view of its excitation electrode construction (b). The resonator is formed with rotation angle ϕ about the x-axis from Z plate and rotation angle θ about the z'-axis of the new axis for the z-axis. In addition, the present resonator has a dimension of width x_0, thickness z_0 and length y_0.

§4 Results and Discussion

The calculated and measured values are shown, they are, in detail, discussed. First, a relationship between cut angles(ϕ, θ) and piezoelectric constant e_{16} to determine an intensity of excitation for a torsional resonator at TT(X_1, X_2)-Cut is shown. Next, a relationship between cut angle ϕ and frequency constant ($f \cdot y_0$) is shown, when the cut angle θ is taken as a parameter, then, a relationship between thickness-to-width ratio Rzx and frequency constant ($f \cdot y_0$) is, as well, shown, when the cut angle θ is taken as a parameter, and based on these relationships, a relationship of actual resonant frequency f versus length y_0 is shown, when Rzx is taken as a parameter. Next, a relationship between cut angle ϕ and thickness-to-width ratio Rzx wherein the first order temperature coefficient α reaches zero, is shown, when cut angle θ is again taken as a parameter, simultaneously, a value of the then second order temperature coefficient β is shown, as a result, it is shown that there are the cut angles(ϕ, θ) whereat the absolute value of β becomes approximately one third of the flexural mode resonator. The then frequency temperature behavior is also shown. Finally, a relationship of series resistance R_1 and a quality factor Q versus Rzx is shown, when cut angle θ is taken as a parameter and also, a typical value for the obtained electrical equivalent circuit parameters is shown. In an analysis of frequency temperature behavior, the constants shown in Appendix of the reference [12] are utilized. Let us explain in detail below.





(b)

Fig. 2 Resonator shape (a) and sectional view of electrode construction (b).

4.1 Piezoelectric constant

Figure 3 shows a relationship between cut angle ϕ and piezoelectric constant e_{16}, when cut angle θ is taken as a parameter. When $\theta = 0°$, e_{16} reaches zero. Therefore, the present resonator is found not to be ex-

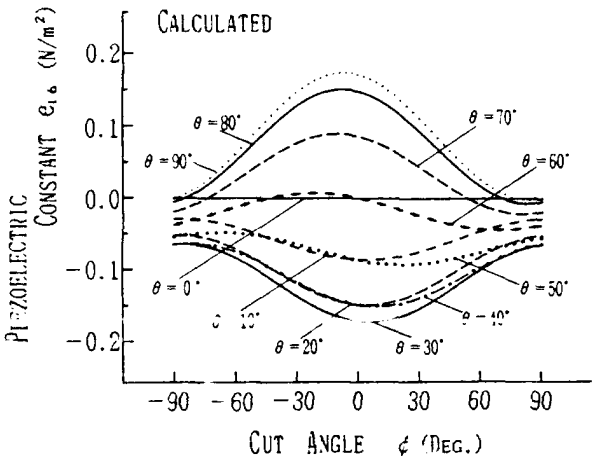


Fig. 3 Relationship between piezoelectric constant e_{16} and cut angle ϕ when cut angle θ is taken as a parameter.

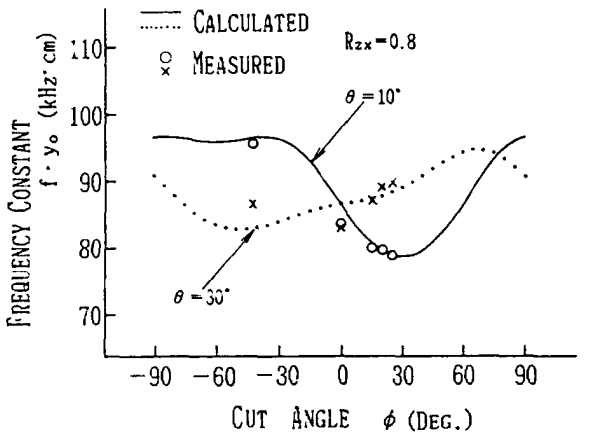


Fig. 4 Relationship between frequency constant ($f \cdot y_0$) and cut angle ϕ when cut angle θ is taken as a parameter.

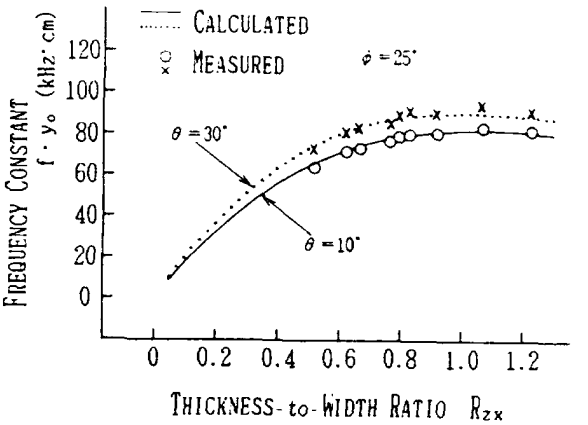


Fig. 5 Relationship between frequency constant ($f \cdot y_0$) and thickness to width ratio Rzx when cut angle θ is taken as a parameter.

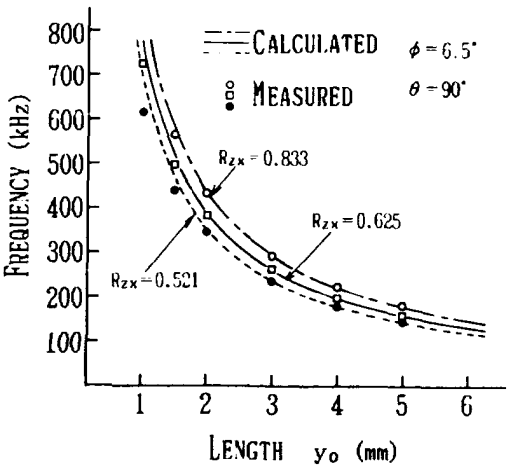


Fig. 6 Relationship between resonant frequency f and length y_0 of vibration arm when thickness to width ratio Rzx is taken as a parameter.

cited by the excitation electrode illustrated in Fig. 2 (b). In addition, it is sufficiently understood from Fig. 3 that e_{16} varies markedly with a change of θ. Since the e_{16} decides a value of series resistance R_1 for a resonator, it goes without saying that cut angles (ϕ, θ) whereat the absolute value of e_{16} becomes as large as possible, must be chosen so as to obtain a small R_1.

4.2 Frequency constant

Figure 4 shows a relationship between cut angle ϕ and frequency constant($f \cdot y_0$) of a resonator with thickness to width ratio Rzx=0.8, when cut angle θ is taken as a parameter. The solid and dotted lines are the calculated values for θ=10° and 30°, and the signs ○, × are the then measured values. Both results are in sufficiently good agreement. Then, the frequency constant varies with cut angle ϕ, when a miniaturized resonator of low frequency is required, TT(X_1)-Cut is found to prefer to TT(X_2)-Cut for θ=10°, because the constant for TT(X_1) Cut with a positive value of ϕ is smaller than that for TT(X_2) Cut with a negative value of ϕ.

Figure 5 shows a relationship between thickness to width ratio Rzx and frequency constant ($f \cdot y_0$) of a resonator with cut angle ϕ=25°, when cut angle θ is taken as a parameter. As is obvious from equation (9), since frequency constant ($f \cdot y_0$) is approximately proportional to thickness-to-width ratio Rzx for thickness z_0 smaller than width x_0, as is shown in Fig. 5, the frequency constant ($f \cdot y_0$) increases gradually according to gradual increase of Rzx. However, according to further increase of Rzx, it shows gradual saturation. Furthermore, frequency constant ($f \cdot y_0$) has a value of 78.9kHz·cm in the calculation and 78.8kHz·cm in the measured value for cut angles ϕ=25°, θ=10° and Rzx=0.8. Both results agree well. This value is between frequency constants for tuning fork type quartz crystal resonator and a length extensional quartz crystal resonator.

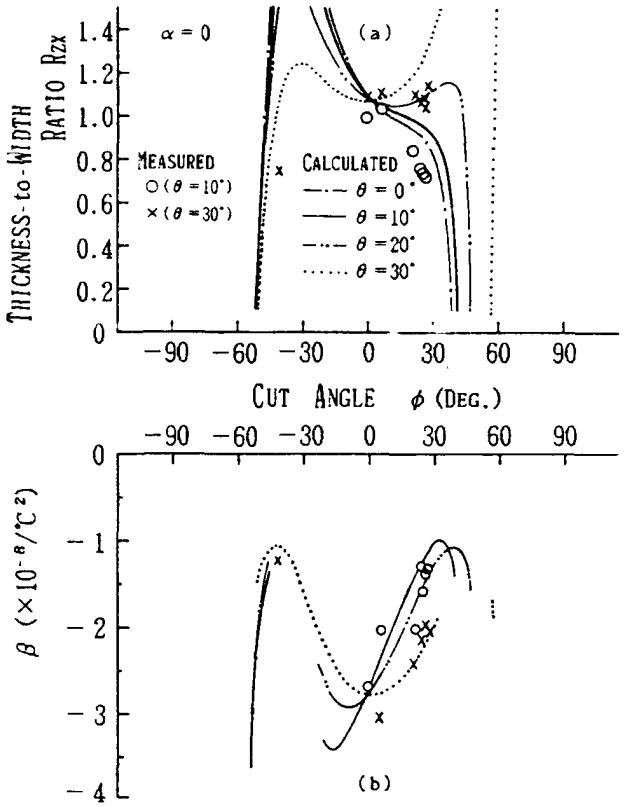


Fig. 7 Relationship between thickness-to-width ratio Rzx and angle ϕ with the first order temperature coefficient $\alpha = 0$ when cut angle θ is taken as a parameter, and the then β value.

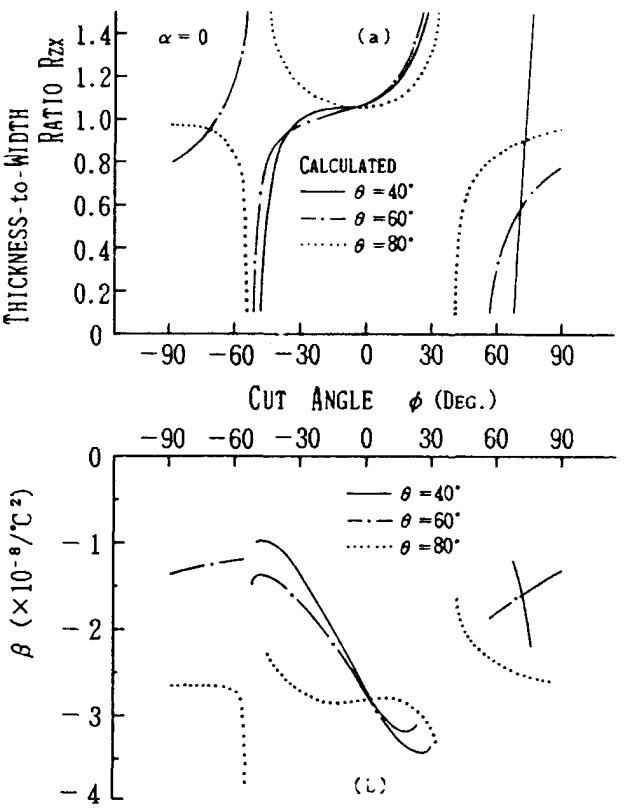


Fig. 8 Another example of Relationship between thickness-to-width ratio Rzx and angle ϕ with the first order temperature coefficient $\alpha = 0$ when cut angle θ is taken as a parameter, and the then β value.

Therefore, when resonator's size is taken into account, the present resonator is effective in a frequency range of 200kHz to 500kHz particularly as will be exhibited in Fig. 6.

Figure 6 shows a relationship of resonant frequency versus length y_0 of a resonator with cut angles $\phi = 6.5°$, $\theta = 90°$, when thickness to width ratio Rzx is taken as a paremeter. In this condition the thickness-to-width ratio Rzx is 0.95, which gives the first order temperature coefficient $\alpha = 0$. Accordingly, the resonant frequency varies from 631kHz to 143kHz, when length y_0 varies from 1mm to 5mm at Rzx = 0.521. In addition, in order to house the resonator in a resonator unit of 2mm $\phi \times$6mm length which is used widely for a tuning fork-type quartz crystal resonator, length y_0 must be smaller than 3.0mm. From the result, resonant frequency higher than 233kHz is obtained, but, because y_0 can not be taken to be very small, when electrical characteristics are taken into consideration, the upper limit of frequency for a tuning fork type torsional quartz crystal resonator, therefore, seems to be approximately 840kHz. In order to realize a resonator of high frequency, as was already-described, it is understood that a free-free bar type resonator is more suitable because its frequency constant becomes large.

4.3 Frequency temperature coefficients

Figure 7 shows a relationship between thickness to width ratio Rzx and cut angle ϕ of a resonator where the first order temperature coefficient α reaches zero (a) and the then second order temperature coefficient β (b). As is apparent from Fig. 7, it is understood that numerous relationships exist between Rzx and ϕ whereat α reaches zero. However, a relationship between Rzx and ϕ whereat an absolute value of β becomes comparatively small, exists finitely. When taking into account the cut angle whereat a resonator is piezoelectrically excited sufficiently, for TT(X_1)-Cut with a positive value of ϕ, the absolute value of β becomes comparatively small at $\theta = 10°$ and $\phi = 25°$ to 28°, and has a value of 1.16×10^{-8}/℃² to 1.25×10^{-8}/℃² in the calculation and 1.29×10^{-8}/℃² to 1.58×10^{-8}/℃² in the experiments. These results agree sufficiently well. On the other hand, for TT(X_2) Cut with a negative value of ϕ, the absolute value of β becomes comparatively small at $\theta = 30°$ and $\phi = 41°$ to $-43°$, and has a value of 1.06×10^{-8}/℃² to 1.07×10^{-8}/℃² in the calculation and 1.23×10^{-8}/℃² in the experiment for $\phi = -42°$. Thus, the absolute value of β for TT(X_1) Cut and TT(X_2) Cut has one third value of the tuning fork type quartz crystal reso-

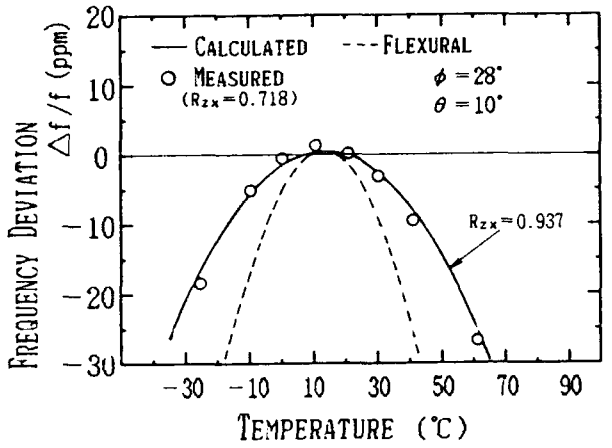


Fig. 9 Frequency temperature behavior.

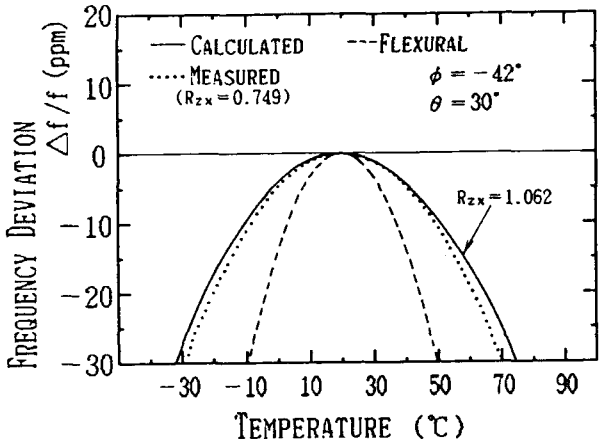


Fig. 10 Another example of frequency temperature behavior.

nator. In addition, when $\phi=28°$ and $\theta=10°$, thickness-to width ratio Rzx to give $\alpha=0$ at 20℃ is 0.705 in the experiments and 0.93 in the calculation, though an error of about 24% between both results takes place, this is probably because of ignorance of piezoelectric effect and elastic stiffness c_{16} in the derivation of the frequency equation, an error of width dimension caused by an etching process and so on.

Figure 8 shows the same relationship as described in Fig. 7, when θ varies from 40° to 80°. The absolute value of β is found to become small also for $\phi=40°$ to 60°. Although a relationship between Rzx and ϕ which takes β to be small exists numerously, this is decided by goodness of piezoelectric efficiency, ease by a chemical etching process and so on.

4.4 Frequency temperature behavior

Figure 9 shows an example of frequency temperature behavior of a resonator with $\phi=28°$, $\theta=10°$ and length $y_a=2$mm. The solid lines show the calculated value when Rzx=0.937, while the sign ○ shows the measured values when Rzx=0.718. Though the error of about 23%

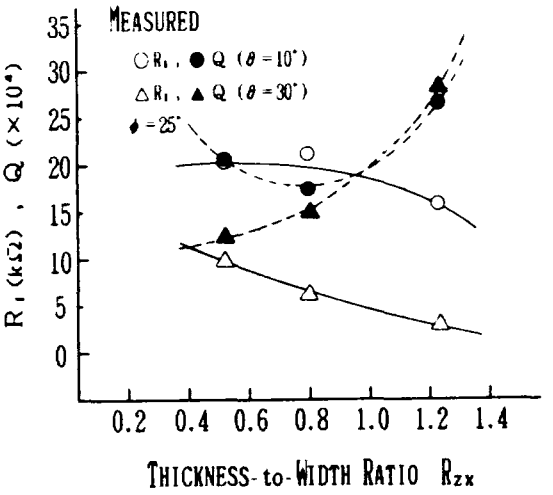


Fig. 11 Relationship of R_1 and Q versus thickness to width ratio Rzx when cut angle θ is taken as a parameter.

Table 1 Typical values of electrical equivalent circuit parameters.

φ (DEG.)	θ (DEG.)	f (kHz)	R_1 (kΩ)	L_1 (kH)	Q (x10³)	C_0 (pF)	r
0	30	444.1	2.2	0.30	378	0.425	984
26	10	385.8	14.4	1.64	276	0.434	4190

takes place in the calculated value and the measured one of Rzx, this reason was already described in Fig. 7. In addition, the second order temperature coefficient β of the resonator with $\alpha=0$ approximately is 1.16×10^{-8} /℃² in the calculation and -1.29×10^{-8}/℃² in the experiments, these results agree sufficiently well. The dotted line shows the frequency temperature behavior of a tuning fork-type flexural quartz crystal resonator with $\beta=-3.5\times10^{-8}$/℃² whose absolute value is comparatively large. The present torsional quartz crystal resonator, therefore, shows more excellent frequency temperature behavior than that of the tuning fork type flexural quartz crystal resonator.

Figure 10 shows another example of frequency temperature behavior of a resonator with cut angles $\phi=-42°$ and $\theta=30°$. It is easily predicted that turn over temperature point (T·p) shifts, because α varies markedly with a value of Rzx. In the condition of the present resonator, when Rzx=1.062 in the calculation and Rzx=0.749 in the experiments, α reaches zero at 20℃, so that T·p can be set up in the vicinity of room temperature. Furthermore, the then β is -1.07×10^{-8}/℃² and -1.23×10^{-8}/℃² respectively, as well as the resonator in Fig. 9, a torsional quartz crystal resonator is theoretically and experimentally obtained with a small β in an absolute value. Hereafter, the detail confirmation including another cut angles will be performed in the experiments.

4.5 Electrical characteristics

Figure 11 shows a relationship of series resistance R_1 and a quality factor versus thickness to width ratio

Rzx of a resonator with cut angles ϕ 25°, θ 10° and 30°, and length y_0 2mm, when width x_0 and thickness z_0 are varied. As already described in Fig.3, because the piezoelectric constant differs at θ 10° and 30°, that is, piezoelectric constant e_{16} at θ 30° is larger than that at θ 10°, R_1 has a remarkably small value and a high quality factor is obtained. It is understood from Fig.11 that R_1 and Q value hardly vary versus thickness to width ratio Rzx.

Table 1 shows typical values of electrical equivalent circuit parameters for resonators with cut angles ϕ 0° or 26° and θ 10° or 30°, and length y_0 2mm. The values are sufficiently satisfactory as a resonator.

§5 Conclusions

In this paper, a new cut torsional quartz crystal resonator, which was designated "TT Cut" classified into "TT(X_1) Cut" and "TT(X_2) Cut", was proposed, and a study was performed with a view to clarifying its frequency characteristics, frequency temperature behavior and electrical characteristics. First, in an analysis procedure, equation of motion was derived from an energy method and the frequency equation was derived under the boundary condition of "free-free bar" or "clamped free bar". Next, frequency characteristics were clarified versus cut angle ϕ and thickness to-width ratio Rzx from the obtained frequency equation. Specifically, it was shown in the calculation and in the experiments that a region wherein frequency constant becomes small for TT(X_1) Cut exists. Simultaneously, the calculated and measured values were found to agree comparatively well.

In addition, it was theoretically shown that a combination of thickness to width ratio Rzx and cut angles (ϕ, θ) where the first order temperature coefficient reaches zero exists numerously. In the combinations, an absolute value of the second order temperature coefficient was found to become small at both regions of TT(X_1)-Cut and TT(X_2) Cut, the calculated values were then compared with the measured ones for TT(X_1)-Cut. As a result, it was shown that the calculated and measured values of the second order temperature coefficient for $\alpha = 0$, agree sufficiently well. Furthermore, an example of frequency temperature behavior was shown and also it was shown in the calculation and in the experiments that the tuning fork type torsional quartz crystal resonator has a frequency deviation versus temperature smaller than the tuning fork-type flexural quartz crystal resonator.

Finally, electrical equivalent circuit parameters for the present torsional quartz crystal resonator were examined experimentally. Consequently, the present resonator was found to be a resonator with a low frequency of a few hundred kHz, a small R_1 and a high quality factor. Therefore, the present resonator is very suitable for handy type products such as a wristwatch and so on.

For the coming subject, a study on possibility of improvement in frequency temperature behavior will be performed.

Acknowledgement

We wish to express our thanks to Mr. K. Sunaga for his help in the calculation of the numerical value by computer.

References

[1] H. Kawashima:"A Variational Analysis of a GT Cut Quartz Crystal Resonator with the Supporting Portions at the Ends", Trans. IECE(A), J68-A, 8, pp. 771-778 (1985-08)(in Japanese).

[2] H. Kawashima:"Variational analysis of GT cut quartz crystal resonators with the supporting portions at the ends", Proc. the 40th Annual Symp. on Frequency Control, pp. 193-200(1986).

[3] H. Kawashima:"An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New Shape GT Cut Quartz Crystal Resonator", IEEE Ultrasonics Symposium Proc., 1, pp. 465-470(1989).

[4] H. Kawashima and M. Matsuyama:" An Analysis of Electrical Equivalent Circuit Parameters for a GT Cut Quartz Crystal Resonator with the Supporting Portions at Both Ends", J. IEICE(A), J71-A, 6, pp. 1217-1224 (1988-06)(in Japanese).

[5] W. P. Mason:"A New Quartz Crystal Plate Designated the GT, which Produces a Very Constant Frequency over a Wide Temperature Range", Proc. IRE, pp. 220-223 (1940).

[6] M. Nakazawa and S. Kozima:"A Study of GT Type Quartz Crystal Plates", IEEE Trans. Sonics&Ultrason., SU-29, 3, (May 1982).

[7] E. Momosaki, S. Kogure, M. Inoue and T. Sonoda :" New Quartz Tuning Fork with Very Low Temperature Coefficient", Proc. the 33rd Annual Symp. on Frequency Control, pp. 247-251(1979).

[8] H. Kawashima:"Vibration Analysis of Flexural Torsional Mode Tuning Fork Type resonator", Trans. IECE (A), J68-A, 7, pp. 680-687(1985-07)(in Japanese).

[9] H. Kawashima:"Vibration Analysis of Coupled Flexural Torsional Mode Tuning Fork Type Quartz Crystal Resonator", Proc. the 42nd Annual Symp. on Frequency Control, pp. 45-52(1988).

[10] J. Hermann:"DT Cut Torsional Resonators", Proc. the 31st Annual Symp. on Frequency Control, pp. 55-61 (1977).

[11] I. G. Vasin and P. G. Pozdnjakov:"Quartz Resonator", US Patent 3376439, awarded April 2, 1968.

[12] H. Kawashima:"Vibration Analysis of GT Cut Quartz Crystal Resonator Taking Account of Dissipation Term", J. IEICE(A), J72-A, 4, pp. 633-639 (1989-04)(in Japanese).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Thickness-Shear Mode Shapes and Mass-Frequency Influence Surface of a Circular and Electroded AT-Cut Quartz Resonator.

Y-K Yong*, J.T. Stewart*, J. Detaint†, A. Zarka‡, N. Capelle‡, and Y. Zheng‡
*Dept. of Civil/Environmental Engineering, Rutgers University,
P.O. Box 909, Piscataway, NJ 08855-0909: † CNET dept
PAB/BAG/MCT/CMM, Bagneux, France; ‡ L.M.C. Universite, Paris, France.

ABSTRACT.

Finite element solutions for the fundamental thickness shear mode and the second anharmonic overtone of a circular, 1.87 Mhz AT-Cut quartz plate with no electrodes are presented and compared with previously obtained[1] results for a rectangular plate of similar properties. The edge flexural mode in circular plates, a vibration mode not seen in the rectangular plate, is also presented. A 5 Mhz circular and electroded AT-cut quartz plates is studied. A portion of the frequency spectrum is constructed in the neighborhood of the fundamental thickness-shear mode. A convergence study is also presented for the electroded 5 Mhz plate. A new two dimensional technique for visualizing the vibration mode solutions is presented. This method departs substantially from the three dimensional "wire-frame" plots presented in the previous analysis. The two dimensional images can be manipulated to produce nodal line diagrams and can be color coded to illustrate mode shapes and energy trapping phenomenon. A contour plot of the mass-frequency influence surface for the plated 5 Mhz resonator is presented. The mass-frequency influence surface is defined as a surface giving the frequency change due to a small localized mass applied to the resonator surface.

I. Introduction.

The use of quartz crystal plates in ultrasonics is well studied experimentally. Analytical closed form solutions of a freely vibrating anisotropic elastic plate can in general only be obtained for either a one dimensional problem, or in an infinite domain. For a two-dimensional finite domain, solutions may be obtained either approximately, or for certain specific types of boundary conditions and geometries. To perform a practical analysis of the AT-cut quartz resonator, it is necessary to solve the free vibration problem for a finite plate. In this paper, such a problem is solved using the finite element method.

The high frequency vibrations as well as the anisotropicity of quartz demands that a general thick plate theory be used. In this study, Mindlin's two dimensional plate equations are truncated to a first order approximation. The first order approximation is chosen because it is the lowest order formulation which includes the major modes of interest, namely, the fundamental thickness-shear modes[2]. Real AT-cut resonators generally contain a thin electrode plating on a portion of each face, across which a time varying voltage is applied to drive the vibration. The effect of this electrode plating was studied analytically by Mindlin [3] for an infinite plate. Mindlin's equations for the electroded plate are incorporated into the present analysis so that the effects on frequencies and mode shapes of this electrode plating can be studied. The formulation for the plated problem is also useful for assessing the mass loading sensitivity of a given plate. Yong and Vig[4] proposed that surface contamination may be a contributor to random frequency fluctuations of a thickness shear resonator, and a useful parameter in predicting the magnitude of these frequency fluctuations is the frequency change due to molecular mass loading on the plate's surface. This mass loading effect is highly position dependent. The variation of the change in frequency due to a unit mass loading at a certain position is given by the frequency influence surface, which, in this paper, is calculated for an electroded 5 Mhz circular plate with an aspect ratio (ratio of diameter to thickness) of 24.0.

The finite element formulation for the first order plate theory is derived using a non-conforming bilinear quadrilateral element. The element is chosen because its formulation is relatively simple and it performs well as a thick plate element[1]. A finite element program is written to solve the resulting eigenvalue problem. The program includes an efficient eigenvalue solver employing the Lanczos algorithm [5,6] which allows the user to find resonant frequencies within a specified bandwidth. Using this program, the frequency spectrum and mode shapes in the region around the fundamental thickness-shear mode are studied. The computer work was performed on a CRAY-2 supercomputer at the National Center for Supercomputing Applications (NCSA).

CH2965-2/91/0000-137 $1.00 © 1991 IEEE

II. Mindlin's Plate Equations for Thickness-Shear, Thickness-Twist, and Flexural Vibrations in an AT-cut, Electrode Plated Quartz Plate.

Figure 1 shows the general plate problem considered: a circular plate of diameter $2a$, thickness $2b$ and mass density ρ with optional electrode plating of diameter $2a'$ thickness $2b'$ and mass density ρ'. Mindlin's two dimensional plate equations are obtained by expanding the three dimensional equations of motion for an elastic continuum into an infinite series of powers of the thickness coordinate (x_2). The resulting equations are then integrated through the thickness, eliminating the x_2 dependence. The approximate plate equations are then obtained by truncating[2] this series to the first order plate equations. For economy of space, the finite element equations and their derivations are not repeated here and the interested reader is referred to reference [1]. We list here the partial differential equations for the boundary value problem of an AT-cut quartz resonator solved numerically using finite elements:

a) Stress equations of motion.

The stress equations of motion accommodating thickness-shear, thickness-twist, and flexural vibrations are

$$Q_{1,1} + Q_{3,3} = 2b\rho(1+R)\ddot{u}_2, \tag{1}$$

$$M_{1,1} + M_{5,3} - Q_1 = \frac{2}{3}b^3\rho(1+3R)\ddot{\psi}_1, \tag{2}$$

$$M_{5,1} + M_{3,3} - Q_3 = \frac{2}{3}b^3\rho(1+3R)\ddot{\psi}_3, \tag{3}$$

where the terms Q_i $(i=1,3)$ are the transverse shear forces per unit length, and M_i $(i=1,3)$, are the bending moments per unit length, and, M_5 is the twisting moment per unit length. The term $R = 2\rho'b'/(\rho b)$ is ratio of mass of electrodes to mass of plate per unit area. Displacement u_2 is the transverse displacement in the x_2 direction, and ψ_1, and ψ_3 are shear rotations about the x_1, and x_3 axes, respectively.

b) Stress-strain relations.

For an AT-cut quartz plate, the elastic constants exhibits monoclinic symmetry when the x_1 axis coincides with one of the digonal axes of crystal. The constitutive relations are

$$Q_1 = 2b\bar{k}_1^2\bar{C}_{66}\gamma_1, \tag{4}$$

$$Q_3 = 2b\bar{k}_3^2\bar{C}_{44}\gamma_3, \tag{5}$$

$$M_1 = \frac{2}{3}b^3\left(\hat{C}_{11}\chi_1 + \hat{C}_{13}\chi_3\right), \tag{6}$$

$$M_3 = \frac{2}{3}b^3\left(\hat{C}_{13}\chi_1 + \hat{C}_{33}\chi_3\right), \tag{7}$$

$$\text{and } M_5 = \frac{2}{3}b^3\hat{C}_{55}\chi_5, \tag{8}$$

where γ_i $(i=1,3)$ are the shear strains, and χ_1, χ_3, and χ_5 are the x_1 bending curvature, x_3 bending curvature, and twisting curvature, respectively. The coefficients $\bar{k}_1^2$ and $\bar{k}_3^2$ in Eqns. (4)–(5) are shear correction factors which include mass loading effects[3], that is,

$$\bar{k}_1^2 = \frac{\pi^2}{12}\frac{(1+3R)}{(1+R)^2}, \tag{9}$$

$$\text{and } \bar{k}_3^2 = k_3^2\frac{(1+3R)}{\left[1+\frac{12k_3^2}{\pi^2}R\right]^2}, \tag{10}$$

where

$$k_3^2 = \frac{\pi^2}{24}\frac{\left\{C_{22}+C_{44}-\left[(C_{22}-C_{44})^2+4C_{24}^2\right]^{1/2}\right\}}{\bar{C}_{44}}. \tag{11}$$

The terms C_{pq} are elastic stiffnesses for AT-cut quartz, and $\bar{C}_{pq}$, and $\hat{C}_{pq}$ are functions of the elastic stiffnesses, namely,

$$\bar{C}_{pq} = C_{pq} - \frac{C_{p2}C_{2q}}{C_{22}}, \tag{12}$$

$$\hat{C}_{pq} = \bar{C}_{pq} - \frac{\bar{C}_{p4}\bar{C}_{4q}}{\bar{C}_{44}}, \tag{13}$$

$$\text{and } \hat{C}_{55} = C_{55} - \frac{C_{56}^2}{C_{66}}. \tag{14}$$

Eqns. (1)–(3) were uncoupled from extensional motions by setting $\bar{C}_{14} = \bar{C}_{34} = C_{56} = 0$. In this paper, we employ the values of elastic stiffnesses given by Bechmann, Ballato and Lukaszek[7].

c) Strain-displacements relations.

The strain-displacement relations are

$$\gamma_1 = u_{2,1} + \psi_1, \tag{15}$$

$$\gamma_3 = u_{2,3} + \psi_3, \tag{16}$$

$$\chi_1 = \psi_{1,1}, \tag{17}$$

$$\chi_3 = \psi_{3,3}, \tag{18}$$

$$\text{and } \chi_5 = \psi_{3,1} + \psi_{1,3}. \tag{19}$$

III. Visualization of Solutions

The post-processing phase of our finite element analysis may be as important as the actual solution phase. It is not only necessary to compute the frequency spectrum for a given plate, but also to study the complex vibration mode shapes which result. For resonator design purposes, it would be very useful if accurate and concise visual data is available to aid in the siting of device mountings, and to study the effects of plate geometry and electrode geometry and its position. Some of this data can be obtained experimentally through the use of stroboscopic X-ray topography, which requires an elaborate and costly laboratory setup.

To visualize the finite element solutions obtained, a two-dimensional technique is employed. Previously, mode shapes were plotted as three-dimensional surfaces with hidden lines removed[1]. In the present study, use is made of a Silicon Graphics Iris workstation to produce 2–D Gouraud shaded images of the vibration mode shapes. A simple FORTRAN program is written which uses v2f and c3f calls to the Iris GL library[8] to construct 2–D polygons with colors specified at each vertex. In our case, the polygons are four-node quadrilateral elements employed in the finite element analysis, and the vertices are the element nodes. The colors are supplied to the i^{th} vertex ($node_i$) via a 3–D vector (R_i,G_i,B_i), where R_i, G_i and B_i denote fractions of red, green, and blue that make up the desired color, namely, each one of the three primary colors can vary from a value of zero to one. To produce a particular image, the nodal displacements from the finite element eigenvalue problem are normalized into the interval [0,1], and the color proportions are computed by mapping this interval into the proper values of R_i, G_i and B_i. Three useful images of mode shapes can be created using this technique:

1. Color images. To produce a linearly graded color mapping ranging from blue representing the minimum displacement to green (zero) and red (maximum), the following transformation is applied to a nodal displacement value $d_i \in [0,1]$:

$$\begin{aligned}
&\text{for } d_i \le 0.25, \begin{cases} R_i = 0 \\ G_i = 4d_i \\ B_i = 1 \end{cases} \\
&\text{for } 0.25 < d_i \le 0.5, \begin{cases} R_i = 0 \\ G_i = 1 \\ B_i = -4d_i + 2 \end{cases} \\
&\text{for } 0.5 < d_i \le 0.75, \begin{cases} R_i = 4d_i - 2 \\ G_i = 1 \\ B_i = 0 \end{cases} \\
&\text{for } 0.75 < d_i, \begin{cases} R_i = 1 \\ G_i = -4d_i + 4 \\ B_i = 0 \end{cases}
\end{aligned} \tag{20}$$

2. Grey scale images. To produce a grey scale mapping with black and white representing minimum and maximum of the displacement values, respectively, the following transformation is applied to the same value d_i:

$$R_i = G_i = B_i = d_i \tag{21}$$

3. Nodal line images. To produce a nodal line image in which the curves representing the zeros of the vibration mode are shown in black, and the rest of the displacement values is represented in light grey or white, the following transformation is applied to d_i:

$$R_i = G_i = B_i = |2d_i - 1|^p \tag{22}$$

where p is a value (< 1) chosen by the user. The smaller the value of p the finer the nodal lines. The value of p used for this analysis is 0.3.

The figures presented in this paper are grey scale and nodal-line images of some of the important vibration modes. The grey scale images are comparable to the experimental X-ray stroboscopic images taken of the resonator. The nodal line images will be useful in the design process because they locate areas where displacements are minimal for mounting purposes. Since each vibration mode has a unique nodal line pattern, these images also may be used to identify quickly the modes of vibration. Color images are not presented here because they cannot be reproduced in a black and white medium.

IV. Frequency Spectrums and Mode Shapes Of Circular Plated And Unplated AT-Cut Quartz Plates

A uniformly impressed electric field over the AT-Cut quartz plate will excite modes that have

1. u_2 odd in x_1 and even in x_3,
2. ψ_1 even in x_1 and even in x_3, and
3. ψ_3 odd in x_1 and odd in x_3.

The resonator can thus be modeled using a quarter plate and imposing the essential boundary conditions $u_2 = \psi_3 = 0$ along the x_3 axis, and $\psi_3 = 0$ along the x_1 axis. This results in a substantial savings in storage and computation as well as reducing the number of modes to be computed within a specified bandwidth. Figure 2 shows a typical finite element discretization with optional electrode plating and mass loading.

(a) 1.9 Mhz Unplated AT-Cut Quartz Resonator.

A 1.9 Mhz unelectroded, circular plate is studied for aspect ratios ranging from 15.000 to 16.000 and vibration frequencies in the neighborhood of the fundamental thickness-shear mode. The problem was modeled with 16,707 displacement degrees of freedom. Figure 3 shows a portion of the frequency spectrum ranging over the aspect ratios of 15.000 to 16.000, and with the natural frequencies normalized by $\overline{\omega} = \frac{\pi}{2b}\sqrt{\frac{C_{66}}{\rho}}$, which is the cutoff frequency of an infinite plate of the same thickness.

The nodal line plot for the fundamental thickness-shear mode (TS-1) of the circular plate with an aspect ratio of 15.500 is shown in Fig.4, and is compared with Fig.5 of a rectangular plate with the same aspect ratio. The mode for the circular plate corresponds to point Aa in Fig.3. The first anharmonic overtone of the fundamental thickness-shear mode (TS-3) for the two plates are also compared in Figs.6 and 7. The mode of Fig.6 corresponds to point Ab of Fig.3. Based on these four figures, we make the following observation:

1. The assumption originally made by Mindlin and Gazis[9] for straight crested wave solutions in rectangular AT-cut plates vibrating in the fundamental thickness shear and its anharmonic overtones is fairly good as seen by the fairly straight nodal line patterns over a large area of the plate. There are some exceptions, namely, for the u_2 displacement of Fig.5 along the boundary at $x_1 = \pm a$, and the ψ_3 shear of Figs.5 and 7 along the boundary at $x_3 = \pm a$.
2. Straight crested wave solutions for circular plates are unrealistic. The nodal lines are neither straight nor axis-symmetric.

(b) 5 Mhz Unplated and Plated AT-Cut Quartz Resonator

A 5 Mhz circular plate with and without electrode plating is studied for aspect ratios ranging from 24.000 to 25.000 in the neighborhood of the fundamental thickness-shear mode (TS-1). A convergence study of a plated circular plate with an aspect ratio of 24.0 was performed for the TS-1 mode (5.07 Mhz) and the TS-3 mode (5.12 Mhz). The results of this study are presented in Fig.8. The frequencies converge monotonically from below.

Figures 9 and 10 show respectively, the frequency spectra for the unplated and plated resonators, ranging over the aspect ratios discussed above. The calculations were performed using an R value of $6.73\text{x}10^{-3}$, a quarter plate and 20,655 displacement degrees of freedom. We observe the two spectra are quite similar, except for the TS-1 and TS-3 branches which are lower in frequency and more pronounced for the plated resonator.

Figure 11 shows a grey-scale image of the TS-1 mode shape for the plated 5 Mhz resonator with an aspect ratio of 24.375. This mode corresponds to point Ca in Fig.10. Shown in Fig.12 is a nodal-line image for the same plate. Figures 13 and 14 are grey-scale and nodal-line images, respectively, for the TS-3 mode which corresponds to point Cb in Fig.10. The u_2 and ψ_3 mode shapes are more intricate than their counterparts in Figs.4 and 6.

Figure 15 exhibits a mode of vibration not previously observed in rectangular plates. We named this mode the edge flexural mode, due to the fact that the strain energy is concentrated along the circular plate edges and the predominant displacement is the transverse flexural displacement u_2 component. This image corresponds to point Cc of Fig.10. The edge flexural mode may be excited by electrode fingers of alternating voltages distributed along the circumference. Such a resonator may have useful applications and its resonant frequency is limited only by the number of electrode fingers along the circumference. Hence, very high resonant frequencies may be obtained. The resonator can be mounted at the center. Unwanted modes may be suppressed by strong damping near the center of the plate.

V. Mass Frequency Influence Surface of a 5 Mhz Plated Circular AT-Cut Quartz plate

The finite element program is employed to construct the mass frequency influence surface for the 5 Mhz electroded circular plate with an aspect ratio of 24.000 and vibrating in the fundamental thickness-shear mode. This surface gives the frequency effect of a small, discrete, mass loading as a function of position on the plate's surface. For a very small mass loading $m(x_1, x_3)$ $\mu g/mm^2$ such that $m(x_1, x_3)/(2b\rho) \sim 10^{-5}$, the frequency change is a linear function of $m(x_1, x_3)$, and it does not matter whether the mass layer is on the top or bottom surface of the plate. The frequency change for the resonator is an algebraic sum of the frequency effects of $m(x_1, x_3)$ over the entire plate. Hence, we define the mass frequency influence surface $F(x_1, x_3)$ $Hz/\mu g$ as a function such that

$$\int_A F(x_1, x_3)\, m(x_1, x_3)\, dA = -\Delta f \qquad (23)$$

where Δf is the change in resonator frequency in Hz, and A the plate area in mm^2.

Figure 16 shows a contour plot of the mass frequency influence surface of the 5 MHz electroded plate. A maximum frequency change of 398 $Hz/\mu g$ is observed just off center along the x_1 axis. There are a few locations along a line passing through the plate center in the x_1 direction that show high sensitivity to mass loading. The area of plate surface which is most sensitive to mass loading is located near the center of the electrodes.

Acknowledgement

The first two authors gratefully the support by the U.S. Army Research Office contract no. DAAL03-91-G-0018 and the National Center for Supercomputing Applications in Urbana-Champaign, Illinois.

References.

1. "Mass Frequency Influence Surface, Mode Shapes, and Frequency Spectrum of a Rectangular AT-Cut Quartz Plate", Y-K Yong and J. T. Stewart, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 38, No. 1, 1991, pp. 67–73.
2. An Introduction To the Mathematical Theory of Vibrations of Elastic Plates, R. D. Mindlin, Monograph prepared for The U.S. Army Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey, 1955.
3. "High Frequency Vibrations of Plated Crystal Plates", R. D. Mindlin, Progress In Applied Mechanics, Macmillan, New York, 1963, pp. 73–84.
4. "Resonator Surface Contamination — A Cause of Frequency Fluctuations ?", Y-K Yong and J. R. Vig, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 36, No. 4, 1989, pp. 452–458.
5. "An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators", Lanczos, C., Journal of Research of the National Bureau of Standards, 45, 1950, pp. 255–281.
6. "The Lanczos Algorithm with Selective Orthogonalization", Parlett, B. N., and Scott, D., Mathematics of Computation, 33, No. 145, 1979, pp. 217–238.
7. "Higher-Order Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha-Quartz," Bechmann, R., Ballato, A.D., and Lukaszek, T.J., Proc. IRE, 50, 196[illegible] pp. 1812–1822.
8. Graphics Library Reference Manual, FORTRAN 77 Edition , Ver. 1.0, Silicon Graphics, Inc., Mountain View, California.
9. "Strong Resonances of Rectangular AT-Cut Quartz Plate," Mindlin, R., and Gazis, D.C., Proc. of the 4th U.S. Nat. Congress of Appl. Mech., Vol.1, 1962, pp. 305–310.

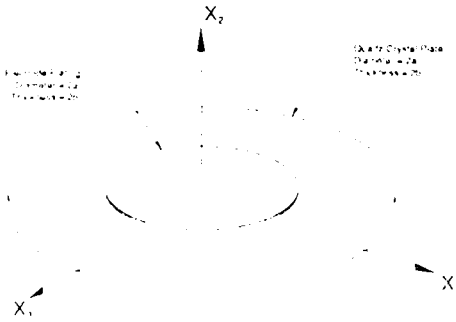


Fig.1 Orientation and dimensions of a circular electroded plate.

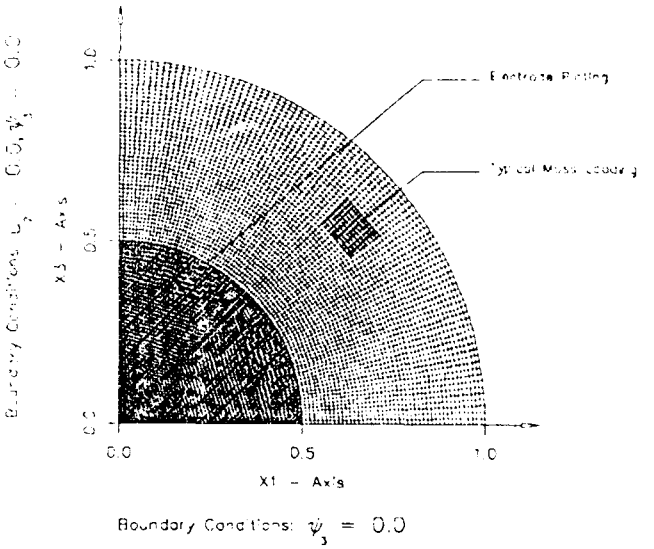


Fig.2 Finite element mesh of quarter circular plate with essential boundary conditions at the plate edges along the coordinate axes.

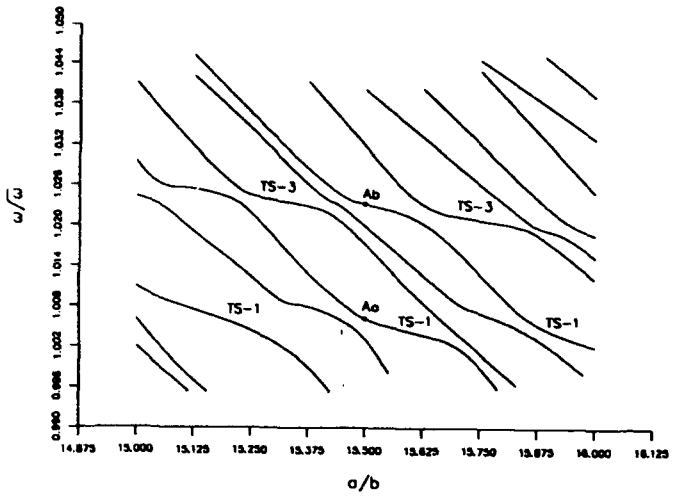


Fig.3 Frequency Spectrum for a 1.87 MHz AT-Cut Plate Without Electrodes (thickness=0.883 mm).

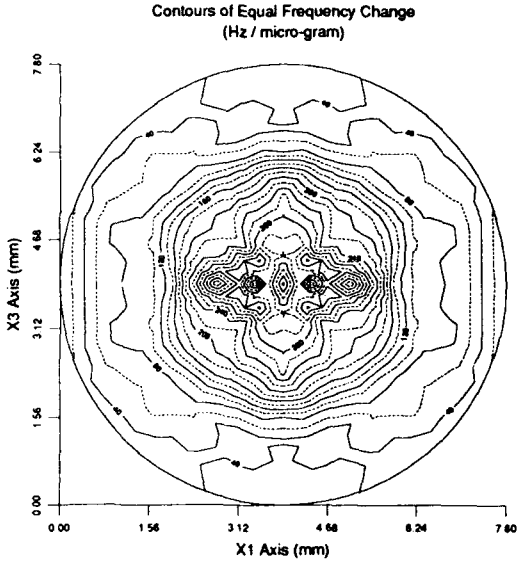


Fig.16 Mass Frequency Influence Surface For a plated 5 Mhz AT-Cut Circular Plate (thickness=0.325 mm, plate diameter=7.8 mm, electrode diameter=3.9mm).

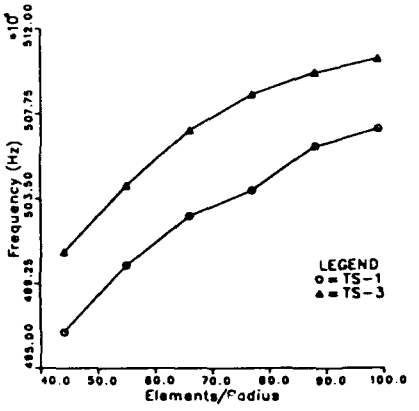


Fig.8 Convergence study for an electroded quarter circular plate.

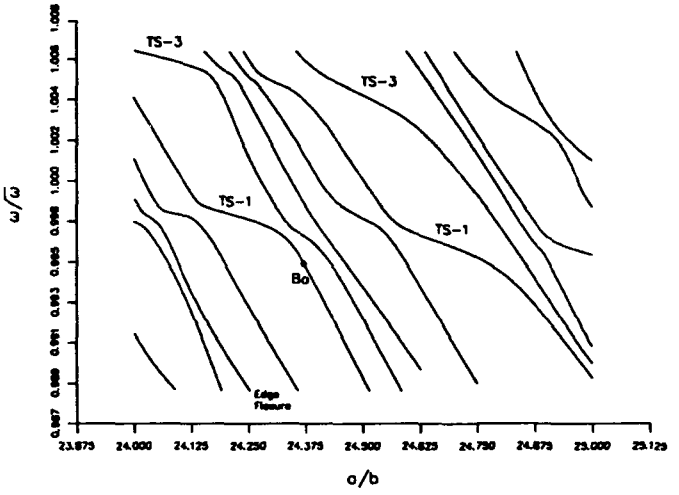


Fig.9 Frequency Spectrum for the unplated 5 Mhz AT-Cut Plate (thickness=0.325 mm).

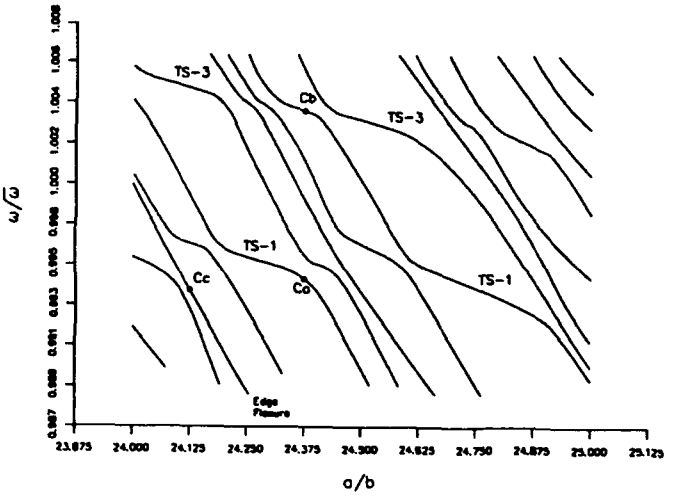
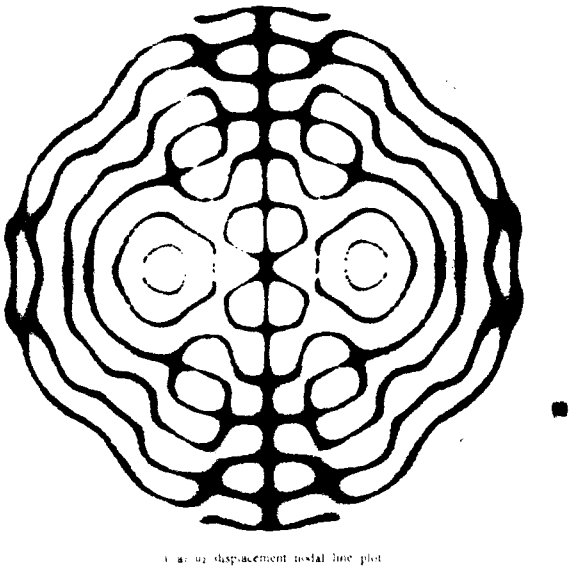


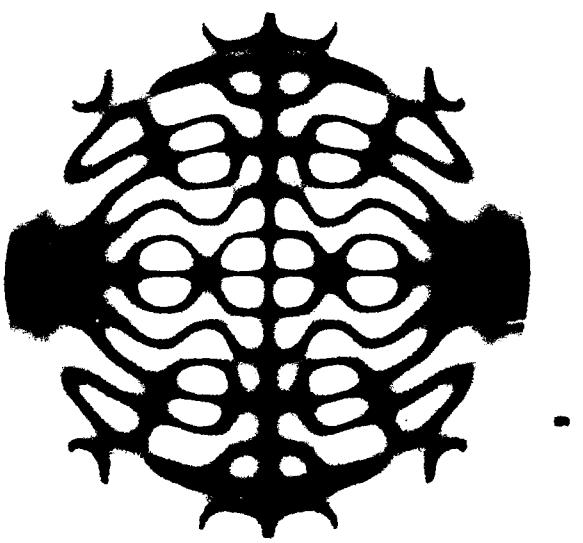
Fig.10 Frequency Spectrum for the plated 5 Mhz AT-Cut Plate (thickness=0.325 mm,electrode diameter=plate diameter/2.0).



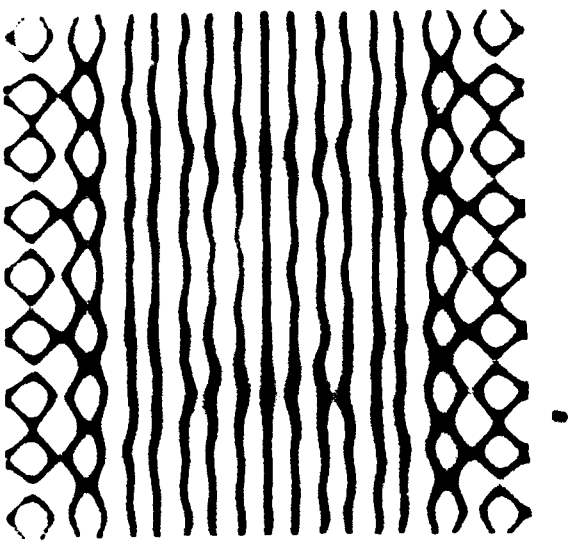
(a) u_3 displacement modal line plot



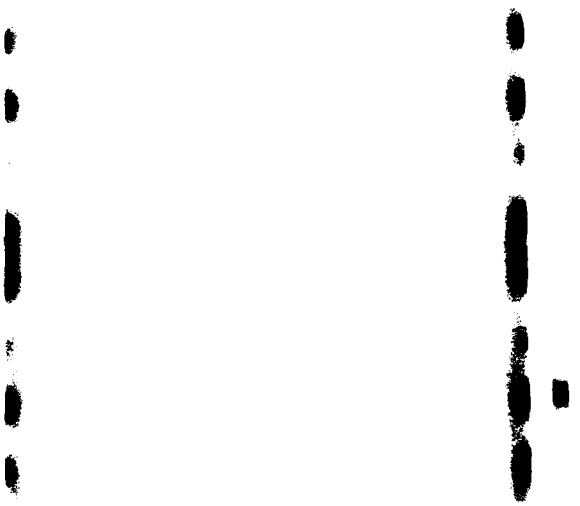
(b) ψ_1 displacement modal line plot



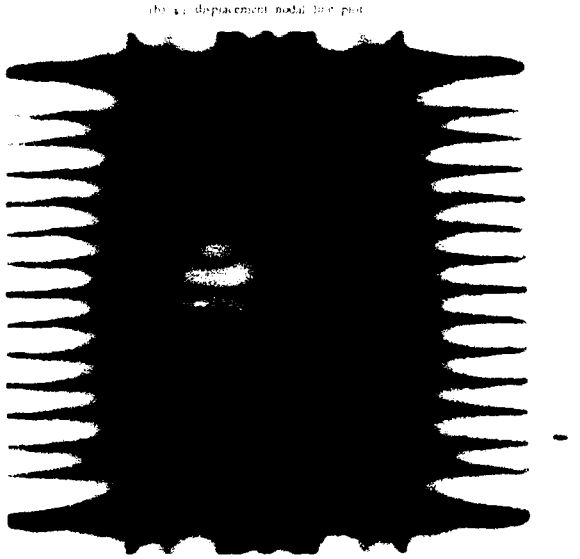
(c) ψ_2 displacement modal line plot



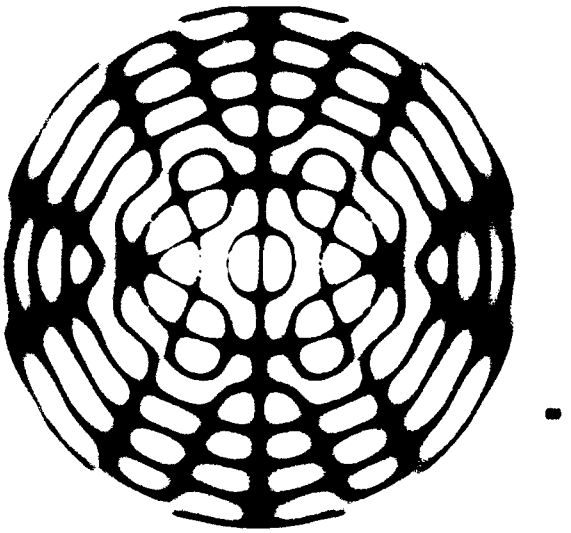
(a) u_3 displacement modal line plot

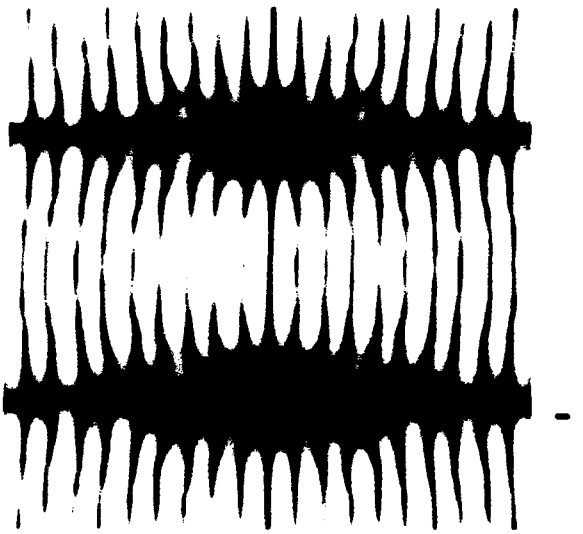


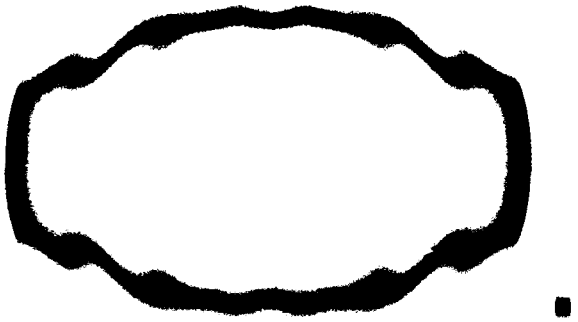
(b) ψ_1 displacement modal line plot

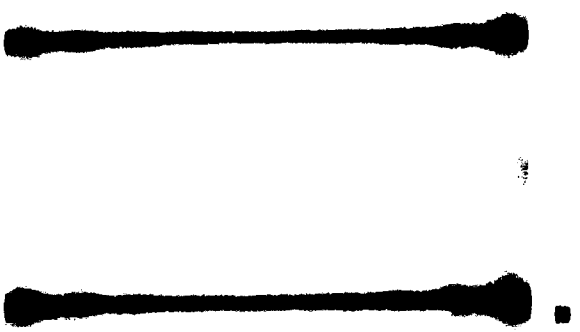


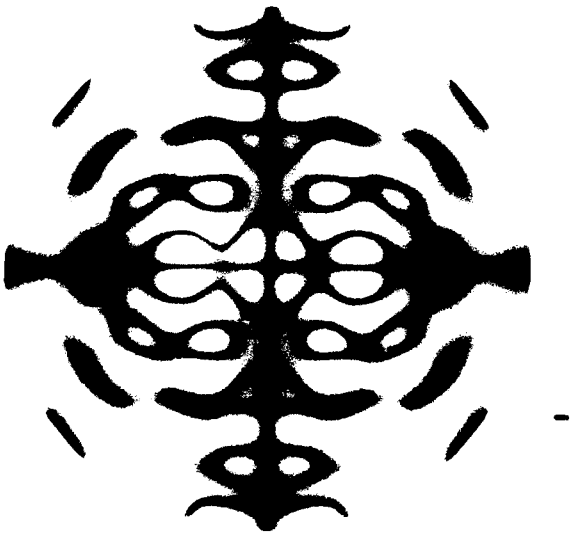
(c) ψ_2 displacement modal line plot



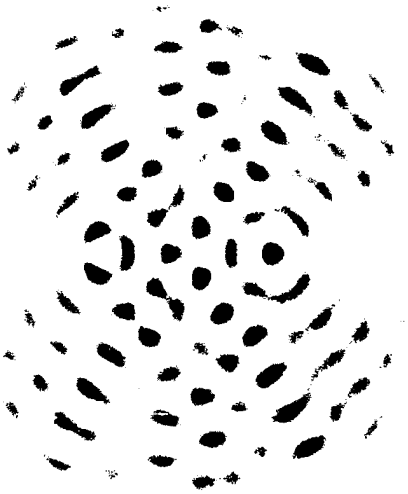


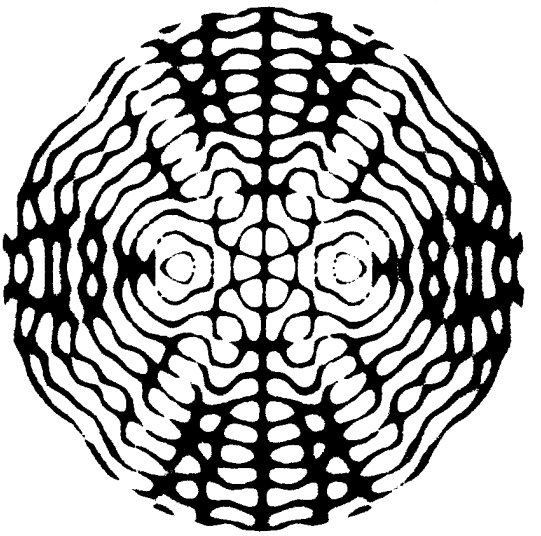


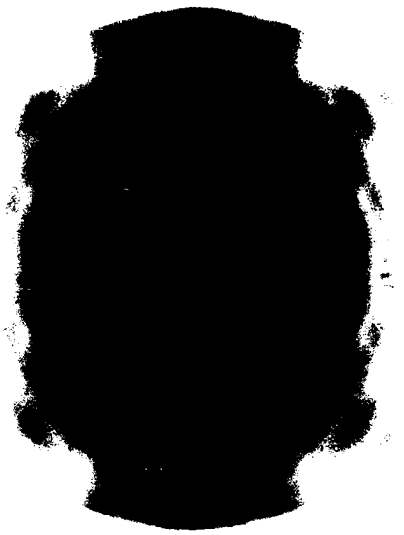














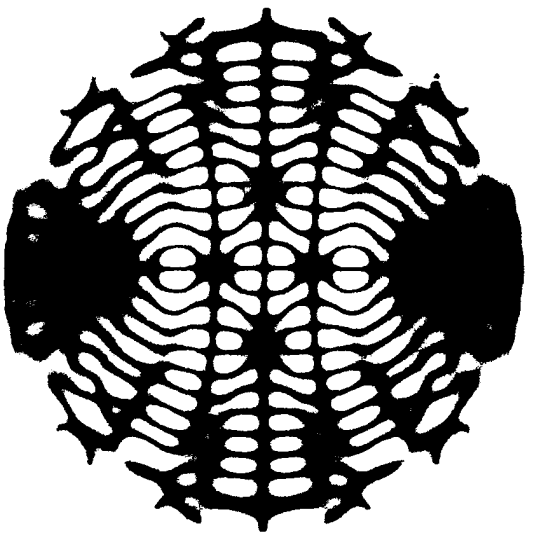
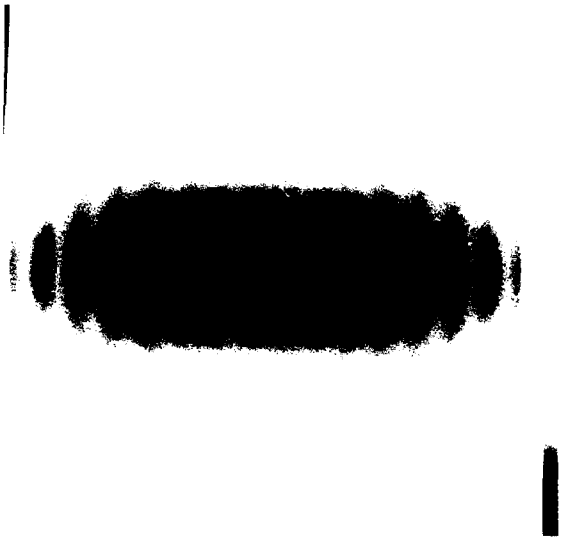


Fig. 14 Grey scale plot of (a) u_3, (b) v_3 and (c) w_3 displacements of the first antisymmetric overtone of fundamental thickness shear mode for the circular plate with electrodes, $a/h=24.4$ [8], frequency 8.11 MHz



(a) u_3 displacement grey scale plot

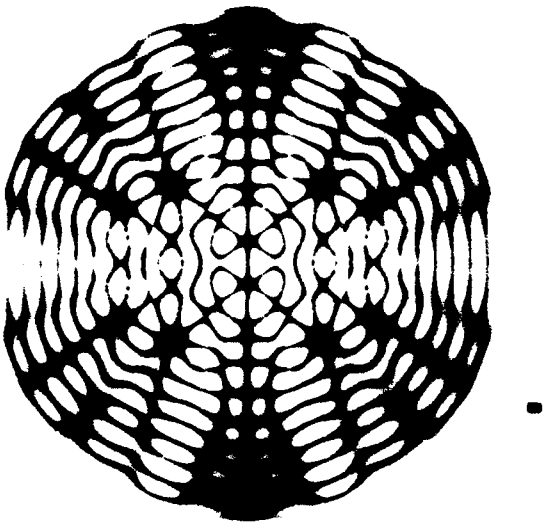


(b) v_3 displacement grey scale plot

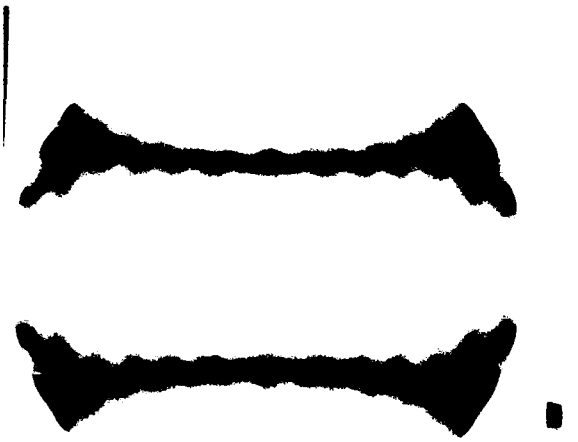


(c) w_3 displacement grey scale plot

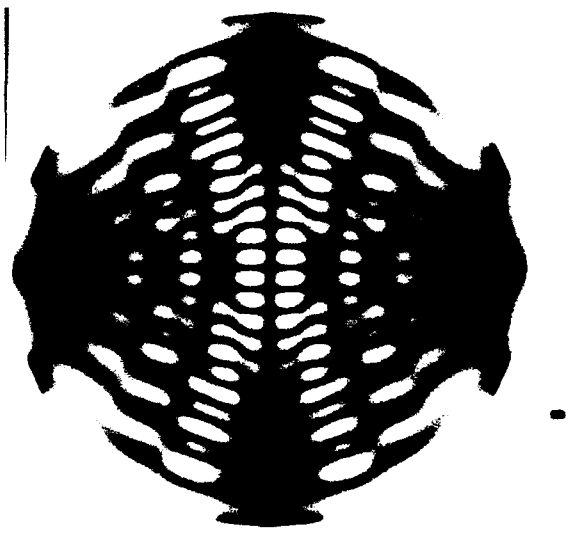
Fig. 14 Nodal line plots of (a) u_3, (b) v_3 and (c) w_3 displacements of the first antisymmetric overtone of fundamental thickness shear mode for a circular plate with electrodes, $a/h=24.4$ [8], frequency 8.11 MHz



(a) u_3 displacement nodal line plot

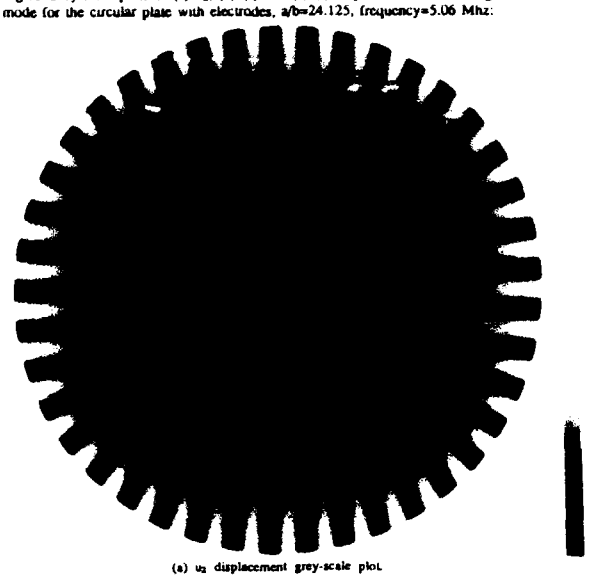


(b) v_3 displacement nodal line plot

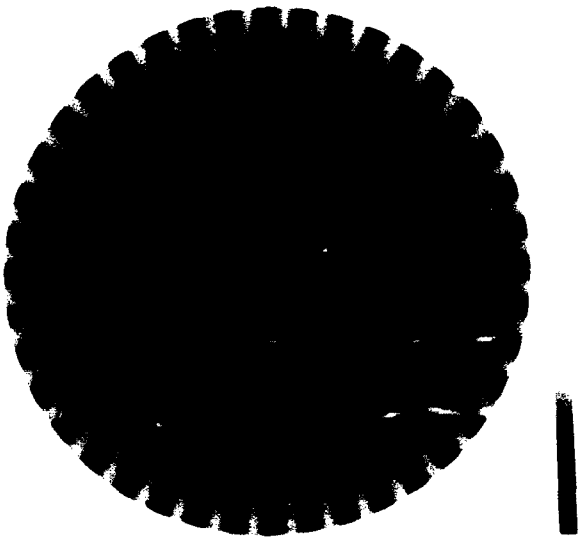


(c) w_3 displacement nodal line plot

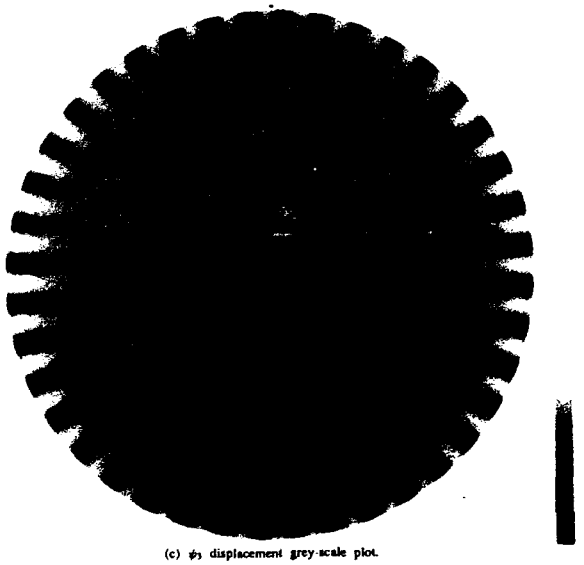
Fig.15 Grey-scale plots of (a) u_2, (b) ψ_1 and (c) ψ_3 displacements of the edge flexural mode for the circular plate with electrodes, a/b=24.125, frequency=5.06 Mhz:



(a) u_2 displacement grey-scale plot.



(b) ψ_1 displacement grey-scale plot.



(c) ψ_3 displacement grey-scale plot.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A Simple Model for Quartz Resonator Low Level Drive Sensitivity

Lawrence Dworsky
Robert Kinsman

Motorola, Inc.

Abstract

The nature of the observed phenomenon of "starting resistance" is reviewed, along with the accompanying experimental evidence of high drive-level altering and sometimes "curing" of the malady. A simple phenomenological model is proposed, and then it is shown that this model does indeed predict all of the observed properties of starting resistance with the exception of high drive curing. Device changes due to high drive which would account for the observed characteristics and fit within the assumptions of the model are contemplated.

Introduction

The phenomenon of drive power-level sensitivity of quartz resonators has been recognized for several decades. It has been described by many names, Starting Resistance, Sleeping Sickness and Second Level of Drive being just a few. The problem relates to the fact that the parameters of a linear device should not vary with excitation, or drive, level. Instead, it was often found that, at drive levels too low to normally produce non-linear elastic phenomena, the "motional" resistance and series resonance frequency were a function of the drive level.

The application problems caused by starting resistance are well known to oscillator engineers. It is these engineers who probably coined the very descriptive expression "starting resistance" in the early days of electronic circuit development. Typically, they found that an oscillator which had more than adequate gain to operate with the expected crystal resistance did not start or "turn on." The problem occurred both after repeated use of a crystal and with a new crystal which exhibited low enough resistance at operating drive levels. Investigation of the problem showed either an increase in crystal resistance as the drive level was lowered or a peak in resistance at some specific drive level. If the (increased) resistance was high enough, the oscillator circuit gain was inadequate and oscillation would not occur. Marginal crystal units were particularly frustrating in that they may or may not oscillate at any given time. The usual "cure" in this case was to switch the power on and off until oscillation started. Sometimes, driving the crystal at a level many times the designed operating level, or "high driving", would "cure" the problem; sometimes high driving would have no effect. Often the high drive cure was permanent, sometimes (particularly when the crystal is not operated for a period of time) the problem returned (hence the descriptor "sleeping sickness".

The purpose of this paper is to briefly review the literature in this area, and then to suggest a single physical model which can account for all of the observed phenomena. To this latter end, a simple, phenomenological, model which couples a sliding small mass to the main resonator vibrating mass is proposed. Non-linearity is introduced in terms of inelastic collisions of these masses and deformations of the small mass. Experimental data is included and it is shown that this simple model does indeed replicate all of the observed situations except for the "high drive cure." Mechanisms for this latter case which are consistent with the model are suggested.

Historical Background

The problems caused by crystal starting resistance were a common experience to early radio users. Bottom reports that the pressure-mount, unsealed, crystals manufactured before and during WW II were often found to be "dead", particularly after a storage period[1]. It was apparently routine practice at that time to periodically open the crystal and clean the quartz wafer with soap and water or a

CH2965-2/91/0000-148 $1.00 © 1991 IEEE

solvent. It was also sometimes found that the problem could be cured, at least temporarily, by driving the crystals at very high levels. In 1954, Gerber published results which showed that the problem existed in plated as well as pressure-mount crystals [2]. In his paper, Gerber described a process in which drive sensitive crystals were reworked by removing their metal electrodes, cleaning them, and then etching them using ion bombardment just prior to replating them. The reworked crystals were found to behave linearly and the theory was then proposed that the drive sensitivity was caused by poor adhesion of the electrodes to the quartz surface.

In a landmark 1967 paper, Bernstein described experiments with crystals which had been coated with small irregular particles held in place by a film of oil or other sticky material[3]. These crystals were found to be very drive sensitive. When the oil film was removed with an appropriate solvent, the drive sensitivity disappeared. Also, when clean crystals were covered with particles, no significant drive sensitivity was found. The control of oil films was an industry wide problem at that time due to back streaming problems caused by the oil diffusion vacuum pumps which were commonly used. Bernstein also stated that "an etch of the finished lapped surface is mandatory to prevent the low power increased resistance effect" and presented data comparing crystals prepared with and without etching.

In 1971, Nonaka, Yuuki, and Hara were able to correlate drive sensitivity with a deep scratch in the active area of the (metal) electrode on a crystal[4]. They also showed results data for a drive sensitive crystal which had a single gold particle stuck to the surface of the electrode. After driving the particle off at a high power level, the crystal was retested and no drive sensitivity was found.

In 1974, Knowles presented a series of optical and SEM photographs showing the presence of loose quartz particles imbedded in the surface of crystal wafers which had not been deeply etched[5]. He proposed that these particles could break loose over time, thereby causing a drive sensitive condition to randomly occur after the crystal was fabricated.

Bottom disclosed the results of his work using extremely small (micron and sub-micron) particles in his 1983 paper[1]. He showed that he could create drive sensitivity in an unsealed crystal by blowing talc from a distance of several feet. Bottom proposed that these very small particles were held in place on the crystal surface by "dipole-dipole" attractive forces which may be equal to many times the weight of the particles. In the case of larger particles, the attractive forces were much smaller (than the weight of the particle) and the resultant drive sensitivity was minimal.

The conclusion which can be reached from these various experiments is that the low-level, non-linear characteristics of crystal resonators are caused by trapped particles on the surface of the quartz and/or (pieces of) electrodes which are free to move when the crystal is excited.

Experimental Results

An examination of experimental data for a variety of drive sensitive crystals reveals two very distinctive behavior patterns. One type of behavior is an essentially monotonic increase in motional resistance, to a maximum plateau level, with decreasing drive level. The other type of behavior is that of a sharp resistance peak at a specific drive level. Data for this second type of crystal is shown in the paper by Nonaka, et al[4]. The resistance peak of the second type of crystal is also often found to exhibit a hysteresis effect, i.e. the level of drive at which the peak occurs varies depending on whether the measurment starts at a low power level or at a high power level. No particular correlation between the drive sensitivity pattern and the suspected cause of the non-linear performance has been established to date.

We have found examples of the first type of behavior in crystals which were fabricated using quartz blanks which received a very limited amount of etching after final lapping. Resistance and frequency data for a typical unit is shown in Figure 1. These particular crystals were third overtone resonators operating in the 45 to 60 MHz range. They were fabricated using unpolished round blanks and were packaged in HC-18 holders. Other units of the same design which were prepared using a longer etch time did not exhibit the same non-linear behavior.

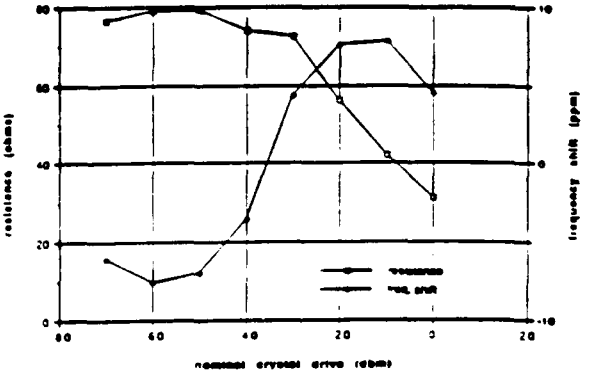


Figure 1. Resistance and Frequency vs Drive Level of a Poorly Etched Crystal

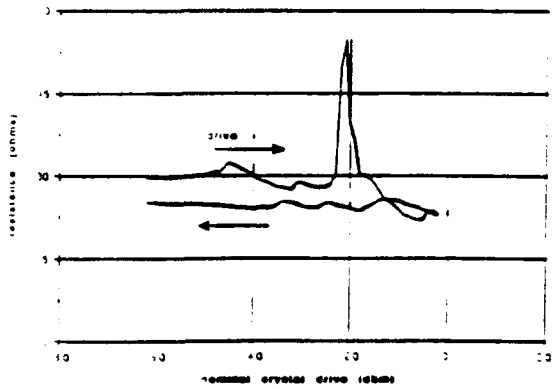


Figure 2. Resistance vs Drive Level of a Crystal with (Suspected) Loosely Attached Particles

An example of the second type of behavior is shown in Figures 2 and 3. The crystal in this case is also a 3rd overtone resonator in an HC-18 package. The exact cause of the non-linear performance has not been identified. This crystal exhibits reasonably linear behavior except for a very narrow drive level region in the vicinity of -20 dbm. It is interesting to note that the non-linear behavior only occurs when the crystal drive is increasing.

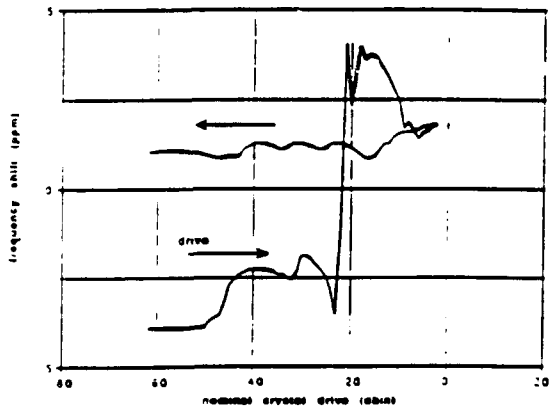


Figure 3. Frequency Shift vs Drive Level of a Crystal wit. (Suspected) Loosely Attached Particles

Another aspect of the non-linear performance is shown in figures 1 and 3. Here it is seen that the resonance frequency (series resonance in this case) also changes with drive level. It is this effect that has led some workers in the field to generalize the term "starting resistance" to the more encompassing "drive level dependency."

Empirical Model

In this section we will demonstrate that the two types of starting resistance variation (Figures 1 and 2) and the starting frequency variation (Figures 1 and 3) can be predicted by a single physical model which in turn can be plausibly related to the experimentally observed causes of these phonema.

Consider a quartz crystal resonator, crudely, as a single mass M_1, spring K_1, and an internal loss mechanism R_1. The resonator is driven by a force F(t). This system (Figure 4) is described by the equation

$$F(t) = M_1 \frac{d^2u_1}{dt^2} + R_1 \frac{du_1}{dt} + K_1 u_1 \qquad [1]$$

where u_1 is the displacement of the mass from its equilibrium position.

Equation [1] can equivalently be written in terms of the mass velocity, $v = du_1/dt$

$$F(t) = M_1 \frac{dv}{dt} + R_1 v + K_1 \int v dt \qquad [2]$$

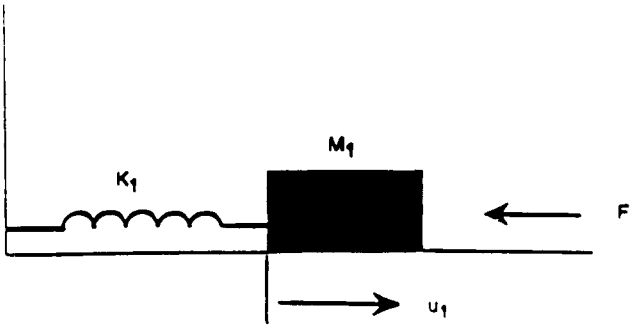


Figure 4. Simple Spring -Mass Mechanical Resonator

Equation [2] also describes a series R-L-C circuit, with $L = M_1$, $C = 1/K_1$, $R = R_1$, and current v. Note that L and C are not the equivalent motional parameters of the resonator, but merely parameters that parrot the simple mechanical model.

Now, consider the situation where a small mass M_2 is somehow caught in a crevice in M_1 (Figure 5). As M_1 moves, M_2 will somehow be "dragged" along. In other words, there will be both a restoring force $K_2(u_2 - u_1)$ and a frictional force $R(v_2 - v_1)$. Figure 5 also shows a second case which would be described by similar relations - a small mass M_2 pivoting at the end of an elastic arm. This situation crudely describes a ragged electrode edge caused by, say, a scratch.

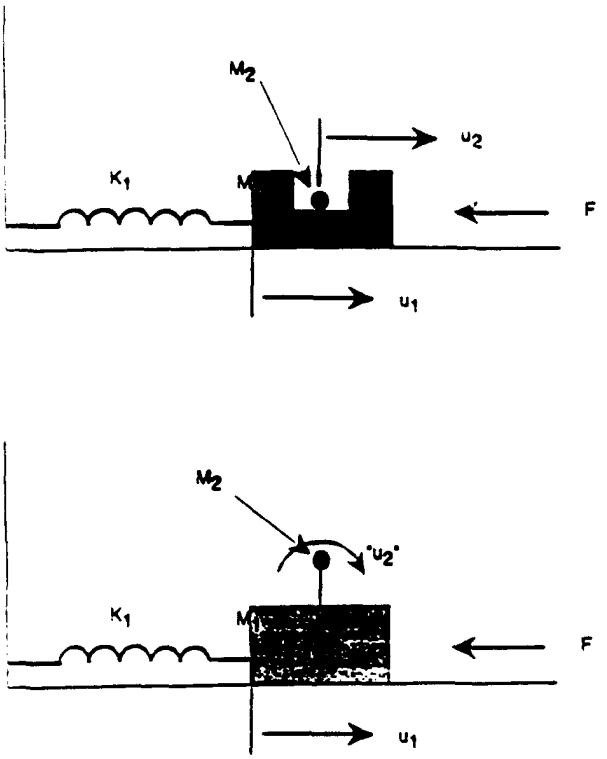


Figure 5. Coupled (2) Resonator Systems:
a. Particle in Crevice on Surface
b. Loosely Attached Particle on Surface

The total situation is described by the set of equations

$$F(t) = M_1\frac{d^2u_1}{dt^2} + R_1\frac{du_1}{dt} + K_1u_1 + K_2(u_1 - u_2) + R\frac{d(u_1 - u_2)}{dt} \qquad [3]$$

$$0 = M_2\frac{d^2u_2}{dt^2} + K_2(u_2 - u_1) + R\frac{d(u_2 - u_1)}{dt} \qquad [4]$$

These equations also describe the electrical circuit of Figure 6.

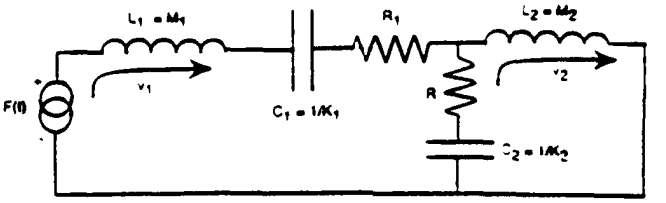


Figure 6. Equivalent Circuit of 2 Resonator System

Since an actual crevice has a finite width, the particle may be expected to experience inelastic (and probably deforming) collisions with the crevice walls at high drive levels. This may be modeled by having the resistance R be a function of $|u_2 - u_1|$ which rises suddenly and sharply at some threshold level. In a real device there are many particles, and also many crevices, all having different parameters. The aggregate function R may therefore be thought of as being a monotonically increasing function of $|u_2 - u_1|$. A suitable function for R would be of the form

$$R = R_0\left[1 + \left(\frac{u_2 - u_1}{g}\right)^n\right] \qquad [5]$$

where R_0 is the low (drive) level frictional resistance, 2g is the width of the crevice and n is an even integer.

Equations [4] and [5] then become

$$F(t) = M_1\frac{d^2u_1}{dt^2} + R_1\frac{du_1}{dt} + K_1u_1 + K_2(u_1 - u_2) + R_0\left[1 + \left(\frac{u_1 - u_2}{g}\right)^n\right]\frac{d(u_1 - u_2)}{dt} \qquad [6]$$

$$0 = M_2 \frac{d^2 u_2}{dt^2} + K_2 (u_2 - u_1)$$

$$+ R_0 \left[1 + \left(\frac{u_2 - u_1}{g} \right)^n \right] \frac{d(u_2 - u_1)}{dt} \qquad [7]$$

Equations [6] and [7] are highly non-linear, and difficult (if at all possible) to solve. However, considerable insight to the (non-linear) steady state behavior may be had by noting that we would expect R to essentially be R_0 at "low enough" drive level, and to increase (on a time average) with increasing drive level.

The input impedance of the circuit of Figure 6 is determined by the series R_1 - L_1 - C_1 (original linear resonator) parameters and the parameters of the non-linear (particle + coupling) impedance. By inspection, this latter impedance is given by:

$$Z = (j\omega L_2) \,||\, (R - \frac{j}{\omega C_2}) = \frac{j\omega L_2 (R - \frac{j}{\omega C_2})}{j\omega L_2 + R - \frac{j}{\omega C_2}} \qquad [8]$$

Separating the real and imaginary parts,

$$Re(z) = R_n = \frac{R \left(\frac{\omega}{\omega_2} \right)^4}{(\omega R C_2)^2 + \left(\left(\frac{\omega}{\omega_2} \right)^2 - 1 \right)^2} \qquad [9]$$

and

$$Im(z) = X_n = \frac{R^2 C_2 \omega \left(\frac{\omega}{\omega_2} \right)^2 - \omega L_2 \left(\left(\frac{\omega}{\omega_2} \right)^2 - 1 \right)}{(\omega R C_2)^2 + \left(\left(\frac{\omega}{\omega_2} \right)^2 - 1 \right)^2}$$

where

$$\omega_2 = \frac{1}{\sqrt{L_2 C_2}} \qquad [10]$$

is the L_2C_2 resonance frequency.

Consider an example: Set $\omega = 1$, $L_2 = .5$, $C_2 = 1$. Note that, in this case, $\omega < \omega_2$. Figure 7 shows R_n as a function of R. In this figure R begins at 0. In the actual model, R begins at R_0. This means that, depending upon the value of R_0, the observed resistance vs. drive level could either:

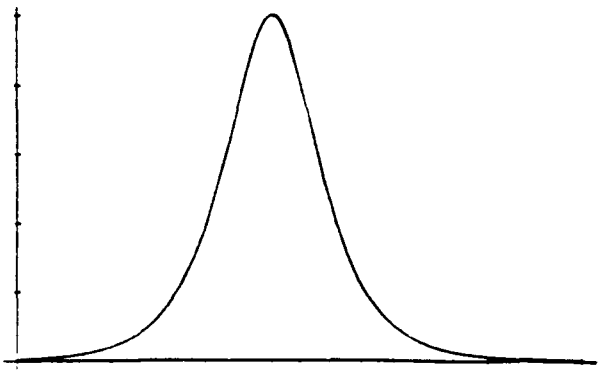


Figure 7. R_n vs Log(R)

Start low, increase, and then decrease (R_0 small), or,

Start high and then increase (R_0 near the peak of Figure 7), or

Start low and stay low (R_0 high).

Interestingly enough, a high R_0, which we would intuitively assume to be due to a very lossy system, would produce a device with no observable starting resistance.

It would appear that all 3 cases above, which correspond to the observed cases, are derivable from the same physical system - with only 1 adjustable parameter needed.

Continuing, Figure 8 shows X_n as a function of R. Note that X_n is positive (net inductive reactance) and decreases with increasing R - over the same range of R where R_n is varying. Since X_n appears in series with the L_1 - C_1 arm of the circuit, the observed resonance frequency will increase with increasing R.

While the above observations represent, in some sense, expected circuit observations, they are based upon the assumption that R will, again in some sense, increase with increasing drive level. In order to justify this assumption we must solve, or at least approximately solve, the set of non-linear equations in the sinusoidal steady state. The

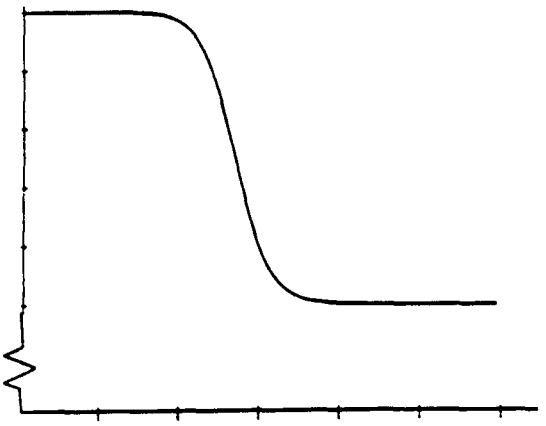


Figure 8. X_n vs Log(R)

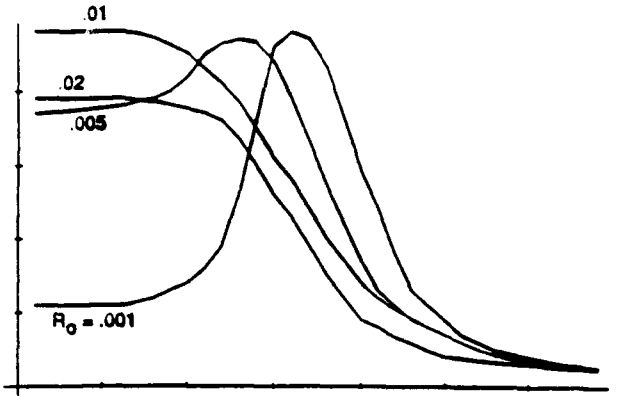


Figure 9. R_{in} vs Log(Drive Level)

details of this solution are given in the appendix.

Figure 9 shows the results of the solution as derived in the appendix, for the parameters shown. Since R_1 simply adds to the input resistance, it is not shown - this is why R_{in} can approach 0 in the figures. As may be seen, for the lower values of R_0 we see one "signature" of starting resistance (resistance starting low, peaking, and then dropping), while for the higher values of R_0 we see the other signature (resistance starting high and then dropping). Also, note that as R_0 gets large, the starting resistance falls. When $R_0 \sim .1$, (not shown in Figure 9), essentially no variation of R_{in} with drive level occurs. Apparently when there is enough loss in the M_2 system, very little (relative) motion occurs and there are no observed variations with drive level.

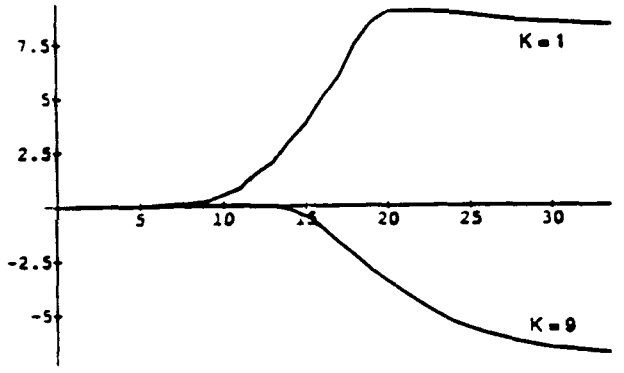


Figure 10. dF/F vs Log(Drive Level)

Figure 10 shows the frequency change (arbitrary units) for the same parameters as (one of the cases of) Figure 9, and also for the same case with K_1 changed from 1 to 9. This simulates looking at the same quartz resonator at its fundamental and 3rd harmonic overtone responses. Note that the frequency changes are in different directions.

Figure 11 shows the input resistance for the same choice of parameters as in figure 10. Note that a much higher starting resistance peak at the 3rd harmonic overtone than at the fundamental frequency, at approximately the same drive level, is predicted.

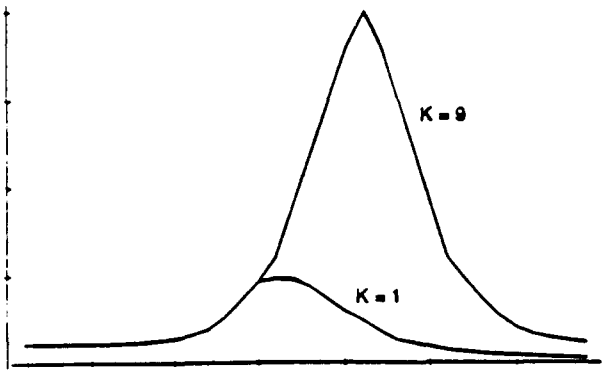


Figure 11. R_{in} vs Log(Drive Level), $R_0 = .001, K = 1, 9$

Summary and Discussion

The simple model described above predicts all of the observed resistance and frequency variation forms seen in actual devices, with only one adjustable parameter required. While this is in no way a proof that the observed phenomena are indeed explained by the model, the results are somewhat compelling.

Missing from the model are the results of high-drive. However, within the framework of the model and experimental evidence we may look for the following possible scenarios:

In some cases, loose particles on the surface and/or poorly attached electrode segments are blown away. If these particles are not contained within a package, the cure should be permanent.

Poorly attached electrode segments are welded down to the adhering electrode by the high localized heat generated by high driving the crystal. The cure might be permanent, but thermal and/or mechanical disturbances could free the segments.

Loose particles in surface crevices are driven into the electrode metal, from the underside, with sufficient force to "stick." Again, the cure might be permanent, but thermal and/or mechanical disturbances could free the particles.

References

1. V. E. Bottom, "The Behavior of Quartz Resonators at Low Drive Levels." Proc. 5th Quartz Crystal Conference, pp. 245-249, August 1983.

2. E. A. Gerber, "VHF Crystal Grinding." Electronics, Vol. 27, #3, pp. 161-163, March 1954.

3. M. Bernstein, "Increased Crystal Unit Resistance at Oscillator Noise Levels." 21st Frequency Control Symposium Proceedings, pp. 244-258, April 1967.

4. S. Nonaka, T. Yuuki, K. Hara, "The Current Dependency of Crystal Unit Resistance at Low Drive Level." 25th Frequency Control Symposium Proceedings, pp. 139-146, April 1971.

5. J. E. Knowles, "On the Origin of the 'Second Level of Drive' Effect in Quartz Oscillators." 29th Frequency Control Symposium Proceedings, pp. 230-236, May 1975.

Appendix: Non-linear Steady State Solution

We wish to solve equations [6] and [7] for the steady state response to a sinusoidal drive signal. An approximate solution may be had by employing the weighted residual method. That is, we solve the equations

$$0 = \int_0^{\frac{2\pi}{\omega_1}} W(t) \left[-F_1 + M_1 \frac{d^2 y_1}{dt^2} + R_1 \frac{dy_1}{dt} + K_1 y_1 + R_0 \{ 1 + (\frac{y_2 - y_1}{g})^n \} \frac{d(y_1 - y_2)}{dt} \right] dt \qquad [11]$$

$$0 = \int_0^{\frac{2\pi}{\omega_1}} W(t) \left[M_2 \frac{d^2 y_2}{dt^2} + K_2 (y_2 - y_1) + R_0 \{ 1 + (\frac{y_2 - y_1}{g})^n \} \frac{d(y_2 - y_1)}{dt} \right] dt \qquad [12]$$

where

ω_1 = driving frequency,

$F(t) = F_0 sin(\omega_1 t)$ = forcing function,

W(t) = weighting function(s),

y_1 and y_2 = approximations to the solutions (u_1 and u_2).

The Galerkin form of the weighted residual technique consists of using the trial functions themselves as the weighting functions. In the sinusoidal steady state, this is very similar to the "harmonic balance" technique.

Since we know that, in the steady state, currents will flow only at integer multiples of ω_1, we let

$$y_1 = \sum_i (A_i cos(i\omega_i t) + B_i sin(i\omega_i t)) \quad [13]$$

$$y_2 = \sum_i (C_i cos(i\omega_i t) + D_i sin(i\omega_i t)) \quad [14]$$

where i is taken from 1 (inspection of the circuit shows that there can be no D.C., or i = 0, current) to a "reasonable" upper limit.

In this situation, since we are dealing with a very high Q circuit and would not expect significant harmonic currents to flow, we take the "0th" order approximation - that is, only allow i = 1:

$$y_1 = A_i cos(i\omega_i t) + B_i sin(i\omega_i t) \quad [15]$$

$$y_2 = C_i cos(i\omega_i t) + D_i sin(i\omega_i t) \quad [16]$$

We next substitute equations [15] and [16] into [13] and [14]. Letting $W = cos(\omega_1 t)$ gives us 2 equations, and then letting $W = sin(\omega_1 t)$ gives us 2 more equations, in A, B, C, and D. Relying on our single frequency approximation, we drop all terms at frequencies other than ω_1. Without showing the algebra, thia gives

$$0 = -\frac{F_0}{\omega_1} + R_1 A + \left(\omega_1 M_1 - \frac{K_1}{\omega_1}\right) B + R_0 \alpha - \frac{K_2 \beta}{\omega_1} + \frac{R_0(\alpha^3 + \alpha\beta^2)}{4g^2} \quad [17]$$

$$0 = \left(\omega_1 M_1 - \frac{K_1}{\omega_1}\right) A - R_1 B - \frac{K_2 \alpha}{\omega_1} - R_0 \beta - \frac{R_0(\beta^3 + \alpha^2\beta)}{4g^2} \quad [18]$$

$$0 = \omega_1 M_2 D - R_0 \alpha + \frac{K_2 \beta}{\omega_1} - \frac{R_0(\alpha^3 + \alpha\beta^2)}{4g^2} \quad [19]$$

$$0 = \omega_1 M_2 C + \frac{K_2 \alpha}{\omega_1} + R_0 \beta + \frac{R_0(\beta^3 + \alpha^2\beta)}{4g^2} \quad [20]$$

The above (4) equations are a non-linear set which may be solved by the Newton-Raphson iteration technique: Write the equations as

$$0 = \bar{F}(\bar{x}) \quad [21]$$

where

$$\bar{x} = \{A, B, C, D\} \quad [22]$$

Given a starting value for $\bar{x}$, iterate using

$$(\bar{x})_{new} = (\bar{x})_{old} + \bar{y} \quad [23]$$

where $\bar{y}$ is found by solving the linear equation set

$$\bar{\bar{J}}\bar{y} = -\bar{F} \quad [24]$$

with

$$\bar{\bar{J}} = \begin{bmatrix} f_{1,1} & f_{1,2} & f_{1,3} & f_{1,4} \\ f_{2,1} & f_{2,2} & f_{2,3} & f_{2,4} \\ f_{3,1} & f_{3,2} & f_{3,3} & f_{3,4} \\ f_{4,1} & f_{4,2} & f_{4,3} & f_{4,4} \end{bmatrix} \quad [25]$$

and

$$f_{i,j} \equiv \frac{\partial f_i}{\partial x_j} \quad [26]$$

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

TWO-DIMENSIONAL EQUATIONS FOR GUIDED EM WAVES IN DIELECTRIC PLATES SURROUNDED BY FREE SPACE

P.C.Y. Lee and J.S. Yang

Department of Civil Engineering & Operations Research
Princeton University, Princeton, NJ 08544

Abstract

Two-dimensional governing equations of successively higher-order approximations for guided EM waves in an isotropic dielectric plate surrounded by free space are deduced from the three-dimensional Maxwell's equations by expanding the EM vector potential in a series of trigonometric functions of thickness coordinate and in exponentially decaying functions of thickness coordinate in the upper and lower halves of free space. A single system of governing equations is obtained by further satisfying the continuity conditions of the EM field at the interfaces between the plate and free space.

Solutions and dispersion relations are obtained from the two-dimensional approximate equations. Dispersion curves are computed and compared with the corresponding ones obtained from the solutions of the three-dimensional Maxwell's equations for the TE and TM modes of various orders and for different values of the refractive index $\hat{n} = (\frac{\epsilon}{\epsilon_0})^{\frac{1}{2}}$. It is shown that the agreement between the approximate and exact dispersion curves is very close.

I. Introduction

Solutions of straight-crested waves propagating in an isotropic dielectric plate of infinite extent and surrounded by free space is one of the very few simple closed form solutions of the three-dimensional Maxwell's equations.[1,2] However, exact closed form solutions of the free vibrations in finite dielectric plate resonators, rectangular or circular, have not yet been found in published articles.

One of the effective methods to facilitate a systematic study of the vibrations in plate resonators is, first, to derive a system of two-dimensional approximate, governing equations and, then. to solve these equations for closed form, analytical or numerical solutions. This kind of approach has been employed in deriving the equations of motion for elastic plates from the three-dimensional equation of elasticity for a long time and by many investigators, for instance, the classical Poisson's equations for extensional vibrations of elastic plates,[3] Cauchy's equations of flexural vibrations of crystal plates,[4] Mindlin's two-dimensional equations of motion of elastic[5] and crystal plates,[6] and Lee's two-dimensional equations of motion of successively higher-order approximations for piezoelectric crystal plates.[7]

In the present paper, two-dimensional governing equations of successively higher-order approximations for guided EM waves in an isotropic dielectric plate surrounded by free space are deduced from the three-dimensional Maxwell's equation's by expanding the EM vector potential in a series of trigonometric functions of thickness coordinate in the plate (similar to the functions employed in Ref. 7) and in exponentially decaying functions of the thickness coordinate in the upper and lower halves of free space. These three sets of equations for the plate, the upper and lower half free spaces are, then, combined into a single system of equations by satisfying the continuity conditions of the EM field at the interfaces of the plate and free space.

To examine the accuracy of presently derived two-dimensional equations, solutions and dispersion relations are obtained from them. Dispersion curves are, then, computed and compared with the corresponding ones obtained from the solutions of the three-dimensional Maxwell's equations for the TE and TM modes of various orders and for different values of the refractive index $\hat{n} = (\frac{\epsilon}{\epsilon_0})^{\frac{1}{2}}$. It is shown that the agreement between the approximate and exact dispersion curves is very close.

II 3-D Equations and Boundry Conditions

CH2965-2/91/0000-156 $1.00 © 1991 IEEE

We consider a linear and isotropic dielectric plate of infinite extent referred to a rectangular coordinate system x_i with the faces at $x_2 = \pm b$ and with x_1 and x_3 the coordinates of the middle plane of the plate (see Fig. 1). The plate is surrounded by free space.

The governing equations of the EM field in the plate comprise Maxwell's equations

$$\epsilon_{ijk}\, E_{k,j} = -B_{i,t}\,, \quad B_{i,i} = 0\,, \tag{1}$$

$$\epsilon_{ijk}\, H_{k,j} = D_{i,t}\,, \quad D_{i,i} = 0\,, \tag{2}$$

and the constitutive equations

$$D_i = \epsilon\, E_i\,, \qquad B_i = \mu_o\, H_i. \tag{3}$$

In these equations, E_i and H_i are the electric and magnetic field intensities, D_i and B_i are the electric and magnetic flux densities, ϵ is the dielectric permittivity, μ_0 is the magnetic permeability of free space, and ϵ_{ijk} is the unit alternating tensor. We have assumed that there are no free charges and no current and that the dielectric is nonmagnetic.

Since the region of the plate is source free and the field is time-varying, a vector potential function A_i may be introduced from which E_i and B_i can be obtained by[8]

$$E_i = -A_{i,t}\,, \qquad B_i = \epsilon_{ijk}\, A_{k,j}. \tag{4}$$

Substitution of (4) into (1) shows that (1) are identically satisfied. Therefore, in terms of A_i, (2), (3), and (4) become the governing equations of the EM field in the dielectric plate.

We designate $\overline{E}_i, \overline{D}_i, \overline{H}_i$ and $\overline{B}_i$ as the field vectors in the free space. The governing equations of EM field in free space are the same as (2)-(4) except ϵ is replaced by ϵ_o, the permittivity of free space.

The continuity conditions to be satisfied at $x_2 = \pm b$, the interfaces between the plate and free space, are

$$\begin{aligned} E_a\,(x_2 = \pm b) &= \overline{E}_a\,(x_2 = \pm b), \\ B_2\,(x_2 = \pm b) &= \overline{B}_2\,(x_2 = \pm b), \\ H_a\,(x_2 = \pm b) &= \overline{H}_a\,(x_2 = \pm b), \\ D_2\,(x_2 = \pm b) &= \overline{D}_2\,(x_2 = \pm b) \end{aligned} \tag{5, 6}$$

where $a = 1, 3$.

By substituting (4) into (5), it may be seen that conditions (5) may be replaced by the continutity conditions on the tangential components of A_i, i.e.

$$A_a(x_2 = \pm b) = \overline{A}_a(x_2 = \pm b)\,, \; a = 1, 3. \tag{7}$$

III. Two-Dimensional Equations

In the region of the plate, i.e. for $|x_2| \leq b$, components of the vector potential A_i are expanded in an infinite series with their thickness-dependence expressed by the trigonometrical functions:

$$\begin{aligned} A_a &= \sum_{n=0}^{\infty} A_a^{(n)}(x_1, x_3, t) \cos\tfrac{n\pi}{2}(1-\psi), \quad a = 1, 3 \\ A_2 &= \sum_{n=0}^{\infty} A_2^{(n)}(x_1, x_3, t) \sin\tfrac{n+1}{2}\pi(1-\psi), \end{aligned} \tag{8}$$

where $\psi = \frac{x_2}{b}$, $A_i^{(n)}$ are called the two-dimensional nth-order vector potential, for they are functions of x_1, x_3, and t only.

By substituting (8) into (4), we have

$$\begin{aligned} E_a &= \sum_{n=o}^{\infty} E_a^{(n)} \cos\tfrac{n\pi}{2}(1-\psi), \\ E_2 &= \sum_{n=o}^{\infty} E_2^{(n)} \sin\tfrac{n+1}{2}\pi(1-\psi), \\ B_a &= \sum_{n=o}^{\infty} B_a^{(n)} \sin\tfrac{n+1}{2}\pi(1-\psi), \\ B_2 &= \sum_{n=o}^{\infty} B_2^{(n)} \cos\tfrac{n\pi}{2}(1-\psi), \end{aligned} \tag{9}$$

where

$$\begin{aligned} E_i^{(n)} &= -A_{i,t}^{(n)}, \\ B_1^{(n)} &= \tfrac{n+1}{2b}\pi A_3^{(n+1)} - A_{2,3}^{(n)}, \\ B_2^{(n)} &= k(A_{1,3}^{(n)} - A_{3,1}^{(n)}), \\ B_3^{(n)} &= -\tfrac{n+1}{2b}\pi A_1^{(n+1)} + A_{2,1}^{(n)}. \end{aligned} \tag{10}$$

Eqs. (10) are the two-dimensional nth-order field-potential relations. We note that a correction factor k has been introduced in the expression for

$B_2^{(n)}$, $(10)_3$, to permit some minor adjustment of frequency branches as it will be explained in Section IV.

By multiplying $(2)_1$ (for $i = 1, 3$) and $(2)_2$ by $cos\frac{n\pi}{2}(1-\psi)$ and $(2)_1$ (for $i = 2$) by $sin\frac{n+1}{2}\pi(1-\psi)$, integrating the resulting expressions with respect to ψ from -1 to $+1$, and using the identities:

$$\int_{-1}^{1} sin\tfrac{m\pi}{2}(1-\psi)sin\tfrac{n\pi}{2}(1-\psi)d\psi = \delta_{mn} - \delta_{mo}\delta_{no},$$

$$\int_{-1}^{1} cos\tfrac{m\pi}{2}(1-\psi)cos\tfrac{n\pi}{2}(1-\psi)d\psi = \delta_{mn} + \delta_{mo}\delta_{no}, \qquad (11)$$

we have the two-dimensional field equations

$$-H_{2,3}^{(n)} - \tfrac{n\pi}{2b}H_3^{(n-1)} + \tfrac{1}{b}\mathcal{H}_3^{(n)} = D_{1,t}^{(n)},$$

$$H_{1,3}^{(n)} - H_{3,1}^{(n)} = D_{2,t}^{(n)},$$

$$H_{2,1}^{(n)} + \tfrac{n\pi}{2b}H_1^{(n-1)} - \tfrac{1}{b}\mathcal{H}_1^{(n)} = D_{3,t}^{(n)},$$

$$D_{1,1}^{(n)} + D_{3,3}^{(n)} - \tfrac{n\pi}{2b}D_2^{(n-1)} + \tfrac{1}{b}\mathcal{D}_2^{(n)} = 0, \qquad (12)$$

where δ_{mn} is the Kronecker delta, the components of the nth-order magnetic field intensities and electric flux densities are defined by

$$H_a^{(n)} = \int_{-1}^{1} H_a sin\tfrac{n+1}{2}\pi(1-\psi)d\psi,$$

$$H_2^{(n)} = \int_{-1}^{1} H_2 cos\tfrac{n\pi}{2}(1-\psi)d\psi,$$

$$D_a^{(n)} = \int_{-1}^{1} D_a cos\tfrac{n\pi}{2}(1-\psi)d\psi,$$

$$D_2^{(n)} = \int_{-1}^{1} D_2 sin\tfrac{n+1}{2}\pi(1-\psi)d\psi, \qquad (13)$$

and the components of the nth-order face magnetic field intensities and face electric charge density are defined by

$$\mathcal{H}_a^{(n)} = H_a(b) - (-1)^n H_a(-b),$$

$$\mathcal{D}_2^{(n)} = D_2(b) - (-1)^n D_2(-b). \qquad (14)$$

By inserting (9) in (3) and, in turn, into (13), we obtain the two-dimensional constitutive relations of nth-order in terms of the vector potential $A_i^{(n)}$:

$$H_1^{(n)} = \tfrac{1}{\mu_o}(\tfrac{n+1}{2b}\pi A_3^{(n+1)} - A_{2,3}^{(n)}),$$

$$H_2^{(n)} = \tfrac{1}{\mu_o}k(1+\delta_{no})(A_{1,3}^{(n)} - A_{3,1}^{(n)}),$$

$$H_3^{(n)} = \tfrac{1}{\mu_o}(-\tfrac{n+1}{2b}\pi A_1^{(n+1)} + A_{2,1}^{(n)}),$$

$$D_a^{(n)} = -\epsilon(1+\delta_{no})A_{2,t}^{(n)},$$

$$D_2^{(n)} = -\epsilon A_{2,t}^{(n)}. \qquad (15)$$

We see that for the plate ($|x_2| \le b$) the three-dimensional governing equations (2), (3), and (4) are now replaced by an infinite set of two-dimensional equations (12), (15), and (10), respectively.

Further substitution of (15) in (12) leads to the two-dimensional nth-order wave equations for $A_i^{(n)}$:

$$(1+\delta_{no})k(A_{1,aa}^{(n)} - A_{a,a1}^{(n)}) + \tfrac{n\pi}{2b}A_{2,1}^{(n-1)}$$

$$-(\tfrac{n\pi}{2b})^2 A_1^{(n)} - \tfrac{\mu_o}{b}\mathcal{H}_3^{(n)} = \epsilon\mu_o(1+\delta_{no})A_{1,tt}^{(n)},$$

$$A_{2,aa}^{(n)} - \tfrac{n+1}{2b}\pi A_{a,a}^{(n+1)} = \epsilon\mu_o A_{2,tt}^{(n)},$$

$$(1+\delta_{no})k(A_{3,aa}^{(n)} - A_{a,a3}^{(n)}) + \tfrac{n\pi}{2b}A_{2,3}^{(n-1)}$$

$$-(\tfrac{n\pi}{2b})^2 A_3^{(n)} + \tfrac{\mu_o}{b}\mathcal{H}_1^{(n)} = \epsilon\mu_o(1+\delta_{no})A_{3,tt}^{(n)},$$

$$(1+\delta_{no})A_{a,at}^{(n)} - \tfrac{n\pi}{2b}A_{2,t}^{(n-1)} - \tfrac{1}{b\epsilon}\mathcal{D}_2^{(n)} = 0, \qquad (16)$$

for $n = 0, 1, 2, \ldots$. We note in (16) that $\mathcal{H}_a^{(n)}$ and $\mathcal{D}_2^{(n)}$, which are defined in (14), may be regarded as the "forcing functions" specified at the faces of the plate.

To derive the two-dimensional governing equations for the evanescent field in the upper half of free space ($x_2 \ge b$), we assume

$$\overline{A}_i = A_i^+(x_1, x_3, t)e^{-\overline{\eta}b(\psi-1)}, \qquad (17)$$

where $\overline{\eta}(\ge 0)$ determines the rate of decaying of $\overline{A}_i$ in the x_2 direction. We note that $\overline{\eta}$ is not a constant, but depends on the wave number and frequency of the wave represented by (17). For instance, by letting

$$A_i^+ = C_i e^{i(\xi x_1 - \omega t)} \qquad (18)$$

for an evanscent plane wave propogating in the x_1 direction, and substituting (17) and (18) into (4), (3), and (2), we see that Maxwell's equations are satisfied provided

$$\overline{\eta}^2 = \xi^2 - \frac{\omega^2}{c_o^2}, \tag{19}$$

where $c_o = (\epsilon_o \mu_o)^{-\frac{1}{2}}$ is the velocity of the light in free space.

Now, substitution of (17) into (4) gives

$$\overline{E}_i = E_i^+ e^{-\overline{\eta} b(\psi - 1)},$$

$$\overline{B}_i = B_i^+ e^{-\overline{\eta} b(\psi - 1)}, \tag{20}$$

where

$$E_i^+ = -A_{i,t}^+, \qquad B_1^+ = -\overline{\eta} A_3^+ - A_{2,3}^+$$

$$B_2^+ = A_{1,3}^+ - A_{3,1}^+, \quad B_3^+ = \overline{\eta} A_1^+ + A_{2,1}^+. \tag{21}$$

The above are the two-dimensional field-potential relations in the upper half of free space.

By multiplying (2) by $e^{-\overline{\eta} b(\psi - 1)}$ and integrating with respect to ψ from 1 to ∞, we obtain the two-dimensional field equations

$$-H_{2,3}^+ + \overline{\eta} H_3^+ - \tfrac{1}{b} \overline{H}_3(b) = D_{3,t}^+,$$

$$H_{1,3}^+ - H_{3,1}^+ = D_{2,t}^+,$$

$$H_{2,1}^+ - \overline{\eta} H_1^+ + \tfrac{1}{b} \overline{H}_1(b) = D_{3,t}^+,$$

$$D_{a,a}^+ + \overline{\eta} D_2^+ - \tfrac{1}{b} \overline{D}_2(b) = 0, \tag{22}$$

where the components of the two-dimensional magnetic field intensity H_i^+ and electric flux density D_i^+ are defined by

$$H_i^+ = \int_1^\infty \overline{H}_i e^{-\overline{\eta} b(\psi - 1)} d\psi$$

$$D_i^+ = \int_1^\infty \overline{D}_i e^{-\overline{\eta} b(\psi - 1)} d\psi. \tag{23}$$

Two-dimensional constitutive equations in terms of A_i^+ are obtained by inserting (20) and (21) into (3), in which ϵ is replaced by ϵ_o, and the result into (23):

$$D_i^+ = -\frac{\epsilon_o}{2b\overline{\eta}} A_{i,t}^+,$$

$$H_1^+ = \frac{1}{2b\overline{\eta}\mu_o}(-\overline{\eta} A_3^+ - A_{2,3}^+),$$

$$H_2^+ = \frac{1}{2b\overline{\eta}\mu_o}(A_{1,3}^+ - A_{3,1}^+),$$

$$H_3^+ = \frac{1}{2b\overline{\eta}\mu_o}(\overline{\eta} A_1^+ + A_{2,1}^+). \tag{24}$$

Finally, substitution of (24) in (22) gives the two-dimensional wave equations of A_i^+ in the upper half of free space

$$A_{1,aa}^+ - A_{a,a1}^+ - \overline{\eta} A_{2,1}^+ - \overline{\eta}^2 A_1^+ + 2\overline{\eta}\mu_o \overline{H}_3(b) = \epsilon_o \mu_o A_{1,tt}^+,$$

$$A_{2,aa}^+ + \overline{\eta} A_{a,a}^+ = \epsilon_o \mu_o A_{2,tt}^+,$$

$$A_{3,aa}^+ - A_{a,a3}^+ - \overline{\eta} A_{2,3}^+ - \overline{\eta}^2 A_3^+ - 2\overline{\eta}\mu_o \overline{H}_1(b) = \epsilon_o \mu_o A_{3,tt}^+,$$

$$A_{a,at}^+ + \overline{\eta} A_{2,t}^+ + \frac{2\overline{\eta}}{\epsilon_o} \overline{D}_2(b) = 0. \tag{25}$$

In the same manner, two-dimensional equations for the evanescent field in the lower half of free space ($x_2 \leq -b$) are obtained by assuming

$$\overline{A}_i = A_i^-(x_1, x_3, t) e^{\overline{\eta} b(\psi + 1)}, \tag{26}$$

except that the integration with respect to ψ is from $-\infty$ to -1. These two-dimensional equations are summarized as follows:

$$\overline{E}_i = E_i^- e^{\overline{\eta} b(\psi + 1)},$$

$$\overline{B}_i = B_i^- e^{\overline{\eta} b(\psi + 1)}. \tag{27}$$

Field-potential relations:

$$E_i^- = -A_i^-, \qquad B_1^- = \overline{\eta} A_3^- - A_{2,3}^-$$

$$B_2^- = A_{1,3}^- - A_{3,1}^-, \quad B_3^- = -\overline{\eta} A_3^- + A_{2,1}^-. \tag{28}$$

Field equations:

$$-H_{2,3}^- - \overline{\eta} H_3^- + \tfrac{1}{b} \overline{H}_3(-b) = D_{1,t}^-,$$

$$H_{1,3}^- - H_{3,1}^- = D_{2,t}^-,$$

$$H_{2,1}^- + \overline{\eta}\, H_1^- - \tfrac{1}{b} \overline{H}_1(-b) = D_{3,t}^-,$$

$$D_{a,a}^- - \overline{\eta} D_2^- + \tfrac{1}{b} \overline{D}_2(-b) = 0, \tag{29}$$

where

$$H_i^- = \int_{-\infty}^{-1} \overline{H}_i e^{\overline{\eta} b(\psi+1)} d\psi,$$

$$D_i^- = \int_{-\infty}^{-1} \overline{D}_i e^{\overline{\eta} b(\psi+1)} d\psi. \tag{30}$$

Constitutive equations:

$$D_i^- = -\frac{\epsilon_o}{2b\overline{\eta}} A_{i,t}^-,$$

$$H_1^- = \frac{1}{2b\overline{\eta}\mu_o}(\overline{\eta} A_3^- - A_{2,3}^-),$$

$$H_2^- = \frac{1}{2b\overline{\eta}\mu_o}(A_{1,3}^- - A_{3,1}^-),$$

$$H_3^- = \frac{1}{2b\overline{\eta}\mu_o}(-\overline{\eta} A_1^- + A_{2,1}^-). \tag{31}$$

Wave equations of A_i^-:

$$A_{1,aa}^- - A_{a,a1}^- + \overline{\eta} A_{2,1}^- - \overline{\eta}^2 A_1^- - 2\overline{\eta}\mu_o \overline{H}_3(-b) = \epsilon_o \mu_o A_{1,tt}^-,$$

$$A_{2,aa}^- - \overline{\eta} A_{a,a}^- = \epsilon_o \mu_o A_{2,tt}^-,$$

$$A_{3,aa}^- - A_{a,a3}^- + \overline{\eta} A_{2,3}^- - \overline{\eta}^2 A_3^- + 2\overline{\eta}\mu_o \overline{H}_1(-b) = \epsilon_o \mu_o A_{3,tt}^-,$$

$$A_{a,at}^- - \overline{\eta} A_{2,t}^- - \frac{2\overline{\eta}}{\epsilon_o}\overline{D}_2(-b) = 0. \tag{32}$$

Substitutions of (8), (17), and (26) into (7), the continuity conditions of the tangential components of vector potential, gives

$$A_a^+ = \sum_{m=o}^{\infty} A_a^{(m)}, \quad A_a^- = \sum_{m=o}^{\infty} (-1)^m A_a^{(m)}, \tag{33}$$

and the insertion of (16), the continuity conditions on H_a and D_2, into (14) yields

$$\mathcal{H}_a^{(n)} = \overline{H}_a(b) - (-1)^n \overline{H}_a(-b),$$

$$\mathcal{D}_2^n = \overline{D}_2(b) - (-1)^n \overline{D}_2(-b). \tag{34}$$

By substituting $(16)_4$, $(25)_4$, and $(32)_4$ into $(34)_2$, integrating the resulting relation with respect to t, and setting the integration constant to zero for time-varying field, we obtain

$$\frac{1}{2b}[A_2^+ - (-1)^n A_2^-] = \frac{n\pi}{2b}\hat{n}^2 A_2^{(n-1)}$$

$$-(1+\delta_{no})\hat{n}^2 A_{a,a}^{(n)} - \frac{1}{\overline{\eta} b}\sum_{m=o}^{\infty} \hat{\delta}_{mn} A_{a,a}^{(m)}, \tag{35}$$

where $\hat{n} = (\frac{\epsilon}{\epsilon_o})^{\frac{1}{2}}$ is the refractive index with respect to the vacuum, and

$$\hat{\delta}_{mn} = \tfrac{1}{2}[1 + (-1)^{m+n}] = \begin{cases} 1, & m+n = \text{even} \\ 0, & m+n = \text{odd.} \end{cases} \tag{36}$$

By further substituting the rest of (16), (25), and (32) to $(34)_1$, using the relations (33) and (35), and listing the resulting equations together with $(16)_2$, we have

$$(1+\delta_{no})k(A_{3,13}^{(n)} - A_{1,33}^{(n)}) + (\tfrac{n\pi}{2b})^2 A_1^{(n)} - \hat{n}^2 A_{a,a1}^{(n)} - \tfrac{n\pi}{2b}(1 + \delta_{no} - \hat{n}^2) A_{2,1}^{(n-1)} + \tfrac{1}{\overline{\eta} b}\sum_{m=o}^{\infty} \hat{\delta}_{mn}(-A_{1,aa}^{(m)} + \overline{\eta}^2 A_1^{(m)}) + \tfrac{\hat{n}^2}{c_o^2}(1+\delta_{no}) A_{1,tt}^{(n)} + \tfrac{1}{\overline{\eta} b c_o^2}\sum_{m=o}^{\infty} \hat{\delta}_{mn} A_{1,tt}^{(m)} = 0,$$

$$\tfrac{n+1}{2b}\pi A_{a,a}^{(n+1)} - A_{2,aa}^{(n)} + \tfrac{\hat{n}^2}{c_o^2} A_{2,tt}^{(n)} = 0,$$

$$(1+\delta_{no})k(A_{3,13}^{(n)} - A_{3,11}^{(n)}) + (\tfrac{n\pi}{2b})^2 A_3^{(n)} - \hat{n}^2 A_{a,a3}^{(n)} - (\tfrac{n\pi}{2b})(1 + \delta_{no} - \hat{n}^2) A_{2,3}^{(n-1)} + \tfrac{1}{\overline{\eta} b}\sum_{m=o}^{\infty} \hat{\delta}_{mn}(-A_{3,aa}^{(m)} + \overline{\eta}^2 A_3^{(m)}) + \tfrac{\hat{n}^2}{c_o^2}(1+\delta_{no}) A_{3,tt}^{(n)} + \tfrac{1}{\overline{\eta} b c_o^2}\sum_{m=o}^{\infty} \hat{\delta}_{mn} A_{3,tt}^{(n)} = 0, \tag{37}$$

for $n = 0, 1, 2, \ldots$ These are the two-dimensional nth-order equations of $A_i^{(n)}$ for trapped or guided waves in the isotropic dielectric plate surrounded by free space. In (37), the interaction of the field in the plate with the evanescent field in free space has been taken into account by satisfying the continuity conditions at the faces of the plate.

IV Dispersion Relations

In the case of straight-crested waves propagating in the x_1 direction, i.e. $A_i^{(n)} = A_i^{(n)}(x_1, t)$,

(37) and (10) reduce and separate into two independent sets of equations:

$$(1+\delta_{no})kA^{(n)}_{3,11} - (\tfrac{n\pi}{2b})^2 A^{(n)}_3 + \tfrac{1}{\bar{\eta}b}\sum_{m=o}^{\infty}\hat{\delta}_{mn}(A^{(n)}_{3,11} - \bar{\eta}^2 A^{(n)}_3) = \tfrac{\hat{n}^2}{c_o^2}(1+\delta_{no})A^{(n)}_{3,tt} + \tfrac{1}{\bar{\eta}bc_o^2}\sum_{m=o}^{\infty}\hat{\delta}_{mn}A^{(n)}_{3,tt},$$

and

$$B^{(n)}_1 = \tfrac{n+1}{2b}\pi A^{(n+1)}_3,$$

$$B^{(n)}_2 = -kA^{n}_{3,1},$$

$$E^{(n)}_3 = -A^{n}_{3,t}; \tag{38}$$

$$\hat{n}^2 A^{(n)}_{1,11} + \tfrac{n\pi}{2b}(1+\delta_{no}-\hat{n}^2)A^{(n-1)}_{2,1} + \tfrac{1}{\bar{\eta}b}\sum_{m=o}^{\infty}\hat{\delta}_{mn}(A^{(m)}_{1,11} - \bar{\eta}^2 A^{(m)}_1) = \tfrac{\hat{n}^2}{c_o^2}(1+\delta_{no})A^{(n)}_{1,tt} + \tfrac{1}{\bar{\eta}bc_o^2}\sum_{m=o}^{\infty}\hat{\delta}_{mn}A^{(n)}_{1,tt},$$

$$\tfrac{n+1}{2b}\pi A^{(n+1)}_{a,a} - A^{(n)}_{2,aa} + \tfrac{\hat{n}^2}{c_o^2}A^{(n)}_{2,tt} = 0,$$

and

$$E^{(n)}_1 = A^{(n)}_{1,t}, \quad E^{(n)}_2 = -A^{(n)}_{2,t},$$

$$B^{(n)}_3 = -\tfrac{n+1}{2b}\pi A^{(n+1)}_1 + A^{(n)}_{2,1}. \tag{39}$$

In the above, we see that $A^{(n)}_3$ is not coupled to $A^{(n)}_1$ and $A^{(n)}_2$ and is related to field components $B^{(n)}_1, B^{(n)}_2$, and $E^{(n)}_3$ only. Therefore, (38) are the equations for the transverse electric (TE) plane waves. In (39), $A^{(n+1)}_1$ and $A^{(n)}_2$ are coupled together and they are related to field components $E^{(n)}_1$, $E^{(n)}_2$, and $B^{(n)}_3$. Hence, (39) are the equations for the transverse magnetic (TM) plane waves. Also, we note that within each set of these equations $A^{(n)}_i$ and $A^{(m)}_i$ are coupled through $\hat{\delta}_{mn}$, defined by (36), when both m and n are even and when both m and n are odd. These couplings among the modes of different orders are usually very weak. In the remaining portion of this paper, these couplings are omitted by letting

$$\hat{\delta}_{mn} = \delta_{mn}. \tag{41}$$

TE waves

We let

$$A^{(n)}_3 = C^{(n)}_i e^{i(\xi x_1 - \omega t)}, \tag{42}$$

where $C^{(n)}_i$ are the constant amplitudes of the straight-crested waves.

Insertion of (41) and (42) in $(38)_1$ gives the dispersion relations of the nth-order TE waves

$$\alpha X^2 - \beta\Omega^2 + \gamma = 0, \tag{43}$$

where

$$\alpha = (1+\delta_{no})k + \tfrac{2}{\pi\overline{Y}},$$

$$\beta = (1+\delta_{no})\hat{n}^2 + \tfrac{2}{\pi\overline{Y}},$$

$$\gamma = n^2 + 2\tfrac{\overline{Y}}{\pi}, \tag{44}$$

and the dimensionless wave numbers (X and $\overline{Y}$) and frequency (Ω) are defined by

$$X = \frac{\xi}{\frac{\pi}{2b}}, \ \overline{Y} = \frac{\bar{\eta}}{\frac{\pi}{2b}}, \ \Omega = \frac{\omega}{\frac{\pi c_o}{2b}}. \tag{45}$$

By rewriting (19) in the dimensionless form

$$\overline{Y}^2 = X^2 - \Omega^2, \tag{46}$$

we see that α, β, and γ in (44) are dependent not only on the order n, the refractive index $\hat{n}$, and correction factor k, but also on X and Ω.

The frequency Ω, as the roots of (43), may be expressed by the functional relation

$$\Omega = \Omega(X; n, \hat{n}, k). \tag{47}$$

The Ω vs. X curves of (47) are the dispersion curves. They are of the symmetric modes for $n = 0, 2, 4...$ and of the anti-symmetric modes for $n = 1, 3, 5, ...$ We shall regard theses curves as approximate ones since they are the roots of the algebraic

equations (43) which are the approximations of the transcendental equations[1,2] derived from the solutions of the three-dimensional Maxwell's equations.

For the need of comparison, the exact dispersion curves are computed from the transcendental equations[1] of TE waves for the first four branches ($n = 0, 1, 2$, and 3) and are shown in Figs. 2, 3, and 4 for $\hat{n} = 1.5$, 5, and 15, respectively.

We note that the factor k is introduced firstly in the expression for $B_2^{(n)}$ in $(10)_3$ and it appears consequently in the expression for α in $(44)_1$. Its introduction is to allow a slight correction on the slope of the approximate dispersion curves at large values of X. By matching the value Ω calculated from (43) for $n = 0$, $X = 4.0$ with the exact value calculated from the transcendental equation, we obtain

$$k = 0.925 \qquad \text{for } \hat{n} = 1.5$$

$$k = 0.875 \qquad \text{for } \hat{n} = 5 - 15 \tag{48}$$

With the use of these values of k, dispersion curves are calculated from (43) and plotted in Figs. 5, 6, and 7 for $\hat{n} = 1.5$, 5, and 15, respectively.

By comparing the "exact" curves in Figs. 2, 3, and 4, with the "approximate" curves in Figs. 5, 6, and 7, it may be seen that the agreement is very close for various orders of modes and for various values of refractive index.

TM waves

By letting

$$A_1^{(n+1)} = C_1^{(n+1)} e^{i(\xi x_1 - \omega t)},$$

$$A_2^{(n)} = C_2^{(n)} e^{i(\xi x_1 - \omega t)}, \tag{49}$$

and substituting (41) and (49) in $(39)_{1,2}$, we obtain the dispersion relations of the nth-order TM waves

$$(\beta' X^2 - \beta' \Omega^2 + \gamma')\,(X^2 - \hat{n}^2 \Omega^2) + (n+1)^2\,(\hat{n}^2 - 1) X^2 = 0, \tag{50}$$

where n = 0, 1, 2,...and

$$\beta' = \hat{n}^2 + \tfrac{2}{\pi \overline{Y}},$$

$$\gamma' = (n+1)^2 + \tfrac{2\overline{Y}}{\pi}. \tag{51}$$

In a similar manner, the exact dispersion curves are computed from the transcendental equations for the TM waves[1,2] and presented in Figs. 8, 9, and 10 for $\hat{n} = 1.5$, 5, and 15, respectively. The corresponding approximate dispersion curves are calculated from (50) and given in Figs. 11, 12, and 13. We note that no correction is made for these curves since factor k does not appear in (50).

Comparison of the exact curves in Figs. 8-10 with the corresponding approximate ones in Figs. 11-13 shows that the agreement is, again, very close.

In summary, a set of two-dimensional equations is derived for guided EM waves in a dielectric plate surrounded by free space. In these equations, the interaction of the vibrating field in the plate with the evanescent field in free space is taken into account by satisfying the continuity conditions at the faces of the plate.

Dispersion relations for straight-crested waves are obtained from the two-dimesnional equations as a series of algerbraic equations instead of the transcentdental equations from the three-dimensional Maxwell's equations.

It is shown that the agreement is very close between the approximate and exact dispersion curves for various orders of waves and for different values of refractive index.

Aknowledgements

This work was supported by a grant (Contract No. DAAL 03-90-G-0079) from the U.S. Army Research Office.

References

1. D. MARCUSE, *Light Transmission Optics*, (Krieger, Florida, 1989) p. 305.
2. R.E. COLLIN, *Field Theory of Guided Waves*, (McGraw-Hill, New York, 1960) p. 470.

3. S.D. POISSON, "Mémoire sur l'équilibre et le mouvement des corps élastiques," Mém. Acad. Sci., Paris, Ser. 2, **8** pp. 357-570, 1829.
4. A.L. CAUCHY, "Sur l'équilibre et le mouvement d'une plaque élastique dont l'élasticité n'est pas la même dans tous les sens," Excercices de Mathématique **4**, pp. 1-14, 1928.
5. R.D. MINDLIN, "Influence of Rotary Inertia and Shear on Flexural Motions of Isotropic, Elastic Plates," J. Appl. Mech., **18**, pp. 31-38, 1951.
6. R.D. MINDLIN, "High Frequency Vibrations of Crystal Plates", Quart. Appl. Math., XIX, pp. 51-61, 1961.
7. P.C.Y. LEE, S. SYNGELLAKIS, AND J.P. HOU, "A Two-Dimensional Theory for High-Frequency Vibrations of Piezoelectric Crystal Plates with or without Electrodes," J. Appl. Phys., **61** (4), pp. 1249-1262, 1987.
8. M. BORN AND E. WOLF, *Principles of Optics*, 3rd. ed. (Pergarmon Press, New York, 1965) pp. 71-73.

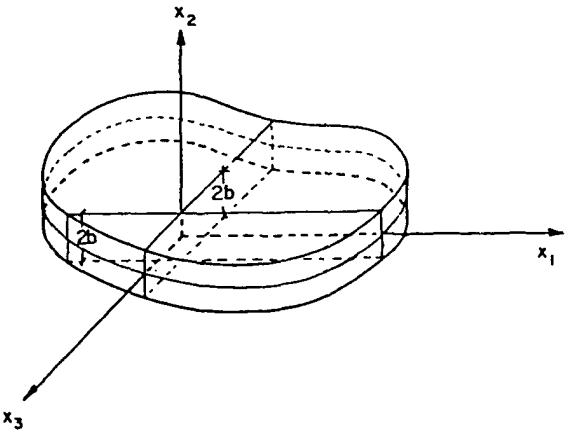


Fig. 1 A dielectric plate surrounded by free space.

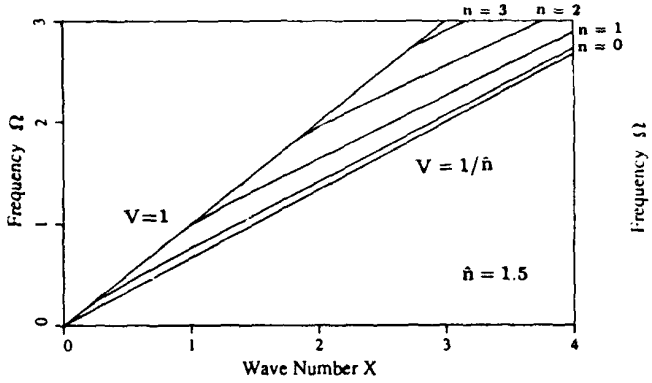


Fig. 2 Dispersion curves for guided TE waves in a dielectric plate surrounded by free space for refractive index $\hat{n} = 1.5$, according to the 3-D equations.

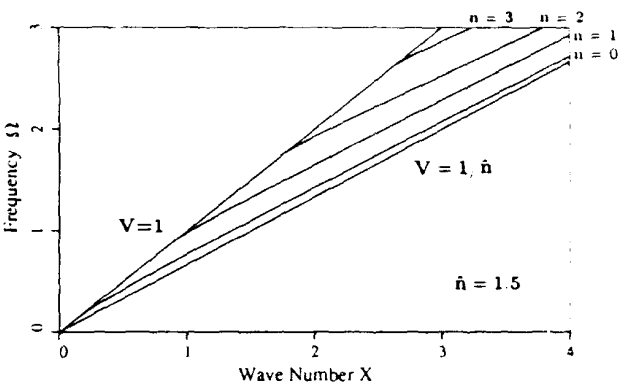


Fig. 5 Ω vs. X curves for TE waves and $\hat{n} = 1.5$, according to the 2-D equations.

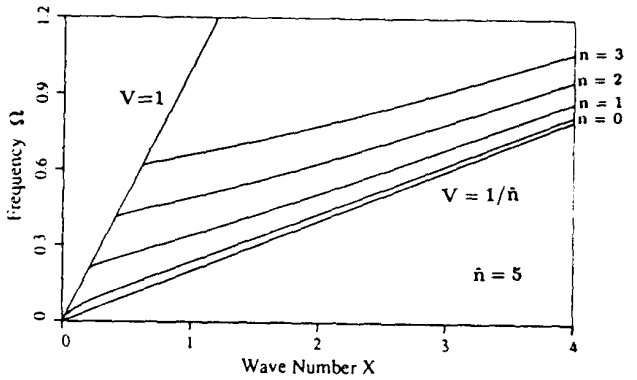


Fig. 3 Ω vs. X curves for TE waves and $\hat{n} = 5$, according to the 3-D equations.

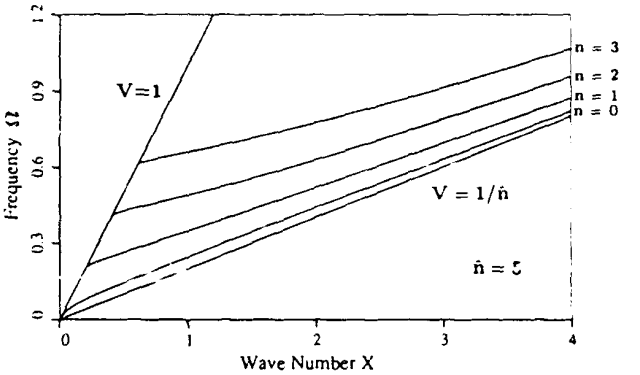


Fig. 6 Ω vs. X curves for TE waves and $\hat{n} = 5$, according to the 2-D equations.

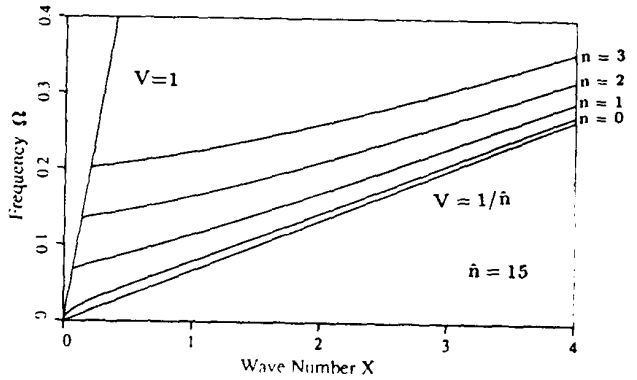


Fig. 4 Ω vs. X curves for TE waves and $\hat{n} = 15$, according to the 3-D equations.

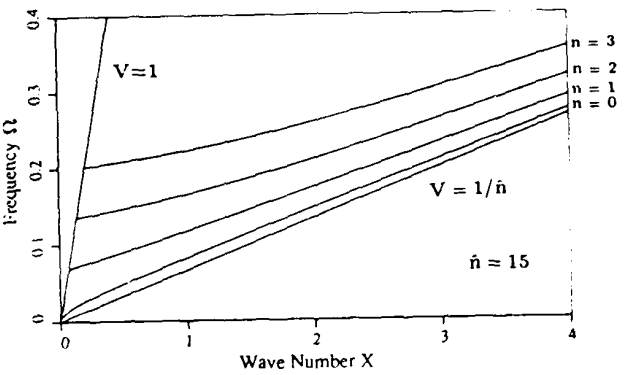


Fig. 7 Ω vs. X curves for TE waves and $\hat{n} = 15$, according to the 2-D equations.

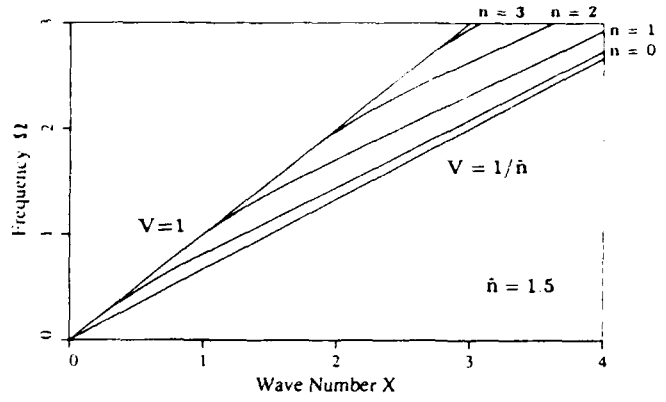


Fig. 8 Ω vs. X curves for TM waves and $\hat{n} = 1.5$, according to the 3-D equations.

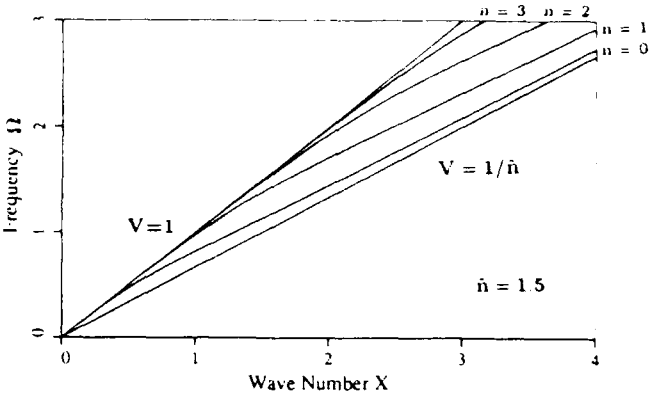


Fig. 11 Ω vs. X curves for TM waves and $\hat{n} = 1.5$, according to the 2-D equations.

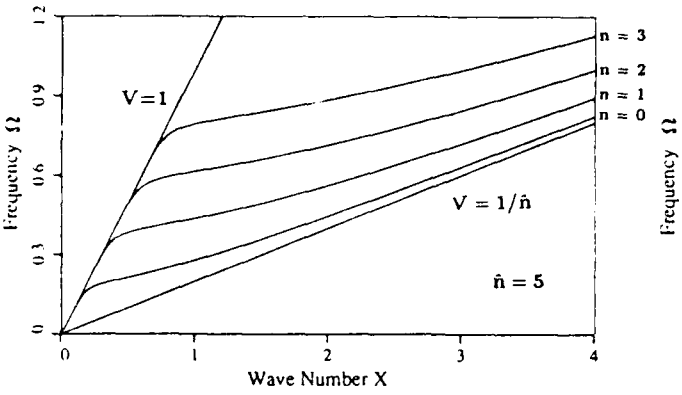


Fig. 9 Ω vs. X curves for TM waves and $\hat{n} = 5$, according to the 3-D equations.

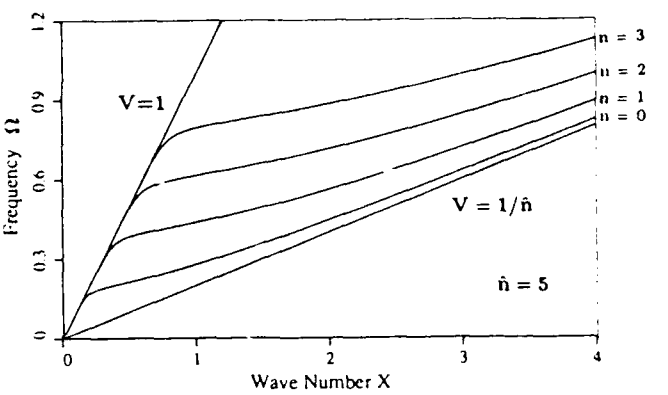


Fig. 12 Ω vs. X curves for TM waves and $\hat{n} = 5$, according to the 2-D equations.

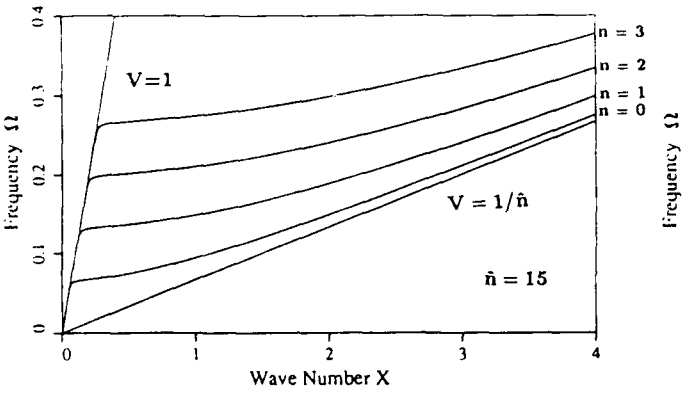


Fig. 10 Ω vs. X curves for TM waves and $\hat{n} = 15$, according to the 3-D equations.

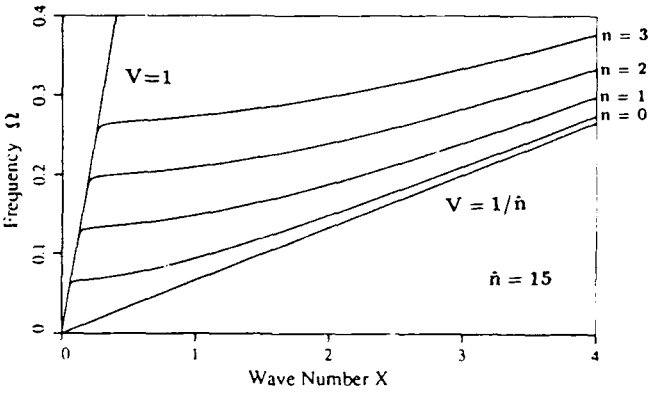


Fig. 13 Ω vs. X curves for TM waves and $\hat{n} = 15$, according to the 2-D equations.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

OPTIMIZATION OF AT BERLINITE AND QUARTZ THICKNESS SHEAR DEVICES FOR VHF FILTER APPLICATIONS

J.Détaint, H.Carru, J.Schwartzel, C.Joly,
and
B.Capelle*, A.Zarka*, Y.Zheng*, E.Philippot**.

C.N.E.T. Centre PAB dept BAG/MCT/CMM 92220 Bagneux France.
*L.M.C. Université Pierre et Marie Curie 75005 Paris France.
**L.P.C.P. Université du Languedoc 34000 Montpellier France.

Abstract.

Systematic investigations concerning the design of V.H.F. berlinite and quartz resonators were made in order to find the best designs for the elements of filters for the intermediate frequency of the future pan-european numerical radiotelephone of the G.S.M. system.

Depending on how the selectivity is distributed along the receiving chain, the I.F. filtering function can be realized, in G.S.M. sets ,when no or little complementary filtering is used, by either 2 cascaded low losses surface wave filters or a 4 poles linearized phase bulk wave filter. When a large amount of complementary filtering is used, either in base band or in a second I.F., the function can be realized by one surface wave filter or by one 2 pole bulk wave filter. For bulk wave filters, the need of a linearized phase response and the high centre frequency (45 to about 200MHz) prescribe to build them with elements having a very reduced sensitivity to the dispersions induced by the elaboration process.

To minimize the sensibility of the response (thermal behaviour) to dispersions of the plate orientation, the thermal variations of the resonance frequency for plates having orientations in the vicinity of the AT cut were studied.

A systematic investigation of the influence of the parameters (electrode shape and surface, mass loading, etc...) of the plane fundamental resonators using the AT cut of berlinite and quartz on the elements of the equivalent scheme and on the anharmonic spectrum was made. This was done using a numerical model based upon the theory of the essentially thickness modes [2]. Several family of designs were found to lead to nearly satisfactory sensitivities and values of the equivalent scheme, but when a constraint was put on the impedance level (inductance) and on the sensitivities to the variations of several parameters that are difficult or costly to totally control in an industrial process, one family of design, using elliptical electrodes having an axis ratio respecting the lateral anisotropy , was found to be much better.

Introduction

Berlinite (AlPO4) is a crystallographic analogous of alpha quartz which has a similarly favourable third order temperature behaviour for the AT cut together with a larger coupling coefficient. This material permits to obtain filters with bandwidths nearly two times larger than with quartz. Berlinite crystals with a high purity and a high Q factor are now available [1]. This material is well adapted to build the I.F. filters of the future pan-european numerical radiotelephone (G.S.M.) system which requires a bandwidth of about 200kHz at a centre frequency in the range of 45 to about 200MHz.

It was previously shown that berlinite has a sufficient coupling coefficient to permit the obtention of bulk wave filters having such bandwidth for centre frequencies above about 45MHz to 55MHz [3] [4]. In the case of quartz it is necessary to use a centre frequency above about 100MHz.

These specifications require to use the fundamental mode for the piezoelectric elements of the filter at a high value of the centre frequencies (most often 70 to 160 MHz). A linearized phase response is also required. In these conditions, the conventional technology would normally lead to have quite large values of the relative dispersions of the parameters of the piezoelectrical elements used in the filters whereas the filter design is made more sensitive by the phase condition. It is thus necessary to improve the technology to reduce, as much as possible, the dispersions.

However, the constraints put on the technology can be somehow relaxed if designs less sensitive to the dispersions can be found for either the filter circuit or its piezoelectrical elements.

In this communication, we consider particularly the research of designs for the resonators components of the filter that minimize the sensitivity of their equivalent scheme towards the dispersions of their physical parameters.

Another important criterion also considered is to obtain resonators with a sufficiently low value of the inductance. It is, indeed, an important advantage at very high frequencies, to keep the iterative impedance of filters as low as possible. This reduces the influence of parasitic capacitances and also permits to lower the complexity and cost of the coils and transformers.

In the following paragraphs, after having considered the specifications of the I.F. filters for G.S.M. sets, we will research how it is possible to reduce the sensitivity of the resonators characteristics to the dispersions introduced by the technological process. In a first step, made in view to choose the crystalline orientation that minimize the thermal drift for berlinite filters, the angular position of AT cut is discussed. Then, a systematic investigation will be made of the influence of the parameters (electrode shape and surface, mass loading, etc....) of the plane fundamental resonators using the AT cut of berlinite and quartz, on the elements of the equivalent scheme and on the anharmonic spectrum. This is done using numerical models based upon the theory of the essentially thickness modes [5]. The particular interest of one family of designs, based upon resonators that take into account the lateral anisotropy of AT plates [6], will be particularly emphasised since this family of designs leads to optimum properties relatively to several criteria.

CH2965-2/91/0000-166 $1.00 © 1991 IEEE

I The I.F. filter for the G.S.M. Radiotelephone.

This system will cover all the west European countries in the next ten years. It uses the frequency bands 890-915 and 935-960MHz and employs digital transmission of voice and data signals in 200 kHz channels shared by 8 subscribers. The possibility of using slow frequency hoping according to a code table is provided for introducing a further improvement of the transmission efficiency (resulting of this kind of frequency diversity). The type of modulation chosen (GMSK) gives a mean frequency spectrum which decreases rapidly from the centre of the channel and is such that most of the energy of the signal is contained in nearly the 2 /3 of the channel width. This system is cellular and is ruled by the GSM specifications (defined by the Groupe Special Mobile of the CEPT) which are mostly of external nature (defining the signals exchanged between the different equipments).

The radio sub-system of a G.S.M. set is typically constituted as indicated in figure 1. The intermediate frequency (I.F.) is not defined by the GSM specifications which only specify the global channel selectivity.

Due to the system characteristics (separation between the transmission and reception sub-bands) and to usual design rules for the R.F. equipments, I.F. situated between 45MHz and approximatively 200 MHz can be chosen with some restrictions resulting of the existence of frequency bands where strong radio signals are present(F.M. and television bands etc..).

The global selectivity requirements for the channel can be divided in different manners between the I.F. filter(s) and a complementary filtering in base band; the choice between the different possible solutions being mostly the result of an arbitration between cost and performances that takes into account the existing solutions for the I.F. filter.

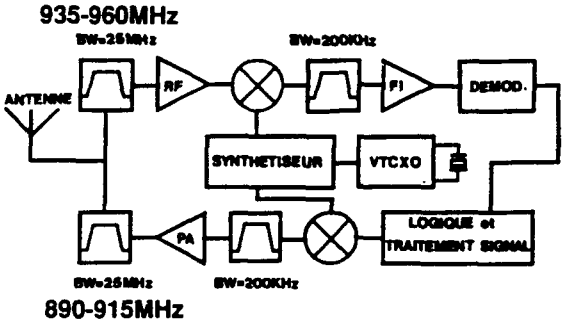


Figure 1: Typical scheme of the radio sub-system of a GSM radiotelephone.

The best technical solution is to place the maximum of selectivity as ahead as possible in the receiver. In this case the total selectivity requirements (figure 2) can be achieved using two low losses surface wave filters or one bulk wave filter with a linearized phase response. The possibility of fulfilling such specifications with a four poles berlinite filter using the fundamental mode was recognized several years ago with filters made at lower frequencies. The response curves given in figure 3 are those of such a filter that was obtained in 1986 [6]. Since that time , other demonstrations were made [3] [4] at higher frequencies (20 and 70 MHz) , and the possibility to obtain, at low cost, the thin berlinite plates necessary to build cheap berlinite filters in the frequency range 45-85 MHz was demonstrated [7]. Bulk wave quartz filters with such bandwidth can be made above about 100MHz. For quartz also, a chemical etching technique exists to produce the very thin plates required [8].

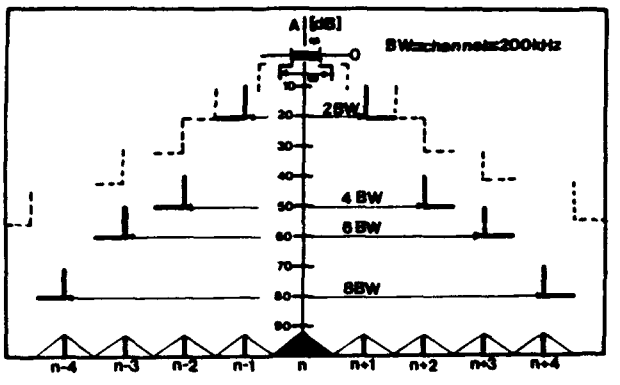


Figure 2: Global specifications of the channel selectivity and relaxed filter specifications (use of complementary filtering).

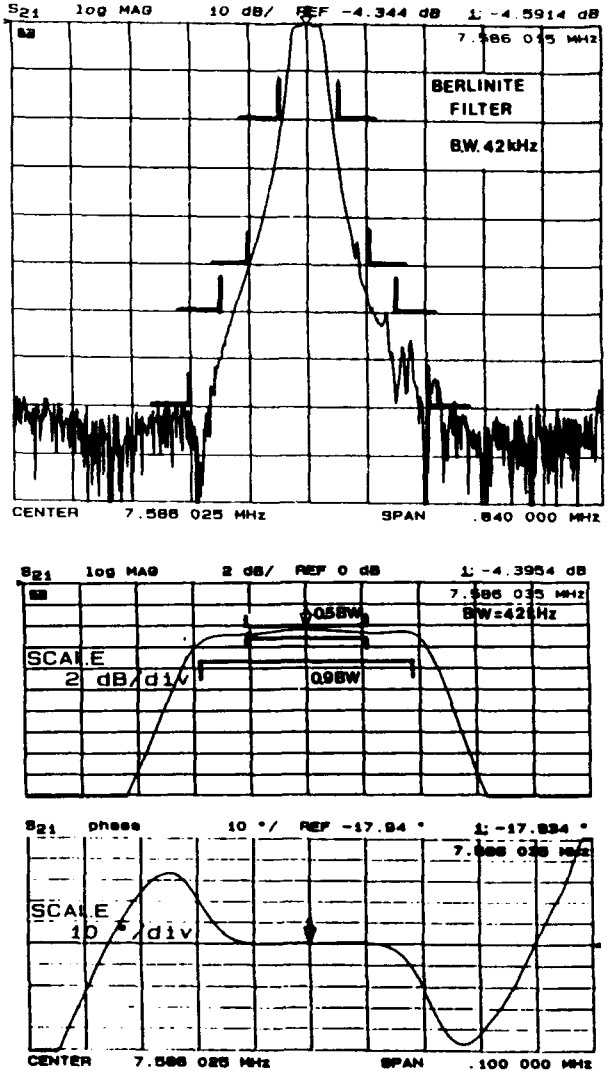


Figure 3: Response of a 4 poles Jaumann berlinite filter with a linearized phase response.

Other choices of the distribution of selectivity in the G.S.M. set appear to be possible, and it seems, in the present state of the development of the receiver circuits, that the minimum selectivity that can be placed in the (first) I.F. filter is some thing like what is also represented in figure 2 (the less stringent specifications). Such specifications can be fulfilled either by one low loss surface wave filter or a two poles bulk wave filter.

Other constraints exist for the components of this system: an important degree of miniaturization and a very reduced cost are required.

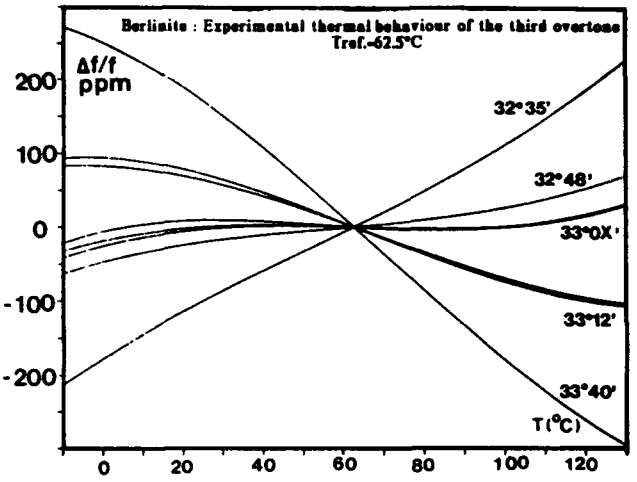


Figure 4: Experimental results of the thermal behaviour of Y rotated berlinite resonators near the AT cut [Third overtone].

II Optimization of the crystalline orientation of berlinite plates.

Using previously obtained results [1] and also the results of new experiments made with resonators operating on the third overtone and having cuts nearly equally spaced in the range Y-32°, Y-34° (examples of such behaviours are given on figure 4), we have computed the variations of the 1st, the 2nd and the 3rd order temperature coefficients of the resonance frequency as a function of the cut angle. The results are represented in the form of relative variations related to the Y-33° cut on the figures 5 to 7 . On these figures we consider the values of the temperature coefficients at the reference temperature of 60°C (this round value is quite close to the inflexion temperatures for the different cuts in the chosen range). As for the AT cut of quartz it was observed that the 1st and the 2nd order temperature coefficients vanish for nearby angles but here for a temperature above the ambient temperature, (or otherwise said, it exists a reference temperature and a cut at which they vanish simultaneously). Also, it can be observed in figure 7, that the third order coefficient presents a maximum in this angle range. Using these temperature coefficients it is possible to simulate the thermal behaviour of 3rd overtone resonators with any orientation in the interval [Y-32°,Y-34°]. This is done on figure 8 for resonators with cuts situated around the AT cut for the third overtone.

Similar measurements were also made for fundamental mode resonators made with the high Q factor crystals now available. As already observed in the early age of synthetic quartz, slight differences of thermal behaviour were observed between these samples and those previously available.

On figure 9 is represented the frequency response of one

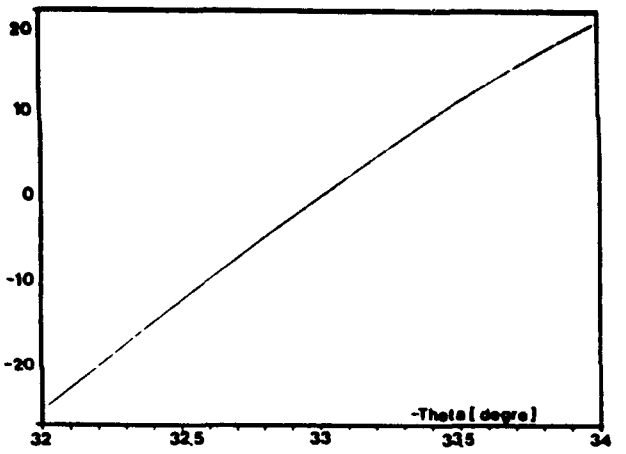


Figure 5: Variation of the first order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone].

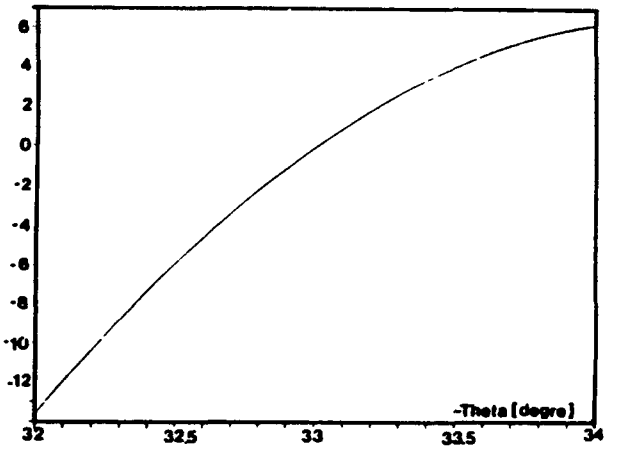


Figure 6: Variation of the second order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone] .

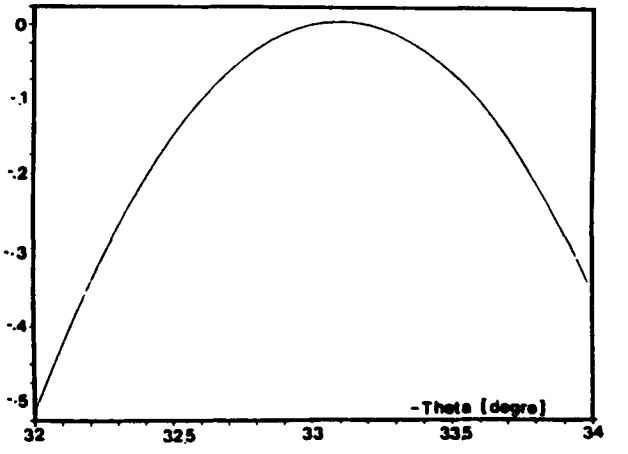


Figure 7: Variation of the third order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone].

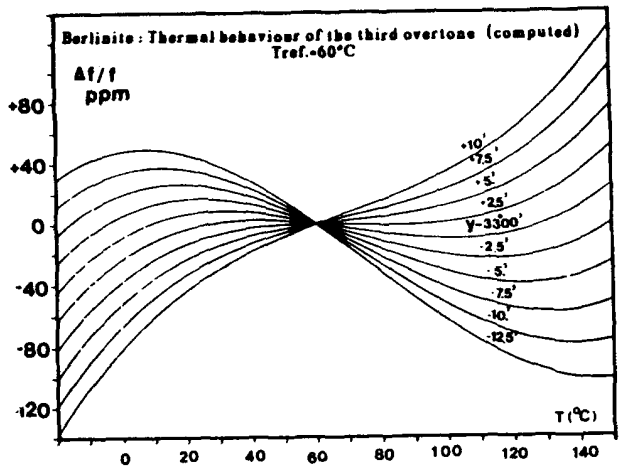


Figure 8: Computed thermal behaviour for the third overtone near AT.

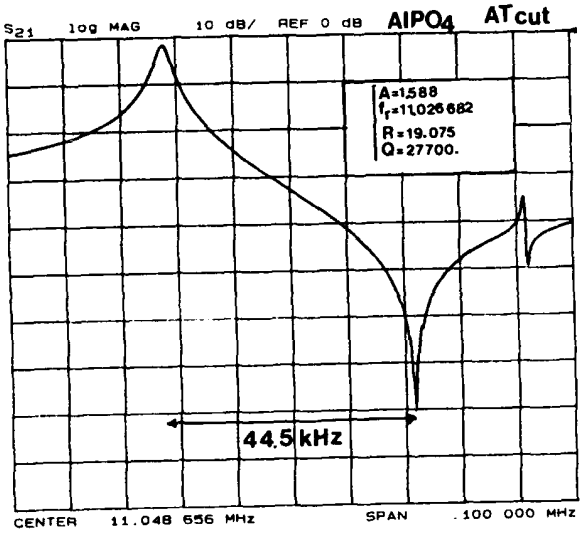


Figure 9: Typical electrical response of a fundamental mode resonator near AT.

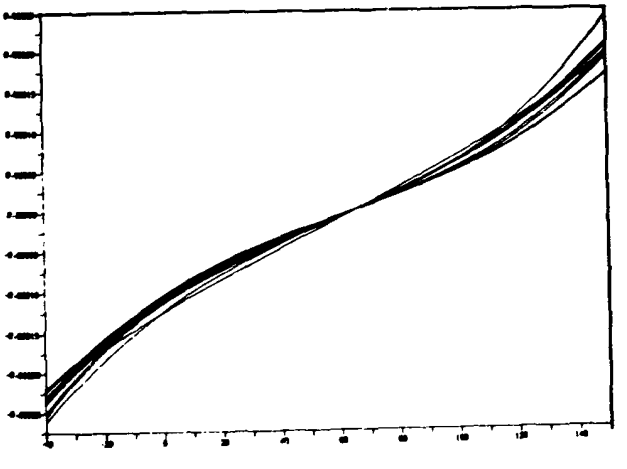


Figure 10: Example of thermal behaviour of fundamental mode resonators.

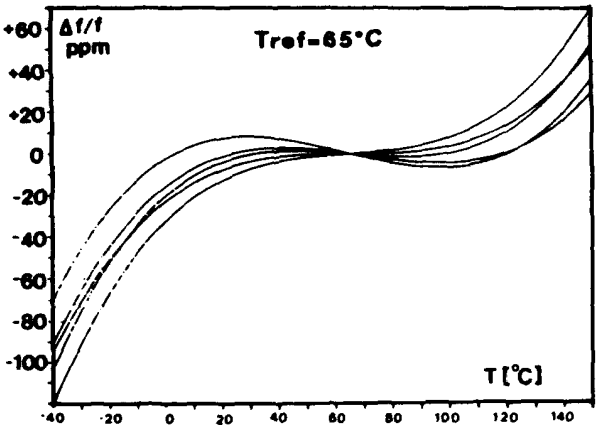


Figure 11: Thermal behaviour near the AT cut for the fundamental mode.

resonator used for these measurements. An example of the obtained results is displayed on figure 10 ; it represents the thermal behaviour of resonators having a cut angle somehow below that giving an horizontal inflexion tangent. These resonators present an angular dispersion resulting of the all the technological operations needed to obtain a V.H.F. plate including a deep chemical thinning. Another example is given on figure 11 where are represented the thermal behaviour of plates operating on the fundamental mode that were cut very near the nominal AT+0° orientation of the fundamental mode and present an angular dispersion that was voluntary increased to obtain more informations around this angle.

III RESONATOR DESIGNS ALLOWING A MINIMISATION OF THE SENSITIVITY OF THE ELECTRICAL RESPONSE TO DISPERSIONS.

III.1 Physical and electrical parameters of a resonator.

The physical parameters of a piezoelectric resonator are listed in table 1 together with the origin of their dispersions.

PHYSICAL	PARAMETERS
Plate orientation	-angular dispersion -temperature coeff.
Thickness (2h)	-parallelism -dispersion of 2h
Electrode Geometry	-dispersions on geometry -misalinement
Mass loading (R)	accumulation of most of the other dispersions
ELECTRICAL	**PARAMETERS**
Resonance Frequency	- very accurate
Inductance	-accumulate dispersions
Static Capacitance	-errors on geometry
Unwanted modes	-most of the parameters

Table 1: Parameters of a thickness shear mode resonator.

The case of the plate orientation and of the corresponding dispersions is particular since it is practically uncoupled of the other; this results of the fact that the temperature coefficients have much faster angular variations than the other properties (dielectric constants, coupling coefficients, effective elastic constants etc...). For the plate orientation the only possible actions are the choice of the best cut and the reduction of the angular dispersion.

The case of the thickness of the plate needs several comments. First, for filter applications a very good parallelism must be obtained unless antisymmetrical anharmonics appear. It was previously shown [6] that for unparallel plates, the main mode is displaced under the electrode in the direction of the least thickness, so that the value of the inductance is greatly modified. Unless to use intricate resonator structures , the only practical choice is to have sufficiently parallel plates (since the possible actions that can reduce the sensitivity of the inductance to the unparallelism will simultaneously increase the importance of antisymmetrical anharmonics). The second comment is to notice that the frequency adjustment operation transforms the dispersion of the thickness into a dispersion of the mass loading. An important dispersion of this parameter constrains to use large values for the mass loading ; we will see below that this prevents to have low inductance and low sensitivity designs. Fortunately berlinite and quartz can be easily thinned chemically so that a precise thickness can be obtained. Another point related to the preceding is the question of the measurement of the plate thickness or more usually of a frequency significative of this parameter with sufficient reproducibility and accuracy . This point is largely discussed in reference [9] where it is indicated that at the very high frequencies large systematic errors can be made if no precautions are taken. The influence of the two other physical parameters (electrode geometry and mass loading) and of their dispersion will be largely discussed in the following paragraphs.

III.2 Sensitivity of a circuit to dispersions of its components.

The electrical responses of electrical circuits are more or less sensitive, depending on their designs, to the dispersion of their constitutive elements. In the case of crystals containing circuits, such as crystal filters, the electrical response is particularly sensitive to the dispersions of the electrical parameters of the crystals which are elements having reactances with fast frequency variations. It is thus important to find designs for these resonators that minimize the influence of the unavoidable dispersions of their physical parameters on the values of the elements of the equivalent scheme.

Definition of the relative sensitivity:

If F(p,X1,X2,...,Xi,..) is a Network Function of the circuit, for example a transfert function (such as $Z_{21}(p)$, $S_{21}(p)$, or H(p), the voltage-ratio transfert function...), and the $X_i(p)$ are electrical variables characteristics of the components used in the circuits, the relative sensitivity of F to variations of X_i is defined by the relation:

$$S_{X_i}(F) = \frac{X_i}{F}\frac{\partial F}{\partial X_i}$$

The variation ΔF of F corresponding to the variationsΔX_i of the components can be obtain by a Taylor development:

$$\Delta F = \frac{1}{1!}\left\{\left(\sum_i \frac{\partial}{\partial X_i}\Delta X_i\right)F\right\} + \frac{1}{2!}\left\{\left(\sum_i \frac{\partial}{\partial X_i}\Delta X_i\right)^2 F\right\} + \frac{1}{3!}\left\{\left(\sum_i \frac{\partial}{\partial X_i}\Delta X_i\right)^3 F\right\} + \ldots\ldots$$

If we consider a 1st order approximation we have:

$$\frac{\Delta F}{F} = \sum_i\left\{\left(\frac{X_i}{F}\frac{\partial F}{\partial X_i}\right)\frac{\Delta X_i}{X_i}\right\} = \sum_i\left\{S_{X_i}(F)\frac{\Delta X_i}{X_i}\right\}$$

If now we express the electrical variables as functions of their physical parameters and of the environmental variables: $X_i = X_i(U_1^i, U_2^i, \ldots\ldots, U_k^i, \ldots\ldots)$, we obtain formally:

$$\frac{\Delta F}{F} = \sum_i\sum_k\left\{\left(\frac{X_i}{F}\frac{\partial F}{\partial X_i}\right)\left(\frac{U_k^i}{X_i}\frac{\partial X_i}{\partial U_k^i}\right)\frac{\Delta U_k^i}{U_k^i}\right\}$$

$$\frac{\Delta F}{F} = \sum_i\sum_k\left\{S_{X_i}(F).S_{U_k^i}(X_i)\frac{\Delta U_k^i}{U_k^i}\right\}$$

This expression contains a sum of products of two kinds of terms. The terms like $S_{X_i}(F)$ are characteristic of the circuit design, whereas the terms like $S_{U_k^i}(X_i)$ are characteristic of the component design. This indicates that, whatever are the "circuit sensitivities" it is sufficient to obtain satisfactory sensitivity properties that the "components sensitivities" can be kept to very low values.

A theoretically better approach would be to obtain a global minimisation of $\frac{\Delta F}{F}$, or of a function of the values taken by this quantity for different frequencies of the pass-band and of the stop-bands of the filter. In such cases, it may be possible to make use of compensation between different terms. This could be made directly using numerical minimisation techniques, but requires that the circuit is no too intricate and allows sufficient designs variations.

In this communication, we would consider only the research of resonators designs that lead to reduced sensitivity of their main electrical variables (inductance. resonance frequency, existence of anharmonics) to dispersions in their physical parameters (thickness, electrode dimensions, mass loading, etc). This approach is justified by the previous remark (sufficient condition) and also by the fact that very little circuit design variations are allowed with linearized phase filters when, simultaneously, very constraining amplitude and phase responses are specified [11].

III.3 Model of energy trapping resonator.

As a basis of this model, we use the scalar equation governing the lateral dependence of the anharmonics in the vicinity of one overtone (number n) of a given thickness mode. This equation was first established for monoclinic plates [12] then for plates of an arbitrary orientation [2]; in this case the equation has a coordinate system with x_2 normal to the thickness and x_1, x_3 chosen to eliminate the mixed derivative [2] the following form :

$$M_n\frac{\partial^2 \bar{u}_1^n}{\partial x_1^2} + P_n\frac{\partial^2 \bar{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 C^*}{4h^2}\bar{u}_1^n + \rho\omega^2\bar{u}_1^n$$

$$= \rho\omega^2(-1)^{(n-1)/2}\frac{e_{26}4V_0 e^{j\omega t}}{c^{(1)}n^2\pi^2} \qquad \{1\}$$

In this equation, $\bar{u}_1(x_1,x_3,t)$ is the lateral dependence of $\hat{u}_1^n(x_1,x_2,x_3,t) = \bar{u}_1^n \sin(n\pi x_2/2h)$.

$\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of ϕ the potential , made to replace the inhomogeneous boundary conditions $\phi = \pm V_0 e^{j\omega t}/2 at x_2 = \pm h$ by the homogeneous conditions $\hat{\phi} = 0$. The inhomogeneous term of the equation results from this transformation.

$$u_1^n = \hat{u}_1^n - \frac{e_{26} V_0 e^{j\omega t}}{c^{(1)} 2h} x_2;$$

$$\phi = \hat{\phi} + \frac{x_2 V_0 e^{j\omega t}}{2h};$$

$$u_1 = \sum_n u_1^n$$

c* is either $\bar{c}^{(1)}$ for the unelectroded part of the resonator or $\hat{c}^{(1)}$ for the electroded part. $\bar{c}^{(1)}$ is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix).

$$\hat{c}^{(1)} = \bar{c}^{(1)}(1 - 8k^{(1)2}/n^2\pi^2 - 2\hat{R})$$

$\hat{c}^{(1)}$ is a constant that includes the electrical effect of the metallization. $4k^{(1)2}/n^2\ \pi^2$ is approximatively the relative frequency lowering due to the electrical effect of the metallization, $\hat{R}$ is the mass loading. $c^{(1)} = \bar{c}^{(1)}(1 - k^{(1)})^2$ is a pseudo ordinary elastic constant which is equal to C_{66} for the Y rotated cuts. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode, this quantity is equal to k_{26} for the Y rotated cuts . M , P are intricate functions of the material constants and of the plate orientation that can be obtained from the method described in reference [5] or in reference [2].

Boundary and continuity conditions : the traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish the equations. On the surfaces limiting the electroded and the unelectroded regions of the resonator we have to specify the continuity of u_1 and of its normal derivative.

In this paper we suppose that the amplitude of the vibration mode near the edges of the plate are negligible so that no boundary condition at the plate edge are taken into account. It was previously shown that, in this case, they induce a negligible modification of both the vibrations modes and the resonances frequencies [13][14].

Eigen mode analysis : the eigen modes of the resonator can be obtained by the simultaneous solution of the equations for the electroded and unelectroded regions taking into account the continuity conditions at the electrode edge. At V=0 the homogeneous form of the approximate equation for the two regions reduces to :

$$M'_n u_{1,11}^n + P'_n u_{1,33}^n + 4\pi^2 \rho (f^2 - f^{*2}) u_1^n = 0 \qquad \{2\}$$

with :

$$f^* = f_{c\,el} = \frac{n}{4h}\sqrt{\frac{\hat{c}^{(1)}}{\rho}}$$

for the electroded region.

or:

$$f^* = f_{cl} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$$

for the unelectroded region.

The coordinate transformations :

$$X_1 = \frac{C^{*1/2}}{M'_n} \cdot \gamma x_1 \qquad X_3 = \frac{C^{*1/2}}{P'_n} \cdot \gamma x_3$$

$$X_1 = r\cos t \qquad X_3 = r\sin t \qquad \{3\}$$

with: $\gamma^4 = \frac{n^2\pi^2}{2h^4}$

$c^* = \hat{c}^{(1)}$ (electroded region)

$c^* = \bar{c}^{(1)}$ (unelectroded region)

leads to the equation :

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,tt} + A u_1 = 0 \qquad \{4\}$$

$$Where: A = \frac{n\pi}{4}\left(\frac{f^2 - f^{*2}}{f^{*2}}\right)$$

that can be separated as $u_1(r,t) = R(r).T(t)$ in :

$$r^2R'' + rR' + R(r^2A - \upsilon^2) = 0 \qquad \{5\}$$
$$T'' + \upsilon^2 T = 0$$

For the electroded region the symmetrical solution bounded at $r = 0$ can be expressed as [6] :

$$u_1 = \sum_{m=0}^{\infty} A_m J_m(r\sqrt{A})\cos mt \qquad \{6\}$$

with m = v even integer

For the unelectroded region the symmetrical solution bounded at r =∞ is (since A is then negative)

$$u_1 = \sum_{m=0}^{\infty} B_m K_m(r\sqrt{A'})\cos mt \qquad \{7\}$$

with: m = v even integer and A' = -A

Jm and Km are respectively the Bessel function of first kind and the modified Bessel function of second kind, of order m.

At all points of the electrode edge we have to specify the continuity of u_1 and of its normal derivative. As previously discussed [15][16][6], this can be approximated by the expression of these conditions at only a discrete number q of points. It the solutions for the two regions are :

$$u_1^I = \left(\sum_{m=0} A_m J_n(r\sqrt{A})\cos mt \sin(n\pi x_2/2h)\right.$$

$$u_1^{II} = \left(\sum_{m=0} B_m K_m(r\sqrt{A'})\cos mt \sin(n\pi x_2/2h)\right. \qquad \{8\}$$

at the point $P(x_1^P x_2^P x_3^P)$ we have :

$$u_1^I(x_1^P x_2^P x_3^P) = u_1^{II}(x_1^P x_2^P x_3^P)$$

$$n_1\frac{\partial u_1^I(P)}{\partial x_1} + n_3\frac{\partial u_1^I(P)}{\partial x_3} = n_1\frac{\partial u_1^{II}(P)}{\partial x_1} + n_3\frac{\partial u_1^{II}(P)}{\partial x_3} \qquad \{9\}$$

If we choose to express these conditions at q points at the electrode edge, we obtain 2q linear relations between the coefficients A_0, A_2, A_{2q-2},..., B_0, B_2, B_{2q-2}, to obtain a solution we have to truncate the series to q terms so that the 2q relations constitute an homogeneous linear system in the 2q coefficients. The determinant of this system has to vanish to have a non trivial (0) solution in the coefficients. This condition constitutes a frequency equation that is solved numerically for the eigen frequencies. For each eigen frequency ($f_{n\mu}$ where n is the rank of the overtone and μ the rank of anharmonic), the coefficients of the eigen modes are found as the solution of the homogeneous linear system $A_{2i}^{n\mu}, B_{2i}^{n\mu}$ where i is an index varying in the interval [0,q-1].

On the whole, the eigen modes for the electroded region are given by :

$$\bar{u}_{1e}^{n\mu} = sin(n\pi x_2/2h)\bar{u}_{1e}^{n\mu}$$

$$\bar{u}_{1e}^{n\mu} = e^{j\omega t}\sum_{i=0}^{i=q-1} A_{2i}^{n\mu} J_{2i}\left(r.\sqrt{A_e^{n\mu}}\right)cos(2it) \qquad \{10\}$$

and for the unelectroded region :

$$\bar{u}_{1l}^{n\mu} = sin(n\pi x_2/2h)\bar{u}_{1l}^{n\mu}$$

$$\bar{u}_{1l}^{n\mu} = e^{j\omega t}\sum_{i=0}^{i=q-1} B_{2i}^{n\mu} K_{2i}\left(r.\sqrt{-A'^{n\mu}_{l}}\right).cos(2it) \qquad \{11\}$$

where the coefficients of the arguments of the Bessel functions are now given by:

$$A_e^{n\mu} = [n\frac{\pi}{4}(f_{n\mu}^2 - f_{ce1}^2)/f_{ce1}^2]$$

$$-A'^{n\mu}_{l} = \left[n\frac{\pi}{4}(f_{cl}^2 - f_{n\mu}^2)/f_{cl}^2\right]$$

The corresponding transformed potentiel is :

$$\hat{\phi}_e^{n\mu} = \frac{e_{26}}{\varepsilon_{22}}\bar{u}_{1e}^{n\mu}\left(\sin\left(\frac{n\pi x_2}{2h}\right) - (-1)^{(n-1)/2}\frac{x_2}{h}\right) \qquad \{12\}$$

the eigen solution must verify the orthogonality relation:

$$\iiint_{V_t} \hat{u}_1^{n\mu}\hat{u}_1^{k\lambda}dV = \delta_{nk}\delta_{\mu\lambda}N^{(n)(\mu)} \qquad \{13\}$$

with :

$$\iiint_{V_t} \hat{u}_1^{n\mu}\hat{u}_1^{k\lambda}dV = \iiint_{V_e} \hat{u}_{1e}^{n\mu}\hat{u}_{1e}^{k\lambda}dV + \iiint_{Vl} \hat{u}_{1l}^{n\mu}\hat{u}_{1l}^{k\lambda}dV$$

V_t = total volume of the plate

V_e = electroded volume

$V_l = V_t - V_e$

Forced vibrations :

The modes forced by the potential $V_0 e^{j\omega t}$ can be obtained by a linear combination of the eigen modes at $V_0 = 0$.

For the potential a supplementary term must be added [17]

$$\hat{u}_1 = \sum_n\sum_\mu H^{n\mu}\bar{u}_1^{n\mu}\sin(n\pi x_2/2h) \qquad \{14\}$$

$$\hat{\phi} = \sum_n\sum_\mu H^{n\mu}\hat{\phi}^{n\mu} + \frac{x_2 V_0 e^{j\omega t}}{2h} \qquad \{15\}$$

where as in {13} $\bar{u}_1^{n\mu}$ is either $\bar{u}_{1e}^{n\mu}$ or $\bar{u}_{1l}^{n\mu}$

$$u_1 = \hat{u}_1 - \frac{e_{26}x_2 V_0 e^{j\omega t}}{2hc^{(1)}}$$

u_1 being the forced solution.

Accounting for the fact that the orthogonality of sin $(n\pi x_2/2h)$ has already been expressed in the forms of equation {1} that are relevant for the electroded and the unelectroded regions we substitute in these equations the lateral dependence of u_1.

Then we multiply these relations by $\bar{u}_1^{k\lambda}$ and integrate respectively on V_e and V_l. The summation of these two relations gives (in the case of the Y rotated cuts of class 32 crystals where $c^{(1)}=C_{66}$ and $k^{(1)}=k_{26}$) :

$$\sum_n\sum_\mu H^{n\mu}\frac{\pi^2}{4h^2 f_{66}^2}(f^2 - f_{n\mu}^2)\iiint_{V_t} \bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dV$$

$$= (-1)^{(n-1)/2}\frac{\rho\omega^2 e_{26}4V_0 e^{j\omega t}}{C_{66}^2 n^2\pi^2}\iiint_{V_e} \bar{u}_1^{k\lambda}dV$$

Using the transformed form of the orthogonality relation accounting for the orthogonality of sinusoidal dependence in x_2:

$$\iiint_{V_t} \bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dV = 2h\iint_{S_t} \bar{u}_1^{n\mu}\bar{u}_1^{k\lambda}dS = 2\delta_{nk}\delta_{\mu\lambda}N^{(n)(\mu)}$$

We obtain:

$$H^{n\mu} = (-1)^{(n-1)/2}\frac{f^2 h e_{26}}{(f^2 - f_{n\mu}^2)n^2\pi^2 C_{66}}\frac{4V_0 e^{j\omega t}\iint_{S_e} \bar{u}_1^{n\mu}dS}{2N^{(n)(\mu)}} \qquad \{16\}$$

where S_e = electroded surface

S_t = total surface

Electrical response :

Using the constitutive equation for D_2

$$D_2 = e_{26}u_{1,2} - \varepsilon_{22}\phi_{,2}$$

$$D_2 = \frac{\varepsilon_{22}}{2h}V_0 e^{j\omega t}(k_{26}^2 + 1) + \sum_n\sum_\mu(-1)^{(n-1)/2}H^{n\mu}\bar{u}_1^{n\mu}\frac{e_{26}}{h}$$

and the relation $Y = \frac{-1}{V_0 e^{j\omega t}}\iint_{S_e} D_2 dS$

We obtain the expression of the admittance

$$Y = j\omega\left[\frac{S_e.\varepsilon_{22}}{2h}(k_{26}^2 + 1) - \sum_n\sum_\mu\frac{f^2 4k_{26}^2\varepsilon_{22}h\left(\iint_{S_e}\bar{u}_1^{n\mu}dS\right)^2}{(f^2 - f_{n\mu}^2)2n^2\pi^2 N^{(n)(\mu)}}\right]$$

The identification with the admittance of the equivalent scheme (that includes in parallel, one serial arm with the "static" capacitance and as many LC serial arms as there is modes), leads to :

$$C_s = \frac{S_e\varepsilon_{22}^S}{2h}(k_{26}^2 + 1)$$

is the capacitance at f = 0 {17}

$$C_{n\mu} = \left[\frac{8k_{26}^2\varepsilon_{22}}{n^2\pi^2 2h}\right] \frac{h\left(\iint_{Se} \bar{u}_1^{n\mu} dS\right)^2}{N^{(n)(\mu)}} \qquad \{18\}$$

the inductance of the forced mode $(n\mu)$ is:

$$L_{n\mu} = \frac{1}{(C_{n\mu} 4\pi^2 f_{n\mu}^2)}$$

As expected for stiffened mode resonators the parallel capacitance contains C_Z and an infinite sum of the opposite of the capacitances of all the modes :

$$C_0 = C_Z - \sum_n \sum_\mu C_{n\mu} \qquad \{19\}$$

In the case of one dimensional modes of an infinite plate, the sum $-\sum_n \sum_\mu C_{n\mu}$ is equal to $-k^2 Se\varepsilon_{22}/2h$ which reduce the parallel capacitance at high frequency to the usual value of $C_0 = Se\varepsilon_{22}^S/2h$ whereas $C_Z = Se\varepsilon_{22}/2h$ is the low frequency capacitance.

IV OPTIMIZATION OF BERLINITE RESONATORS FOR FILTER APPLICATIONS.

IV.1 Results of the computations for resonators having circular electrodes.

The model was applied to compute the resonances frequencies of the fundamental mode and of its anharmonics, the corresponding modes and the values of the equivalent scheme corresponding to each mode for resonator having a thickness of 20 microns, circular electrodes with radius ranging from 0.15mm to 0.4mm and a mass loading ranging from 0.5% to 1.5%. The constants used for the calculations where those measured by D.S.Bailey et al. [18]. The results concerning the resonance frequencies are given on figure 12 in function of the dimensionless quantity $\sqrt{S}/2h$, where S is the area of the electrodes and 2h the thickness of the plate. This representation of the electrodes dimension was chosen for theoretical reasons and also to have an easy comparison between results concerning resonators having electrodes of different shapes.

On Figure 12 we can observe that the sensitivity of the resonance frequency to the electrode dimensions, increase as the mass loading increases and also as the electrode radius is lowered. As well known, it can be observed that anharmonic modes appear, with the largest value of the mass loading, for a much lower value of the electrode radius; so that the interest of using a mass loading as low as permitted by the dispersion of the plate thickness appears again.

We have represented on figure 12 the cut-off frequency of the unelectroded plate and also the cut-off frequencies of the electroded plates for two values (1.% and 1.5%) of the mass loading. The modes noted 2 and 3 are the symmetrical anharmonics with two nodal lines.

On figure 12 we have a direct representation of the frequency lowering due to a given mass loading (difference between the cut-off frequency of the unelectroded plate and the resonance frequency of the partially electroded plate); we can observe that this frequency lowering is greatly dependant of the dimensions of the electrodes.

On figure 13 are represented the same curves using the reduced frequency defined by the relation : $\Omega^2 = (f^2 - f_{ce1}^2)/f_{ce1}^2$ (see the definition of A in relation {4}).

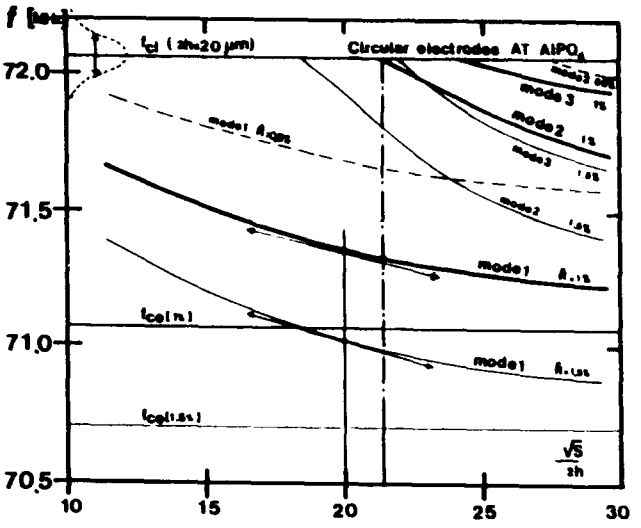


Figure 12: Thickness shear modes of AT berlinite resonator with circular electrodes.(Plate thickness = 20 microns , mass loading 0.5, 1., 1.5 %).

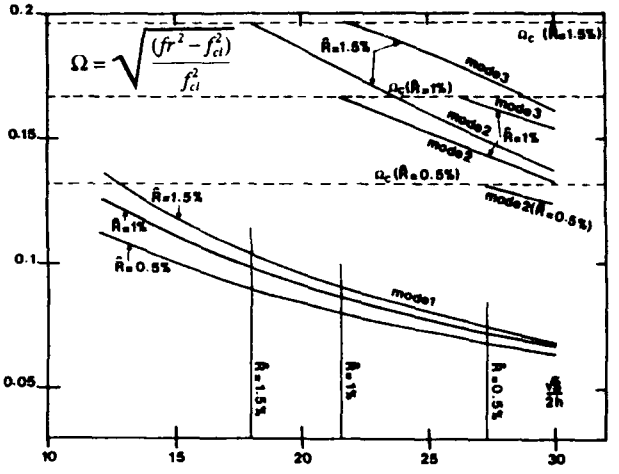


Figure 13: Thickness shear modes of AT berlinite with circular electrodes in reduced frequency coordinates .

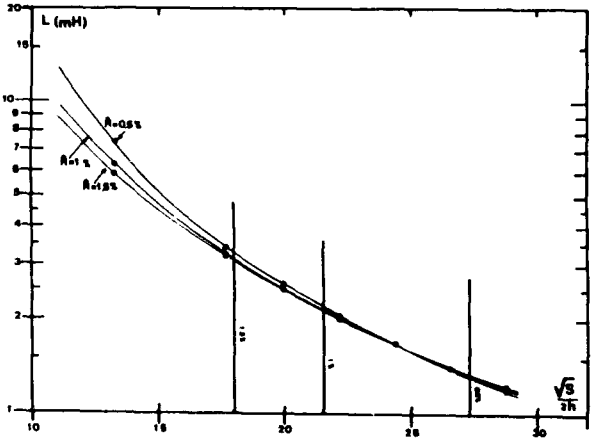


Figure 14: Inductance of AT resonators with circular electrodes.

The inductances corresponding to the same resonators were computed numerically using the relations previously given. On figure 14 we can observe that the value of the inductance is strongly dependant on the electrode radius and is quite independent of the mass loading, at least fo the larger values of the electrode radius. It can also be observe that the sensitivity of this quantity to the dispersion of the electrode dimensions increases rapidly for the small values of the electrode radius.

For the values, here considered, of the mass loading, the sensitivity of the inductance to variations of the mass loading vanishes for a value of the normalized electrode dimension situated near$\sqrt{S}/2h=25$ (depending on the exact value of the mass loading).

IV,2 Results obtained for resonators having elliptical electrodes

Such electrode shapes are interesting for theoretical reasons [6]. The same calculations as those made for circular electrodes were performed for resonators having elliptical electrode with an axis ratio corresponding to the lateral anisotropy of the AT cut of berlinite, and also for elliptical electrodes having an axis ratio comprised between .3 and 2.5. On figure 15 are represented for elliptical electrodes respecting the lateral anisotropy, in reduced coordinates, the variations of the resonance frequencies of the fundamental mode and of its anharmonics as a function of the dimensions of the electrodes. On this figure where, as in figure 13 we have considered three values of the mass loading (0.5%, 1.0%, 1.5%), we can observe that the anharmonic modes appear, for each value of the mass loading, at much greater values of the electrode surface.

On figure 16 are represented the computed modes of a resonator made from a plate 20 microns thick, having such electrodes with a dimension along the x axis 2a=0.55mm and a mass loading of 1.5% that lead to have 3 modes: the fundamental and two anharmonics. The resonance frequencies and the motional inductances of the three modes are indicated on the figure. We can observe that the motional inductance of the second mode was found by the numerical calculation to be extremely high, in fact, this inductance is infinite and the mode is not piezoelectrically excitable, since in the r and t coordinates introduced in § III.3 this mode has a centre of symmetry. This property only appears when a/b is exactly equal to $(M_n/P_n)^{1/2}$, but exists for any values of the mass loading and of the electrode surface.

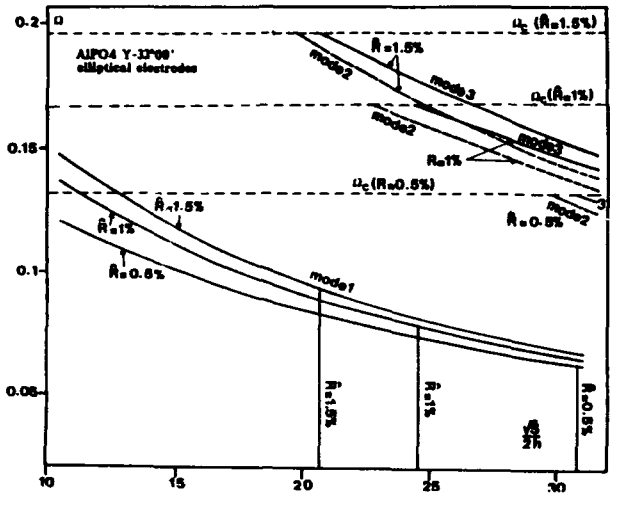


Figure 15: Thickness shear modes of resonators with elliptical electrodes respecting the in plane anisotropy of the AT cut.

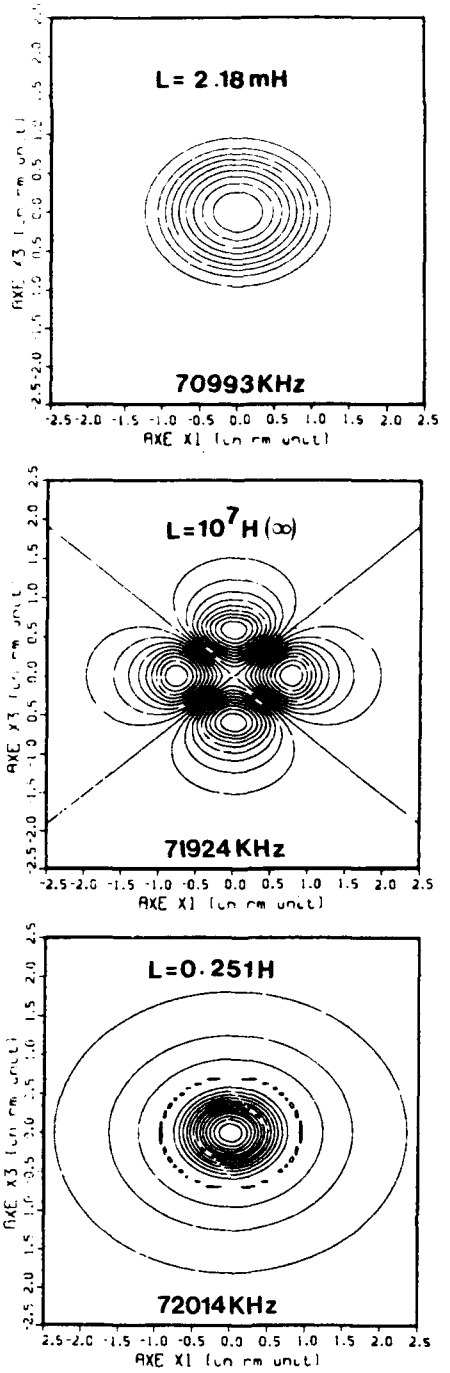


Figure 16: Example of computed modes for AT resonators with elliptical electrodes.

The values of the inductance of the fundamental mode computed for three mass loading (0.5%, 1%, 1.5%) as a function of the reduced electrode dimension is displayed in figure 17. In this figure, we can observe that, at equal surface, the inductances are not very different of those found with circular electrodes. But the elliptical electrodes respecting the lateral anisotropy, permit to have monomodes responses with much larger surfaces, so that, lower inductances can be obtained together with a reduced value of the sensitivity to the dispersion of the electrode dimensions.

Zero of the sensitivity of the inductance to variation of the mass loading.

On figure 18 are represented, in function of the mass loading, the values of the motional inductance of resonators having elliptical electrodes respecting the lateral anisotropy, a thickness 2h=20microns, and two different electrode surfaces. The figure 18a corresponds to the zero sensitivity point observed in figure 17 (at $\sqrt{S}/2h$=22.75 i.e. for a large value of $\sqrt{S}/2h$). We can observe that a very reduced sensitivity to the variations of the mass loading exists for values situated between .75% and 1.15%. On figure 18b is displayed the case of a zero sensitivity appearing for smaller electrodes dimensions ($\sqrt{S}/2h$=15.52), and a higher value of the mass loading. In this case the value of the inductance is much higher than those found with the larger value of $\sqrt{S}/2h$.

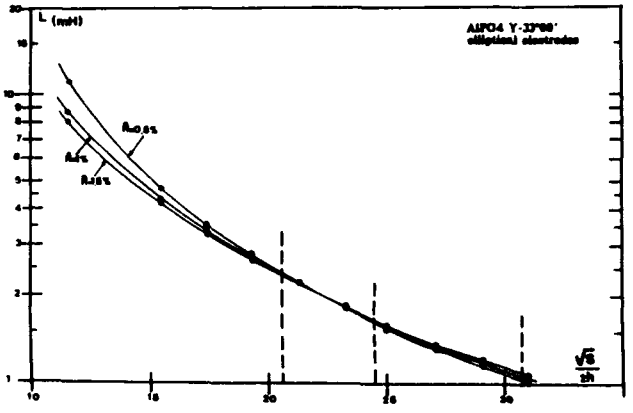


Figure 17: Inductance of AT berlinite resonators with elliptical electrodes.

Existence of a minimum of the "dimension" of the vibration mode:

On figure 19 is displayed a quantity that represents nearly the total radius filled by the mode. The quantity dmax given on this figure is exactly the radius such that $u^2(dmax) < 2.10^{-4} x u^2max$, where u^2max is the maximum of u_1^2. On figure 21 we can observe that dmax, computed here for elliptical electrodes with an axis ratio a/b=1.3, has for the three mass loading considered, a minimum. This means that when the integration of several resonators on the same plate of as reduced dimensions as possible is considered, it exist an optimum choice, for each mass loading , of the electrode dimension and of the distance between resonators.

Influence of the electrode shape on the apparition of anharmonics:

On figure 20 are represented, in reduced coordinates, for the AT cut of berlinite, the resonance frequencies of resonators having a constant surface and a different value of the ratio of the two axis of the ellipse. We can observe on this figure that excepted for very small a/b ratio, the resonance frequency of the fundamental mode is nearly constant, whereas large variations of the resonance frequencies of the anharmonic modes are observed. The spacing between the fundamental mode and the anharmonics is extremal for the values of a/b situated in the interval [1.2, 1.4]. These values of a/b correspond to the most interesting designs from the point of view of the sensibility of the apparition of anharmonics to the variations of either the a/b ratio or the mass loading. When a/b is strictly equal to $(M_n/P_n)^{1/2}$, the 1st anharmonic is not piezoelectrically excited so that a further advantage is gained.

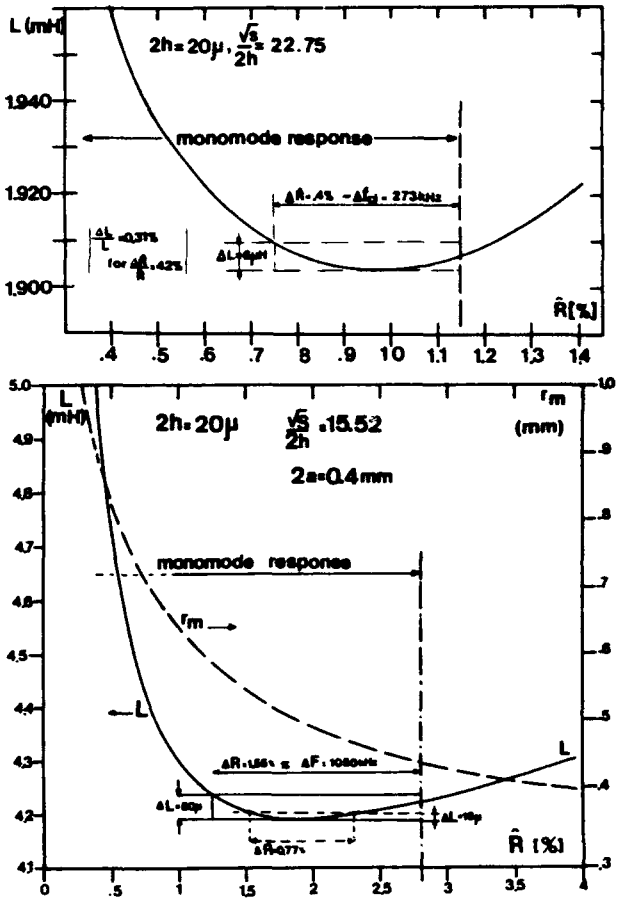


Figure 18: Zero of sensitivity of the inductance to variations of the mass loading. 18a case of large electrodes, 18b case of small electrodes; rm is the "radius" of the mode.

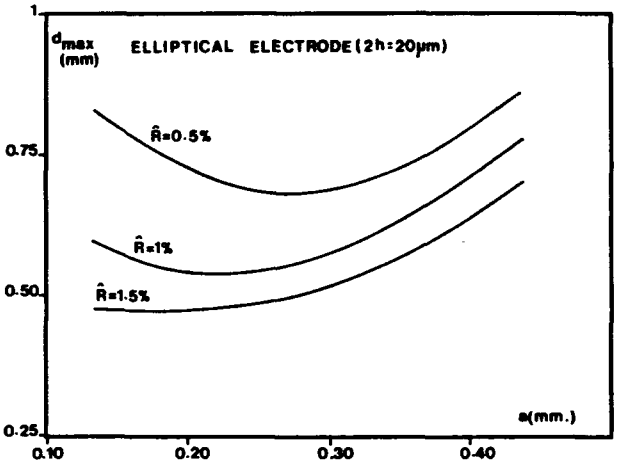


Figure 19: "Dimension" of the mode (elliptical electrodes).

On figure 21 are represented the computed modes of a resonator having a large a/b (2.5) ratio and an usual mass loading (1.0%), these modes are drawn at the same scale as in figure 16 . In this case, there is only one anharmonic mode. We can observe that it has a geometry which is very different of that of the first anharmonic of figure 16. If we compare now the geometry of modes of resonators having the same surface (figure 22), we can observe that the in plane anisotropy of AT

berlinite makes that the mode shape is different of the shape ot the electrodes except for the case where $a/b=(M_n/P_n)^{1/2}$, this is particularly visible in the case of a circular electrode (a/b=1.).

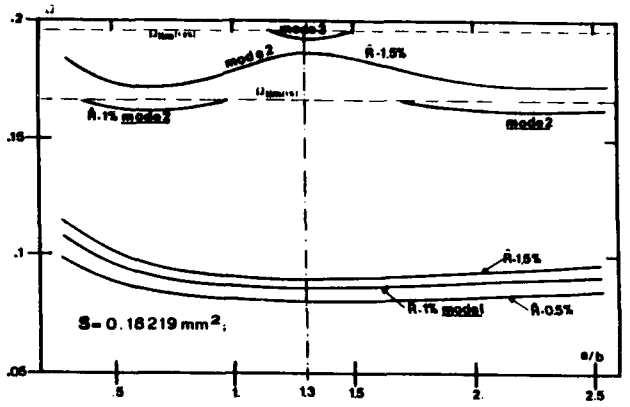
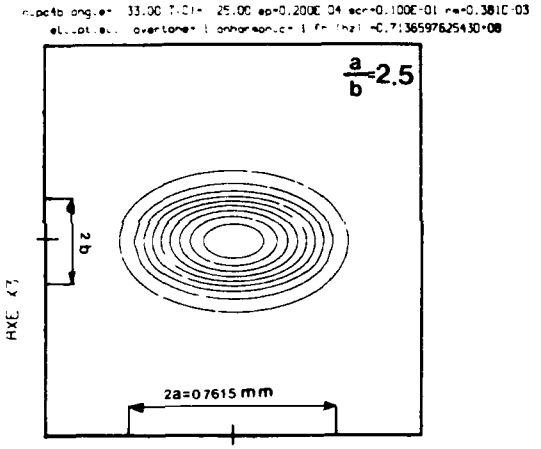


Figure 20: Thickness shear modes of elliptical electrodes with constant surface and varying axis ratio.



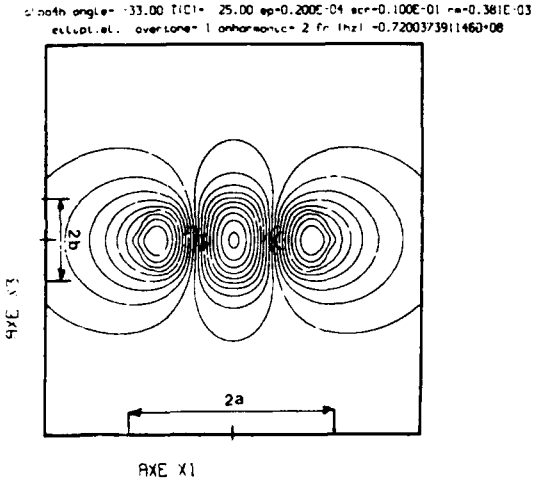
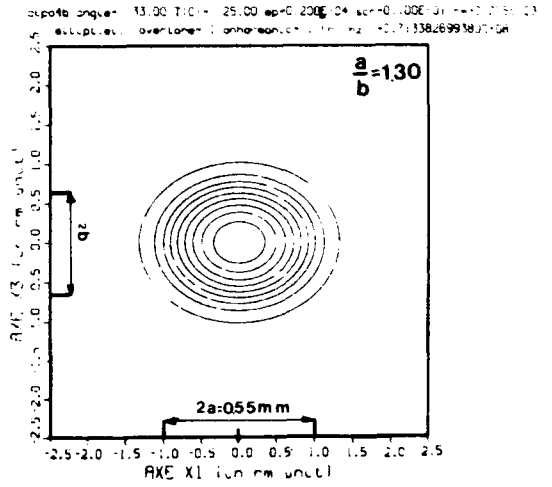
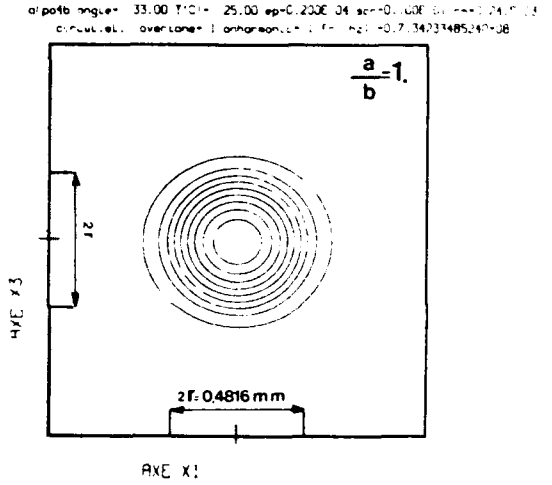


Figure 21: Computed modes of an AT berlinite resonator with elliptical electrodes having a large a/b ratio.





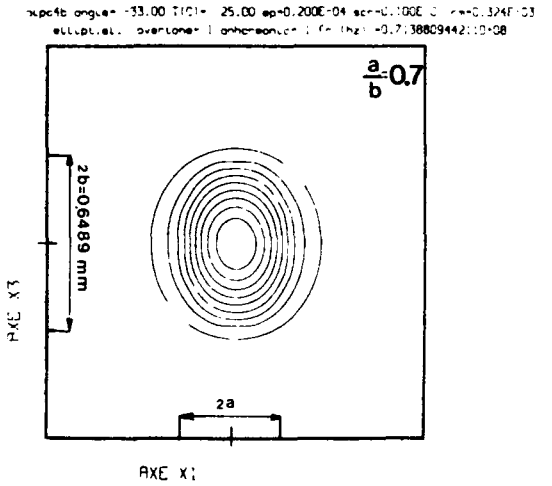


Figure 22: Comparison of the computed fundamental modes of berlinite resonators with elliptical electrodes having a constant surface and different axis ratios.

IV.3 Comparison of resonators having electrodes of different geometry:

In table 2 we compare the computed resonance frequencies and inductances of resonators having the same thickness, the same electroded surface and the same mass loading but different electrode geometries. In this table, we can observe that the resonance frequencies of the fundamental mode and the inductance are not very sensitive to the electrode shape (for these values of the surface and of the mass loading). However the lowest values of these two quantities are observed for the elliptical electrodes that respects the lateral anisotropy. It can be also noticed that the rectangular and square electrodes leads always to slightly higher inductances than the corresponding elliptical or round electrodes. This is due to the inefficiency of the corners in this type of electrodes to give a good contribution to the dynamical capacitance, or, otherwise said, this is due to the fact, that, unless the mass loading is very high, the mode has a more smoother shape than the electrodes.

If we consider now globally, the conditions of apparition of the first symmetrical anharmonic mode, the value of the motional inductance in monomode conditions, and the sensitivities, the electrode shapes having a ratio a/b corresponding to the in plane anisotropy of AT berlinite are much better than the others. Among all the electrode shapes having the proper a/b ratio , the elliptical ones are from far the best.

TABLE 2 Comparison of the inductance of resonators with different electrode geometries (constant surface=.1822mm2)

$S = 0.182\,mm^2$; $\hat{R} = 1\,\%$; $2h = 20\,\mu m$

GEOMETRY	a/b	Frequency (kHz)	Inductance (mH)
Rect	0,7*	71367.	2.45
Square	1	71362.	2.24
Rect.	1,3	71342.	2.22
Rect.	1.9*	71349.	2.22
Ellipt.	0.70*	71388.	2.25
Round.	1	71342.	2.19
Ellipt.	1.3	71338.	2.17
Ellipt.	2.0*	71365.	2.21

* response having anharmonics

V OPTIMIZATION OF AT QUARTZ RESONATORS FOR FILTER APPLICATIONS:

V.1 Resonators having circular electrodes.

The model was applied to compute the resonances frequencies of the fundamental mode and of its anharmonics and the values of the equivalent scheme for resonators having a thickness of 23.0152 microns, circular electrodes with varying radii, and, a mass loading equal to .5%, 1.0% and 1.5%. For the computations we have used the B.B.L. constants of quartz [19]. The results are represented in figures 23 (resonance frequencies) and 24 (motional inductance). All the observations previously made about berlinite concerning the sensitivities are also pertinent in this case.

V.2 Resonators having elliptical electrodes respecting the lateral anisotropy:

In figures 25 and 26 are represented the computed values of the resonance frequencies and of the inductance. In this case we can observe that this electrode shape has the advantage to allow a much larger surface for the electrodes before the apparition of anharmonics. With a low mass loading, we have also again interesting sensitivity properties for the resonance frequency and the inductance.

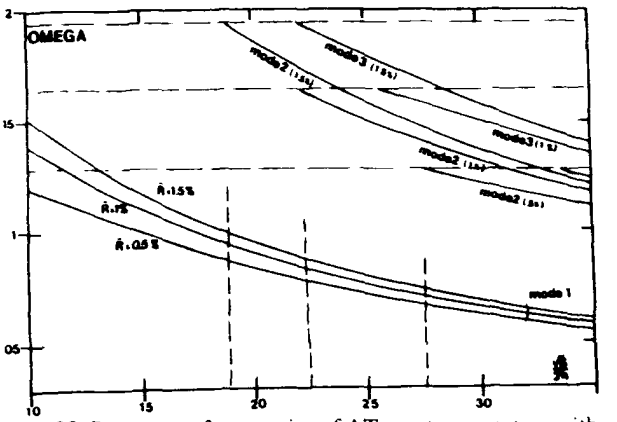


Figure 23: Resonance frequencies of AT quartz resonators with circular electrodes.

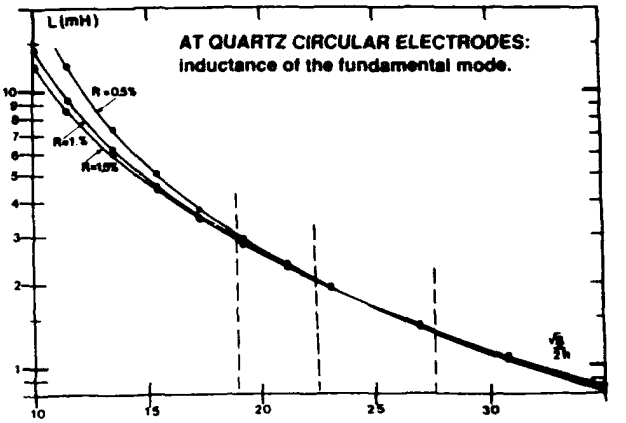


Figure 24: Motional inductance of AT quartz resonators with circular electrodes.

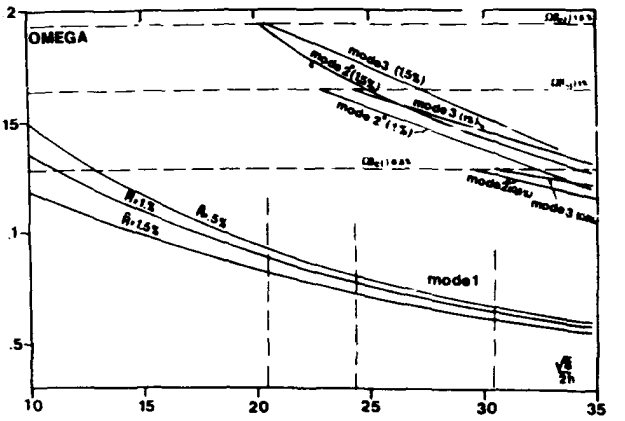


Figure 25: Resonances frequencies of AT quartz resonators with elliptical electrodes.

In order to display some properties of the modes of resonators having such an electrode geometry, we have represented on figure 27 the computed mode shapes of the anharmonics existing of a resonator (2h=23.0152 microns) having very larges electrodes (2a=1mm) and a large mass loading (3%). Four of these anharmonics are centro-symmetrical in the (r,t) coordinates and thus not piezoelectrically excitable.

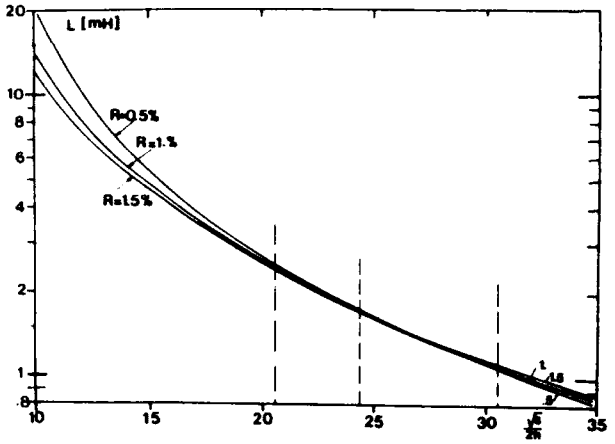


Figure 26: Motional inductance of AT quartz resonators with elliptical electrodes.

"Dimension" of the vibration mode.

On figure 28 are displayed for resonators having the same parameters as in the two preceeding figures the variations of the quantity rc such that $u^2(rc)<=2.10^{-5}.\ u^2(a/2)$. This quantity represents, as dmax in §IV, the radius in which almost all the energy of the resonator is confined. We can observe also, on figure 28 that rc present a minimum, for each value of the mass loading. For the largest electrode dimensions, rc increase proportionnally to the electrode dimensions.

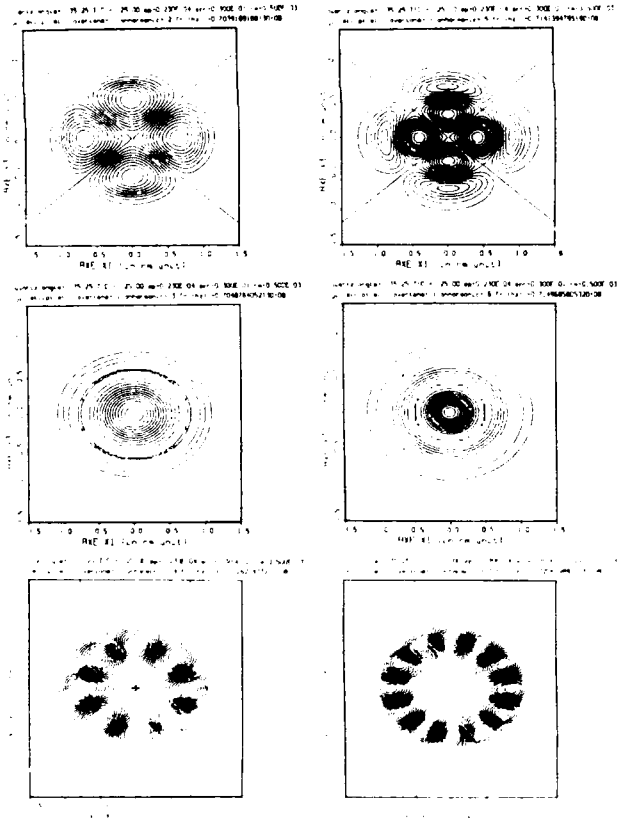


Figure 27: Symmetrical anharmonics of a resonator with elliptical electrodes having a large mass loading (3%) and a large surface (2a=1.mm).

VI. COMPARISON WITH EXPERIMENTAL RESULTS

VI.1 Berlinite resonators.

Computed results were compared to experimental ones for AT berlinite resonators made with resonance frequencies in the range 8-10MHz and round or elliptical electrodes. A quite good agreement was found for the frequencies (the experimental ones are always higher by a few percent [1]) and for the mode shapes. The experimental inductances are systematically lower than the computed ones. This results much probably of a too low value of the computed coupling coefficient (see the expression [18]). The experiments have shown that the computed values of the inductance are to be corrected by a factor of the order of 0.70. The correction factor for the computed coupling coefficient of the shear mode of the AT cut is much probably of the order of 1.2. From other experiments [1] we can estimate that the Cij constants of reference [18] are lower to that of the high Q material now available by a few percent. Much probably, at least one of the piezoelectric constants given in reference [18] is lower to that of the material used in the present experiments.

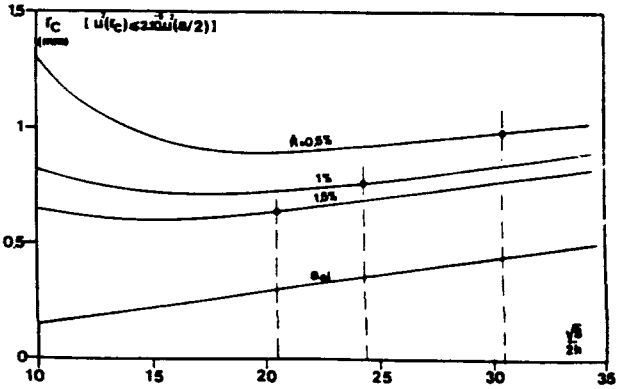


Figure 28: "dimension" of the mode for AT quartz (Fundamental mode with elliptical electrodes). rc here is such $u^2(rc)<2.10^{-5}.u^2(a/2)$. (the curve ael represents the variation of half the great axis of the ellipse).

In figures 29a,b,c are compared experimental and computed results for a resonator having circular electrodes (2h=.152mm, 2r=3.58mm, R=1.40%). The experimental and computed frequencies and motional inductances are respectively fexp.=9.502125MHz, fcal.=9.352265MHz; Lexp.=11.85mH, Lcal.=17.40mH.

In figure 30 is considered the case of a berlinite resonator having elliptical electrodes (2h=.173mm, 2a=2.890mm , 2b=2.216mm , R=1.35%). The experimental frequency is 8.331MHz whereas the computed one is 8.2553MHz. The experimental inductance is 38mH and the computed one is 54mH. As expected from the calculation no anharmonics are present in the response curve (figure 29a), and the mode shape has a geometry very similar to the computed one(figures 29b & 29c).

VI.2 Quartz resonators.

For quartz resonators a very good agreement, always in accordance with the accuracy of the different experimental quantities or parameters used in the calculations, was found between the experimental and the computed results.

In figure 31 are compared the results relative to a 12 MHz resonator having elliptical electrodes (2h=.130mm ,2a=2.90mm , 2b=2.286mm , R=1.32%). The experimental and the computed frequencies and inductances are respectively fe=12.711 ,fc=12.711MHz; Le=17.7 mH, Lc=18.7 mH.

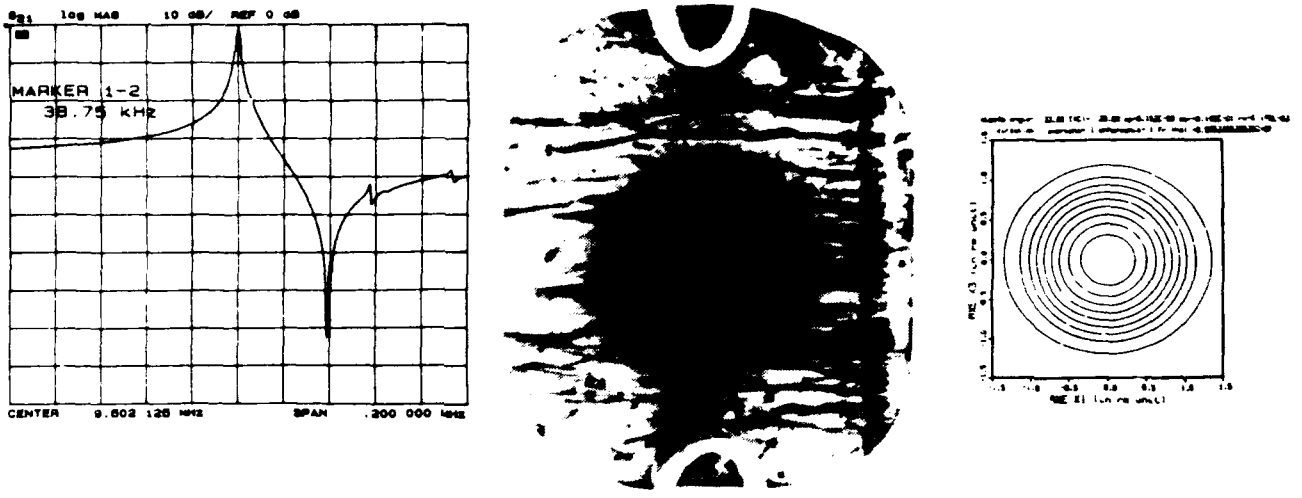


Figure 29: AT berlinite resonator with circular electrodes. 29a:Response curve, 29b:experimental mode shape, 29c: computed mode shape.

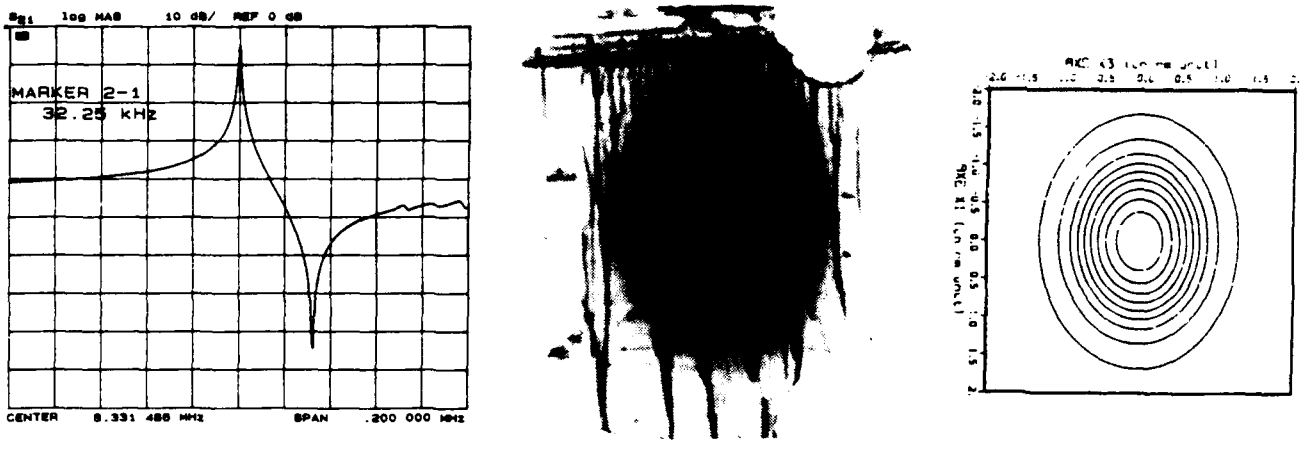


Figure 30: AT berlinite resonator with elliptical electrodes. 30a: response curve, 30b:experimental mode shape, 30c:computed mode shape.

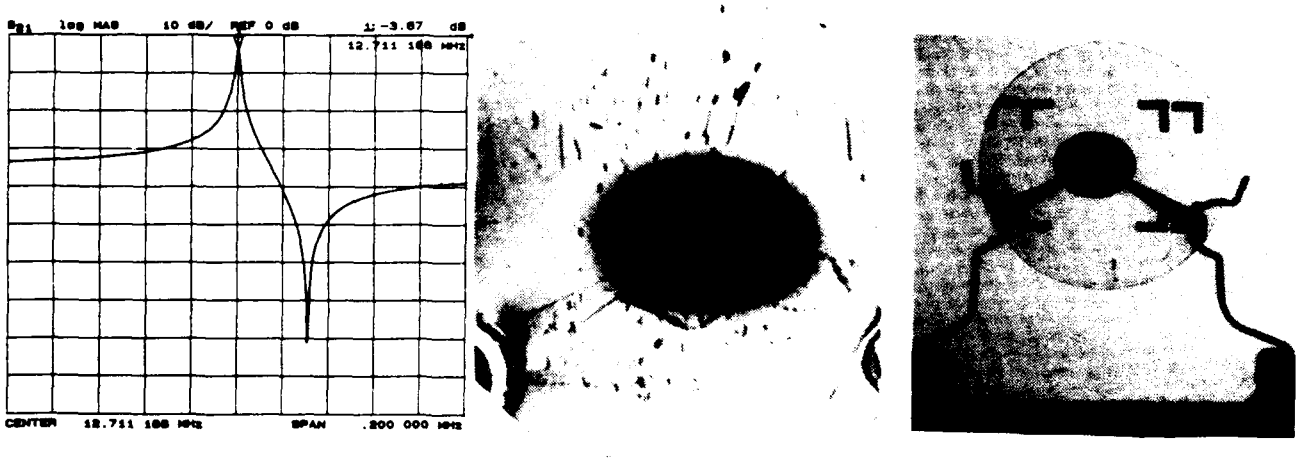


Figure 31: AT quartz resonator with elliptical electrodes. 31a: response curve, 31b: experimental vibration mode, 31c:photography of the resonator.

A comparison was also made for VHF resonators with elliptical electrodes. These resonators were made using chemically milled very thin plates of quartz, commercially available [8]. Elliptical electrodes (2a=0.9mm,2b=0.5mm) with a mass loading of 0.5% were deposited on these plates which have a thickness of about 23.7 microns. On figure 32 is represented the electrical response of one of these resonators together with a photography. It can be observed that as expected from the calculation, no anharmonics are present. The computed inductance is 1.66 mH whereas the experimental values were found in the range 1.52-1.58 mH.

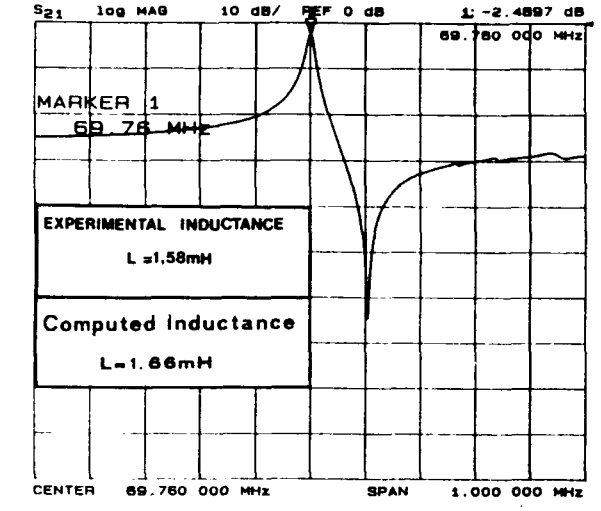


Figure 32: Response curve (32a) and photography (32b) of a V.H.F. resonator with an elliptical electrode (0.9*05mm) having a mass loading of 0.5% giving no anharmonics (AT quartz).

IV Conclusion

An investigation of the properties of V.H.F. resonators using the fundamental mode of AT berlinite and AT quartz was made in view to optimize the design of resonators used in the I.F. Filter for the G.S.M. radiotelephone . In a first step the angular position of the AT cut of berlinite was researched. In a second step, using a model of resonator, the influence of several physical parameters on the values of the elements of the equivalent scheme was studied.

These investigations have shown the interest to use low values of the mass loading together with electrodes as large as permitted by the dispersion of the plate thickness. This permits to obtain low values for the motional inductance and also zero or very reduced sensitivities to several dispersions.

We have shown that elliptical electrodes with a ratio a/b respecting the lateral anisotropy of the material allow to have much larger electrode surface than with any others before the apparition of unwanted anharmonic modes. This permits lower values of the inductance and still better sensitivity properties.

The calculations have indicated that, at least for the smaller values of the mass loading, the total surface occupied by the vibration mode of resonators with such elliptical electrodes presents a minimum. This permits the optimization of filters integrating several resonators on the same plate.

Acknowledgments: The authors wish to acknowledge X.Buisson and R.Arnaud of S.I.C.N. Annecy (France) for kindly providing the high quality berlinite crystals used in this study, They acknowledge the assistance of G.Caumont and A.Daniel for the preparation of the samples and they thank particularly Mrs M.P.Louis for the elaboration of this manuscript.

References:

[1] E.Philippot, A. Ibanez, A.Goiffon, X. Buisson, R. Arnaud, B.Capelle, A.Zarka, Y.Toudic, J.Schwartzel, J.Détaint Proc. 4th European Frequency and Time Forum p.267 (1990).
[2] D.S. Stevens, H.F.Tiersten J.Acoust.Soc.Amer. vol.79 n°6 p1455 (1979).
[3] J.Detaint, J.P.Aubry, J.Ph.Michel, A.Zarka, B.Capelle, Y.Toudic, R.Lefèvre, E.Philippot, J.Schwartzel Proc. 1989 U.R.S.I. International Symposium on Signal Systems and Electronics p.433 (1989).
[4] J.P. Aubry, J.Ph. Michel, J.L.Le Corre, D.Blonde, P. de la Fournière Proc. 3rd European Frequency and Time Forum p.387 (1989).
[5] H.F.Tiersten, R.C. Smythe J. Acoust; Soc. Amer. vol. 65 n°6 p. 1811 (1979).
[6] J.Detaint , J. Schwartzel, C. Joly, E.Philippot proc. 41st Annual Frequency Control Symposium p. 314 (1987).
[7] O.Cambon, A.Goiffon, E.Phillipot, J.P.Aubry, J.Detaint Proc. 4th European Frequency and Time Forum p.610 (1990).
[8] J.Kent Proc.12th Piezoelectric Devices conference vol.1 p1-7 (1990).
[9] T. J.Lukaszek I.E.E.E. trans. Sonics Ultrason. vol. SU 16 n°4 p238 (1971).
[10] J.Detaint, H.Carru, J.Schwartzel, C.Joly, B.Capelle, A.Zarka proc. 43th Annual Frequency Control Symposium p.563 (1989)
[11] Filtres Electriques, M.Hassler, J.Neirynck Presses Polytechniques Romandes Lausanne (1981) p.98.
[12] H.F. Tiersten J. Acoust. Soc. Am. vol. 59 p.879 (1976).
[13] B.Dulmet, H.F.Tiersten Proc. 1986 I.E.E.E. Ultrasonics Symposium p. 355 (1986).
[14] J.Detaint, H.Carru, J.Schwartzel, B.Capelle, A.Zarka proc. 42nd Annual Frequency Control Symposium p. 19 (1988).
[15] K.Nakamura, M. Shimizu proc. 1976 I.E.E.E. Ultrasonics Symposium p.606 (1976).
[16] J.Detaint , A.Zarka, B.Capelle, Y.Toudic, J.Schwartzel, E.Philippot, J.C. Jumas, A.Goiffon, J.C.Doukhan. Proc. 40th Annual Frequency Control Symposium p. 101 (1986).
[17] R.Holland, E.P. Eernisse Design of Resonant Piezoelectric Devices M.I.T.Press (1969).
[18] D.S. Bailey, J.C. Andle, D.L. Lee, W. Soluch, J.F. Vetelino, B.H.T. Chai. Proc. 1983 Ultrasonics Symposium p.325 (1983).
[19] R. Bechmann, A.Ballato, T.J. Lukaszek Proc. IRE vol.50 p.1812 (1962) .

MONOLITHIC FILTERS USING STRONG PIEZOELECTRICS

S.A.Sakharov, I.M.Larionov, V.A.Issaev

Scientific-Research Institute "Phonon"
Moscow 105023,Krasnobogatyrskaya 44, U.S.S.R.

Abstract

This article gives information about developments of monolithic filters using quartz material, lithium tantalate and $La_3Ga_5SiO_{14}$ (langasite). Their principal parameters and angle of cut are reported which ensure obtaining minimum or zero frequency-temperature coefficient over the operating temperature range.

Introduction

At present monolithic filters (MF) are widely used in electronic engineering for frequency selection in communication equipment. The field of application of these devices extends from several megahertz up to hundreds megahertz at the fundamental frequency and up to hundreds megahertz for monolithic filters using inverted mesa and operating at overtones. The range of realizable bandwidths extends from thousands of a percent (monolithic crystal filters) up to 0.3% and from 0.8% up to 4% for lithium tantalate monolithic filters [1] which is shown in Figure 1.

From the figure it follows that over the range of pass bandwidths from 0.3% up to 0.8% a gap existed up to the present time for monolithic filters, these filters were realized as filters with discrete elements with added inductances, which increased considerably their volume and weight and decreased their reliability.

In order to cover this pass band range piezoelectric materials are necessary having intermediate properties regarding electromechanical coupling coefficient and frequency-temperature coefficient (FTC) between quartz and lithium tantalate crystals. The following materials can be attributed to these: berlinite ($AlPO_4$) [2] , lithium tertaborate ($Li_2B_4O_7$) [3] , these crystals being at present applied in piezoelectric engineering. To the new promising materials having intermediate properties belong crystals having the structure of potassium gallium germanite $La_3Ga_5SiO_{14}$, $Ca_3Ga_5GeO_{14}$, $Sr_3Ga_5GeO_{14}$. These trigonal crystals are isotropic to quartz crystals (α - SiO_2), belong to the symmetry class 32 (Group P321) and have as quartz crystals both left-handed and right-handed modification. They are grown by using Tchokhralsky method (T = 1400 - 1500°C) [4] , [5] .

Compared to quartz they exhibit lesser acoustic wave attenuation, their piezoelectric constants e_{11} and e_{14} and diagonal elastic modules being approximately twice as great as those of quartz. Therefore, the existence of cuts with small and zero values of frequency-temperature coefficient is possible for these materials with supposed high electromechanical coupling factor value.

$La_3Ga_5SiO_{14}$ and $LiTaO_3$ - filters

Monolithic filters using langasite crystals have satisfactory frequency-temperature coefficient (approximately equal to $1.3 \times 10^{-6}/1°C$) for this pass bandwidth range and, as it was noted above, have a number of advantages over discrete quartz crystal filters, operating within this range.

For strong piezoelectrics such as lithium niobate and lithium tantalate by using the second type of vibrations it is possible to decrease the temperature coefficient of frequency and the influ-

CH2965-2/91/0000-181 $1.00 © 1991 IEEE

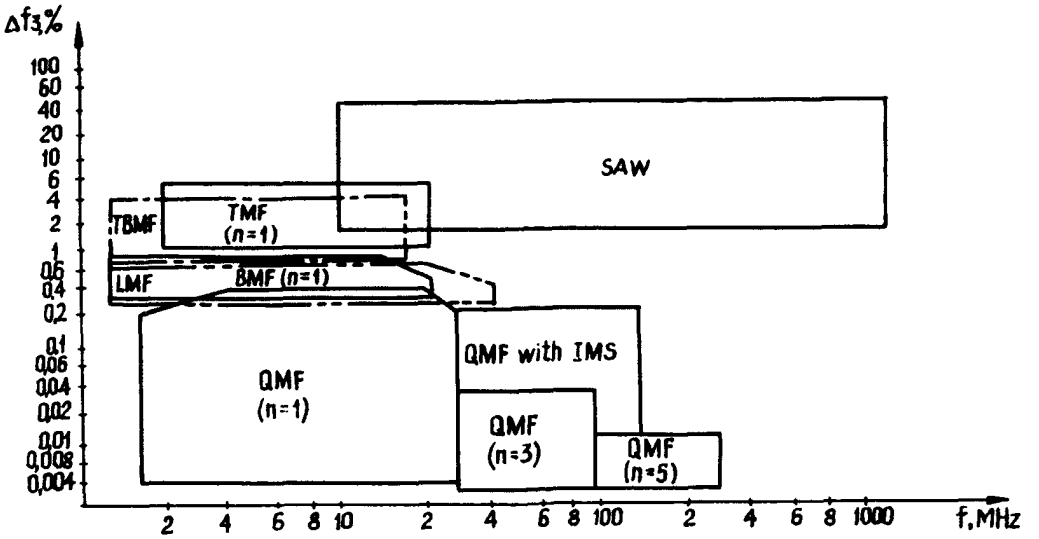


Fig. 1: Frequency ranges and pass bands of monolithic filters using different piezoelectric crystals

ence of piezoelectric effect which enables to design intermediate-band frequency monolithic filters.

Figure 2 shows piezoelectric elements for lithium tantalate monolithic filter of the second order with particular resonators operating in the second type of thickness shear and thickness twist modes. The shown piezoelectric elements have driving electrodes located at one of the main faces of piezoelectric substrate, and "overlapping" them "passive" electrodes of particular resonators are located at the other main face.

In a piezoelectric element shown in Fig. 2a, the thickness-twist shear mode is driven artificially at the expense of corresponding selection of inactive electrode and creation of oppositely directed electrical fields in two underelectrode areas each constituting a particular resonator.

The adjustment of electromechanical coupling factor (k_s) of operational modes with constant electrode area of particular resonators is achieved by changing particular resonator interelectrode distance of particular electrode areas, constituting resonators.

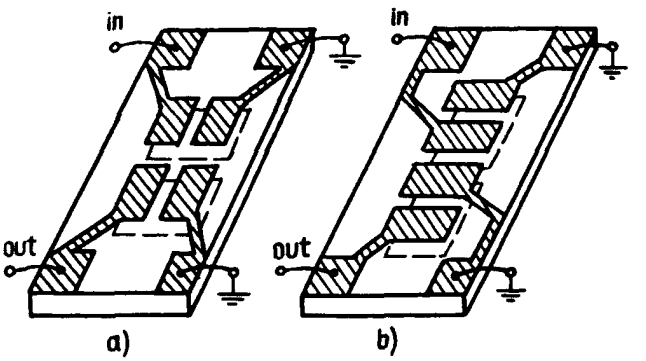


Fig. 2: Second-order monolithic filters operating
a) in thickness-twist mode
b) in thickness-shear mode of 2nd type

In a piezoelectric element shown in Fig. 2b), the change of k_s is maximum, because particular resonators are located in the direction of maximum acoustic coupling coinciding with the Y-axis of a crystal element.

The typical frequency attenuation characteristic of filters shown in Fig. 2a) and b) is given in Fig. 3.

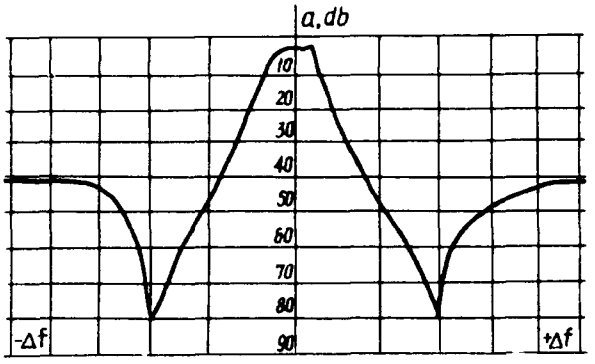


Fig. 3: Typical frequency attenuation characteristic of the 2nd type monolithic filters operating in thickness-twist and thickness-shear modes of the 2nd type

Figure 4 shows temperature coefficient of frequency curves for three monolithic filters manufactured of lithium tantalate.

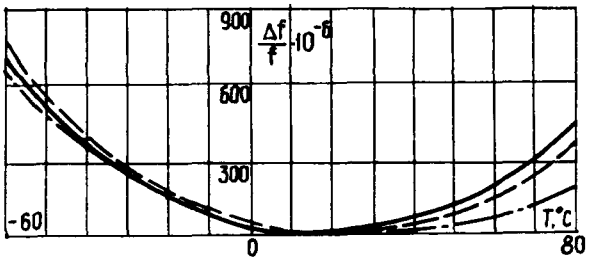


Fig. 4: TCF curves of three different types of lithium tantalate monolithic filters
1 - MF operating in thickness-shear mode of the first order
2 - MF operating in thickness-twist mode of the second order
3 - MF operating in thickness-shear mode of the second order

Conclusion

The authors hope that calcium gallium-germanite crystals and monolithic filters operating in thickness-shear and twist modes of the second type will find their application in piezoelectric engineering.

References

[1] Bronnikova E.G., Larionov I.M. Monolithic filters, Electronic engineering. Series "Radiodetails and Components", issue 3(68), 1987.

[2] Detant J., Zarka A., Capelle B. et al. "Berlinite: characterization of crystal with a low water concentration and design of bulk wave resonators", IEEE, 40th AFCS, pp.101-104, 1986.

[3] Katsuoshi Fukuta, Jisuburo Usizawa, Hitoshi Suzuki, Yasuo Ebata and Sadao Maisumura. Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Applications. Japanese of Applied Physics, vol.22, Supplement 22.2, pp.140-142, 1983.

[4] Kaminskij A.A., Silvestrova I.M., Sarkisov S.E. Investigation of Trigonal $(La_{1-x}Nb_x)Ga_5SiO_{14}$ Crystals-Physica Status Solidi (a), v.50, pp.607-620, 1983.

[5] Andreev I.A., Doubovik M.Ph. New piezoelectric material - $La_3Ga_5SiO_{14}$ with zero frequency temperature coefficient of elastic vibrations, Letter to Journal of Theoretical Physics, v.10, issue 8, 1984.

[6] Bronnikova E.G., Larionov I.M., Issaev V.A., Silvestrova I.M. Pissarevsky Y.V. Monolithic filters and resonators using new piezoelectric material gallium silicate of lanthanum. Electronic Engineering, Series "Radiodetails and Components", issue 2(63), 1986.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INFLECTION TEMPERATURE OF CONTOURED AT-CUT RESONATORS

Joseph Dutka
Crystek Corporation
Fort Myers, FL 33906-6135

John Erasmus, Consultant
5022 Fairfield Drive
Fort Myers, FL 33919

John H. Sherman, Jr., Consultant
2022 Woodcrest Drive
Lynchburg, VA 24503

Abstract

Conveniently programmable design equations are known for the routine design of plano-plano AT-cut resonators. This is not equally true of contoured resonators. A number of effects of contour are discussed in various places in the literature, but some remain uninvestigated. This paper reports an inadequately defined and executed experiment to probe the existence of a variation of the inflection point of the temperature-frequency characteristic of a contoured AT-cut resonator with contour. Two of the authors are now retired and no longer positioned to pursue the matter further. We read this evidence to indicate that the existence of an effect is not here established. The intended experiment and the deficiencies of the one actually performed are described in the hope that a better investigation might be devised and reported by some other investigator.

Introduction

The efficient design of contoured AT-cut resonators is still the esoteric specialty of a few journeymen crystal engineers. For two categories of requirement an inefficient design process is not disabling. These are the very large volume requirement, possibly with somewhat loose temperature-frequency specifications, suitable for pipe bevel contouring, and the very precise resonator as used in frequency standards, for which the cost of fabrication is sufficient that an expensive design effort is well justified. The orphan is the small volume requirement for resonators of only moderate precision. The effort to design these is comparable to that for either of the others. The market for them, however, does not readily support this effort. This is at least one contributing cause for the avoidance of contoured crystals in low volune applications. The availability of low cost frequency divider circuits and synthesizers has made this be not disastrous. There is still a place for these units in our repertoire, however, and good design procedures for them should be in our literature.

Design Elements

It is well known that the motional impedance of a contoured resonator has a minimum possible value [1] which is a function of only thickness and contour. Expressions which allow the calculation of the motional impedance have been derived by Sauerbrey [2] and by Tiersten and his co-workers [3].

It is well known that the effective frequency constant of the quartz is a function of the curvature of the contoured face. Sauerbrey found a simple expression for this. The one-dimensional plano-plano frequency-thickness relation is written as:

$$Ft = nK_t \tag{1}$$

Sauerbrey's expression for the plano-convex resonator with contour radius R is:

$$Ft = K_t(n + a\sqrt{t/R}) \tag{2}$$

His derivation predicted a value of 0.51 for a; his experiments found 0.49. This expression is a convenient design formula. It is equivalent to identifying the effective frequency constant as:

$$K_{eff} = K_t(1 + 0.49\sqrt{t/R}/n) \tag{3}$$

It is well known that the temperature-frequency (T-F) characteristic of AT-cut quartz is described by a cubic equation the slope at the inflection point of which is a function of the angle of rotation of the cut around the X-axis. At the AT-cut angle, -35.25 degrees, the best value for the rate of change of this property with angle is given by Ballato [4] to be 5.08×10^{-4} per degree. The slope of the T-F characteristic at its inflection temperature is controlled in production by controlling the rotation angle.

It is well known that a convex contour on the resonator surface has the effect of decreasing the slope (steepening the negative slope) of the T-F characteristic at the inflection temperature. A curve incorporating this effect was generated by Tyler [5]. His curve plotted the "optimum angle" of a resonator at series resonance as a function of thickness-radius ratio. At that time the term optimum angle referred to that angle yielding the minimum total deviation of frequency across the temperature range -55 to +90 Degrees Celsius, the angle which results in the same frequency at -55 degrees as at the upper turn. This curve is included as Figure 9. Worked backwards from the curve, his

CH2965-2/91/0000-184 $1.00 © 1991 IEEE

contoured resonators had the deviations of a plano-plano resonator at -35 degrees, 17 minutes. The allowance for curvature, the difference of angle in minutes between the values of his curve and 17 minutes for plano-convex crystals, is well fitted by the equation [6]:

$$\Delta\theta = 220t/R + 8.5(1-\exp(-111t/R)) \qquad (4)$$

It is well known that the inflection temperature of the curve of the T-F characteristic of the plano-plano resonator is a function of the angle of rotation. This is illustrated by the range of zero crossings of the universal curves in the Union Thermoelectric Handbook [7]. It is displayed in expressions derived from Bechmann in the variable value of their coefficient "B" [8]. For plano-plano AT-cut resonators the rate of change of inflection temperature with angle is -14.7 degrees C per degree θ [4], practically one degree Celsius per four minutes of angle.

It is not well known, however, if the inflection temperature itself is affected by contour. A variation of inflection temperature with contour could be hidden in Tyler's curve. The temperatures of the turns and the temperature differences between any fixed temperature and the turns all depend upon the inflection temperature. To calculate the design of a resonator to make its turn temperature match an oven, the true value of the inflection temperature should appear in some way in the calculation.

Taking the logarithmic derivative of frequency with respect to temperature in equations (1) and (2) yields expressions (5) and (6), respectively:

$$T'(F) = T'(K_t) - T'(t) \qquad (5)$$

$$T'(F) = T'(K_t) - T'(t) + T'(n + a\sqrt{t/R}) \qquad (6)$$

These appear identical except for that one term in (6) containing the square root of the ratio of thickness to radius of curvature. If quartz were isotropic this term would be identically zero. The thermal coefficient of expansion of quartz is greater along Z- than perpendicular to it, so the contour is truly spherical at only one temperature; the angular orientation of the plano surface of the plano-convex blank changes with temperature; the thickness, which is a little piece of the radius of the contouring sphere perpendicular to the plano surface, deviates from that line at any other temperature. One could conclude from (6), in the absence of an adequate theoretical analysis, that any difference between the inflection temperatures of plano plano and contoured resonators must be quite small.

The cubic characteristic is the characteristic of K_t itself. Observable differences between plano-plano and contour must follow from a subtle dependence of the frequency constant upon the contour. It took great sophistication to compute a dependence of the slope at inflection on contour. The same analysis may predict a variation in the inflection temperature itself. We do not have the sophistication to recognize it, but it may well be there.

The Experiment Performed

Undertaking to measure any systematic variation of inflection temperature with contour and to derive an expression for whatever variation we might find, we performed a rather crude experiment. Our data was taken by contouring an ensemble of 60 blanks of a given diameter, thickness and angle, as drawn from stock, using the range of tools maintained in the tool room, three blanks per tool. These were plated, mounted, cased and measured for T-F characteristic. Everything was done under normal factory conditions and by factory personnel.

Blanks were chosen at a nominal -34 degrees, 49.5 ± 0.5 minutes. They were rounded to 550 mils diameter and lapped to 2305 ± 5 khz. Contouring was done manually on one side of the blanks until the central flat spot (the original surface) just disappeared. They were etched to raise the frequency by a nominal 5 khz and plated with 1000 Angstroms of silver. They were mounted and cemented and sealed in resistance welded envelopes.

In the listing to follow the contours used are stated in tool diopters, followed in parentheses by the contour radius in millimeters. These are nominal values. The tools were drawn from the tool room and had varying degrees of wear. Typically a careless operator will so use the tools that tool wear results in a contour somewhat greater than nominal, while the most conscientious operator will control the tool wear so that the contour becomes somewhat less than nominal. It is practically impossible to have a tool wear by a noticeable amount without having one or the other error develop. The tools used were: 1 dp (530 mm), 2 dp (265 mm), 2.25 dp (235.6 mm), 3 dp (176.7 mm), 4 dp (132.5 mm), 5 dp (106 mm), 5.25 dp (100.9 mm), 6 dp (83.3 mm), 6.5 dp (81.5 mm), 7 dp (75.7 mm),8 dp (66.25 mm), 9 dp (58.9 mm), 9.5 dp (55.8 mm), 9.75 dp (54.4 mm), 10 dp (53 mm), 10.5 dp (50.5 mm), 11 dp (48.2 mm), 12 dp (44.2 mm), 12.25 dp (43.3 mm), 13 dp (40.8 mm),

The T-F characteristics of the finished units were tested in a Saunders Model 2100 Passive Test Set. The measurements were fitted to cubic equations by rote using a least squares program. Those curves which displayed a "goodness of fit" of at least 0.9999 were analyzed to determine their inflection temperatures and slopes. From an initial 60 blanks, 34 curves were analyzed. Of the 26 others 19 curves displayed activity dips. Seven units were dead. No unit contoured with curvature below 4 diopters displayed a characteristic sufficiently smooth to be analyzed. Other contours produced crystals with activity dips.

The inflection temperatures and slopes at that temperature of the 34 "good" crystals are displayed variously in Figures 1 through 4. Figure 1 is a histogram of the inflection temperatures measured. Figure 2 plots the measured inflection temperatures, as points vs. contour. Figure 3 plots the measured slopes at inflection similarly. A curve drawn through the ensemble displays the shape of the curve anticipated for this data drawn from Tyler [5],[6]. Figure 4 shows the measured slopes plotted vs. their

corresponding measured inflection temperatures, as points. These should be related, so the plot contains a line illustrating the slope of the relationship anticipated for this plot.

It is immediately apparent that the experiment has not generated a body of data from which a meaningful design procedure can be extracted.

Mostly as an afterthought, we remeasured the units on higher modes. The measurements were much more difficult on the third mode than on the fundamental, and impossible on the fifth. We got usable data on 40 units on the third mode. These are presented in the same sequence as the fundamental properties in Figures 5 through 8.

Discussion

This angle, -34 degrees, 49 minutes, is sufficiently removed from those angles normally required for plano-plano resonators that these resonators should show inflection temperatures significantly displaced from familiar values if contour does not affect the inflection temperature. In particular, since a plano-plano blank around -35 deg., 13 min. has its fundamental series resonant inflection temperature at 26 degrees C, we should expect a blank at -34 deg., 49 min. to have its inflection temperature six degrees lower, or about 20 degrees C. Only a few curves on the fundamental show inflection temperatures below the familiar 26+ degrees, and none are below 23 degrees.

Since the crystals were not intended to perform on higher overtones, the necessary additional fabrication steps to make them so perform were not done. Performance on the third was marginal at best and the test equipment did not even detect a fifth mode response. The measurement on the third mode yielded a preponderance of curves having inflection temperares clustered about 23 degrees as shown in Figure 8, while the fundamental curves are clustered more loosely around 26 degrees (Figure 1).

This clustering of the inflection temperatures may be the significant positive result of the experiment. The clustering on the fundamental about 26 degrees when plano-plano should come in at 20 indicates that the contouring has raised the inflection temperature by aporoximately 6 degrees, about the same amount that the angle difference would have lowered it.

This is is different from the effect of overtone in achieving an approximation of the intrinsic properties of an isolated resonator. The higher the overtone the more closely the characteristic measured at series resonance matches the intrinsic characteristic of quartz itself. The characteristic of quartz unaffected by the properties of any external material or device is difficult to observe. Series resonance is the performance of the resonator in a peculiarly restrictive though technologically useful circuit application. The characteristic behavior at series resonance differs from the intrinsic behavior of the isolated quartz as a consequence of the variation of the electromechanical coupling with temperature [9]. The difference varies inversely with the square of the overtone order. The effect on the third mode is only some 11% of that on the fundamental, the fifth is 4% and the seventh 2%. The difference between the fundamental bunched temperature and the third overtone bunched temperature can be explained as an overtone effect, but both temperatures are displaced from their plano-plano values by some 6 degrees, ostensibly by the contour.

Although there is no clear pattern to be read in Figure 2, we believe we may see a trend in the cluster of points implying a curve of positive slope or a curve concave upward. Figure 6, the corresponding plot of the third overtone data shows an even wider scatter of the points without a clue of a clearer relationship being apparent. Either could be equally used to bolster an argument that no significant effect such as we are seeking exists.

The slope at inflection should become more negative with contour, in the manner indicated by Tyler's curve, Figure 9. A curve inserted into Figure 3 applies Tyler's curve for the dimensions of these crystals to illustrate the effect on the slope at inflection if the only effect is the well-known change of apparent plano-plano angle due to contour. On the fundamental this aspect of the performance of the crystals is approximately normal, the center of gravity of the point scatter tracking the curve with considerable fidelity, but on the third (Figure 7) the data appears to show the slope increasing rather than decreasing. This does not make these assemblies of scattered points more understandable.

Trying to make more sense of the scatter of the points in Figures 2 and 3, the plot of Figure 4 was made, a point plot of the inflection temperature vs. the slope at inflection. If we see a mixture of normal contoured behavior, whatever that is, and the consequences of angle scatter, it might be inferred from some feature of Figure 4. Plano-plano crystals would show no effect of contour, but would demonstrate angle scatter by generating a line with a slope of (14.7/-5.08 =) -2.89 degrees C per ppm per degree C. passing through the point 0 ppm/degree, 26 degrees. An arbitrary line with this slope is drawn into Figure 4. The figure makes it clear that the scatter of points is not a simple consequence of a scatter of angle in the initial blanks. If it were, the data points would fall into a band having that slope, or deviating from it in some systematic manner. If this plot is of a band of points, then that band has a slope somewhat steeper than the line.

In order further to clarify what we are seeing, we remeasured the crystals on the third mode. This data was harder to get and less satisfying than the fundamental data, being derived from only 9 points per curve, vs 17 for the fundamental measurements. The group of 40 crystals which yielded "good" curves were analyzed by the same approaches. Figure 8 is constructed from third mode measurements as was Figure 4, plotting the inflection temperature vs the slope at inflection. Even though the cubic fir was "good", five points were so far from the rest that they are only indicated, not plotted. This, we are satisfied, is an effective way of demonstrating that the third mode data points are not an obviously more well-behaved single population than are the fundamental mode points.

An object lesson and an example of what we had hoped to see developing as a reasonable extension of our project can be seen in a paper by Adams, et al. [10]. Their reported measurements of angle were made to 5 seconds of angle in some cases, and to no worse than 15 seconds in any case. Their crystals were measured on the 5th mode, ours on the fundamental. For this investigation they used a variety of diameter-thickness ratios as necessary for well-behaved crystals. One of their curves, reproduced here as Figure 10, illustrates the quality of results possible to be had in an adequate experiment, even with contoured crystals. These curves were easier to obtain than ours would ever be, as these crystals were all of a single proven design at a single frequency, differing only in angle. The quartz and the laws of acoustics are both completely reliable. The reason for our imperfect results is merely an imperfect or incomplete experiment.

Conclusions

There was only one fabrication run. Several facts became obvious from this limited experiment:

(1) Visual examination of the results is enough to show that the ensemble represents more than a simple population. The hypothesis of off-angle blanks in the ensemble is not adequate to explain the scatter of results. In another experimental cycle, however, the blanks should be preselected into a tight group, only a few seconds wide, before rounding.

(2) The *single* diameter used for all blanks was not suitable for all contours. This really was predictable. No contour below 4 diopters was usable for data on the fundamental mode. These lower contours would have performed better on larger blanks. Moreover, the principal coupled modes generating activity dips depend on thickness, diameter and contour, while the designed mode's behavior depends only on the central geometry, on the thickness and contour. Several contours higher than 4 diopters also generated unuseable curves. Tyler's work on contoured crystals resulted in a relatively small number of prototype designs which were geometrically scaled to cover a range of frequency. In our experiment, with its variety of contours, we needed a selection of diameters to combat problems with both activity and interfering modes.

(3) We selected an unorthodox angle, one normally better suited for double convex crystals than plano-convex. None of our curves turned; none had a negative slope at inflection. This may actually have been beneficial, as it forced us to depend upon the data itself without being led astray by a subjective feel for the shape of familiar curves, but the wide range of frequency in each measured characteristic made the test set behavior marginal. The test set operated according to a program which did not find the inflection point in any of these curves. All cubic curves have inflection points. Even if it has no point of zero or negative slope, each has a point of zero curvature. in another experiment we would use a rotation some ten minutes greater.

(4) We must accept the possibility that the simple problem we posed for ourselves has no simple solution. Plano-plano crystals are well designed by simple design equations because their behavior is so well described by one-dimensional resonator theory. Contoured crystals are three-dimensional resonators. We simplify the discussion of contoured crystals by using a one-dimensional vocabulary, but speak of "edge effects". These edge effects may well be the source of our data scatter, not reducible in such a simple manner as we tried. Moreover, the effect we tried to measure may not exist, at least as a first order effect.

(5) The experiment might be described more accurately as incomplete rather than as imperfect. Under other circumstances each of the authors would have considered this a good start, and work would have continued.

The results are not as orderly as we hoped. If contour has an effect upon the inflection temperare of the T-F characteristic of contoured resonators, it is certainly small. From this data we can say it is at most a small increase with contour. It can be argued with equal justification that there is no effect, though this would provide no explanation for the distribution of inflection temperatures. The quality of the results is not sufficient for this to be resolved. We would make a more complete project to be reported at a later date if we still had ready access to the necessary facilities.

Appendix 1, The Plotted Points, Fundamental

No.	Dp	Radius	A, ppm/deg	T, deg C
13	4	132.5	1.24906	25.0044
14	4	132.5	1.03388	26.4740
15	4	132.5	1.23744	25.6981
17	5	106	1.06385	25.6022
19	5.25	100.9	1.09468	26.1188
20	5.25	100.9	1.24663	23.6981
21	5.25	100.9	1.16690	25.3089
22	6	83.3	1.05721	25.6225
23	6	83.3	1.17977	25.1056
24	6	83.3	1.01498	25.9747
25	6.5	81.5	1.26957	25.9329
26	6.5	81.5	1.06245	24.9960
27	6.5	81.5	1.21246	23.8020
28	7	75.7	1.00227	28.8675
29	7	75.7	1.12881	27.8176
30	7	75.7	1.13955	27.1753
37	9.5	55.8	1.01055	23.8372
38	9.5	55.8	0.817492	24.3963
40	9.75	54.4	0.790333	24.9076
41	9.75	54.4	0.961604	24.0407
42	9.75	54.4	0.906340	25.6061
46	10.5	50.5	0.997322	26.2465
47	10.5	50.5	0.939358	26.1267
48	10.5	50.5	0.832041	26.5952
49	11	48.2	1.01291	26.4265
51	11	48.2	0.657927	28.1896
52	12	44.2	0.736414	26.6866
54	12	44.2	0.731986	26.1253
55	12.25	43.3	0.842183	27.4259
56	12.25	43.3	0.721713	27.2070
57	12.25	43.3	0.837540	27.9600
58	13	40.8	0.680581	28.0475
59	13	40.8	0.825896	28.1697
60	13	40.8	0.805090	28.2479

Appendix 2. The Plotted Points, Third Mode

No.	Dp	Radius	A_1 ppm/deg	T_i deg C
1	1	530	2.42252	21.2846
7	2.25	235.6	2.48272	21.0132
8	2.25	235.6	2.33709	21.4103
9	2.25	235.6	2.33989	21.7462
10	3	176.7	2.43891	21.2808
11	3	176.7	2.39896	23.0986
12	3	176.7	2.48064	22.4736
17	5	106	2.44540	21.0739
18	5	106	2.63406	20.1839
19	5.25	101	2.41797	24.2434
20	5.25	101	2.60724	21.9196
21	5.25	101	2.46082	24.4372
22	6	88.3	2.49781	23.0816
23	6	88.3	2.57279	18.0658
24	6	88.3	2.47407	24.1556
25	6.5	81.5	2.72146	21.2039
26	6.5	81.5	2.57726	21.9511
27	6.5	81.5	2.70420	22.3773
28	7	75.7	2.53386	22.8711
29	7	75.7	2.50115	22.6261
30	7	75.7	2.69831	21.0617
31*	8	66.25	2.57521	28.7580
33	8	66.25	2.41123	24.5255
35*	9	58.9	2.56037	26.5438
37	9.5	55.8	2.69296	18.9568
38	9.5	55.8	2.53581	19.9563
39	9.5	55.8	2.61678	20.8148
40	9.75	54.4	2.47636	24.3505
41	9.75	54.4	2.65434	22.5330
42	9.75	54.4	2.62007	22.9062
43	10	53	2.51041	19.5382
44	10	53	2.65471	21.1663
45	10	53	2.51943	17.9006
47	10.5	50.5	2.66616	22.9142
48	10.5	50.5	2.61413	22.9752
52	12	44.2	2.56598	24.5728
53*	12	44.2	2.64644	28.2809
54*	12	44.2	2.56439	31.5934
56	12.25	43.3	2.69146	21.5026
60*	13	40.8	2.67057	27.1096

Note: * signifies data points not on plots.

Appendix 3, Inflection Temperature

It was noted above that the program of the commercial test equipment used failed to recognize the inflection temperatures of the T-F characteristics of these crystals which have no region of negative slope. If a cubic curve has two extrema, that is, if it has an upper and a lower temperature turn, the inflection point can be shown to lie mid-way between them. This is not, however, the definition of an inflection point. An inflection point is a point of zero curvature. In analyzing a cubic function for its interesting characteristics, it will be shown that every cubic curve has one inflection point, and if it has one extremum it has two.

If:

$$F = A_0 + A_1T + A_2T^2 + A_3T^3$$

then:

$$F' = A_1 + 2A_2T + 3A_3T^2$$

$$F'' = 2A_2 + 6A_3T$$

At inflection $F'' = 0$, from which it follows that:

$$T_i = -A_2/3A_3$$

The slope at inflection is F' evaluated at T_i, which results in:

$$F'_i = A_1 - A_2^2/3A_3$$

If the cubic has an extremum it occurs where $F' = 0$, at which point(s):

$$T_e = \frac{-A_2 \pm \sqrt{A_2^2 - 3A_2A_1}}{3A_3}$$

If $A_2^2 > 3A_1A_3$ then the expression under the radical is positive and the equation has two roots disposed equally above and below the inflection temperature. If $A_2^2 = 3A_1A_3$ there is only one solution and the slope of the characteristic is zero at the inflection temperature. If $A_2^2 < 3A_1A_3$ there is no point of zero slope. The extrema are imaginary, but there is still one real inflection temperature as derived above.

References

[1] W, G. Stoddard, "Design Equations For Plano-Convex AT-Filter Crystals",Proc. 17th AFCS, 1963, pp 272-282.

[2] G. Sauerbrey and G. Jung, "Schwingungsmoden Planconvexer Quartzplatten",Z. Angew. Phys. 24 bd., heft 2, 1968, pp 100-108.

[3] H. F.Tiersten and R. C. Smythe, "An Analysis of Overtone Modes in Contoured Crystal Resonators", Proc. 31st AFCS, 1977, PP 44-47.

[4] A. Ballato personal communication, 1979.

[5] L. Tyler, "Design of Low Frequency AT-Cut Quartz Resonators",Proc. 14th AFCS, 1960, pp 179-191.

[6] J. H. Sherman, Jr., "Design Routine for Contoured AT-cut Resonators", Proc 5th Quartz Crystal Conference, 1983, pp 131-138.

[7] Union Thermoelectric Division, Quartz Resonator Handbook, Manufacturing Guide for "AT"-Type Units", 1960, AD# PB-171839, pp 90-93.

[8] See ibid, pp 85 and references cited there.

[9] J. H. Sherman, Jr., "On the Orientation of Quartz Resonator Plates", IEEE Trans. UFFC, vol. 33, no. 1, 1986, pp 74-76, or "A Note on Ballato's Angle Increment, Proc 39th AFCS, 1985, pp 427-430.

[10] C. A. Adams, G. M. Enslow, J. A. Kusters and R. W. Ward,"Selected Topics in Quartz Crystal Research", Proc. 24th AFCS, 1970, pp 55-63.

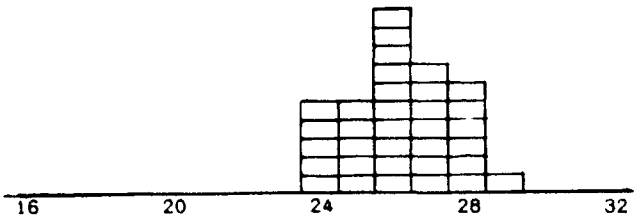


Figure 1. Distribution of 34 Fundamental Inflection Temperatures.

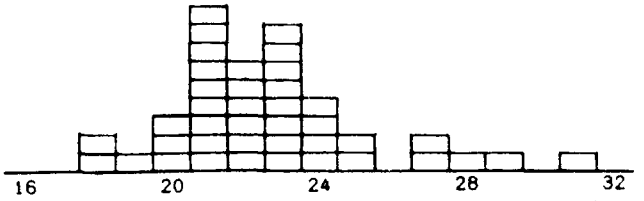


Figure 5. Distribution of 40 Third Mode Inflection Temperatures.

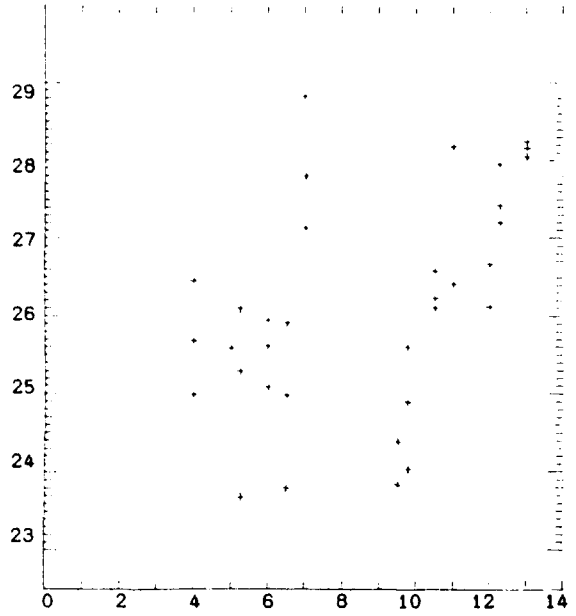


Figure 2. Fundamental Inflection Temperature (deg C) vs. Contour (diopters).

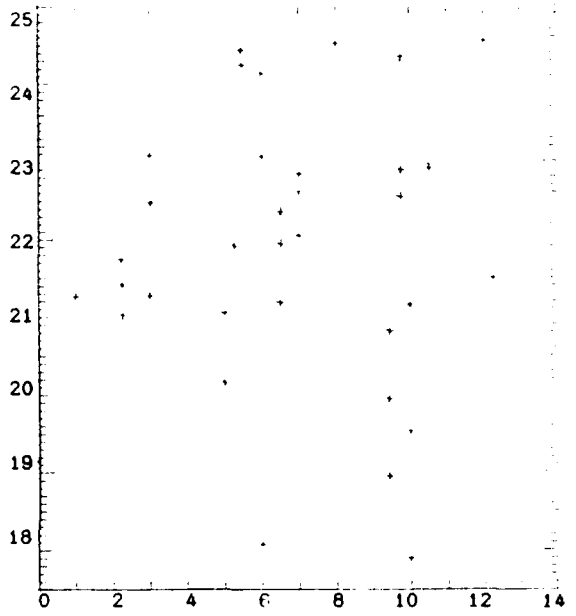


Figure 6. Third Mode Inflection Temperature (deg C) Contour (diopters).

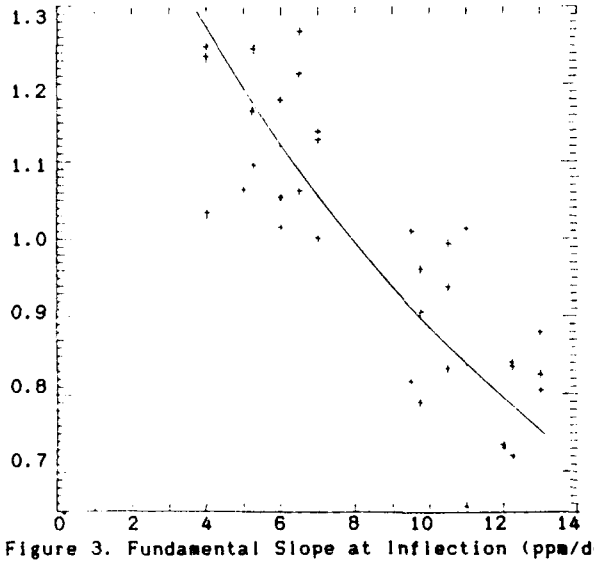


Figure 3. Fundamental Slope at Inflection (ppm/deg) vs. Contour (diopters).

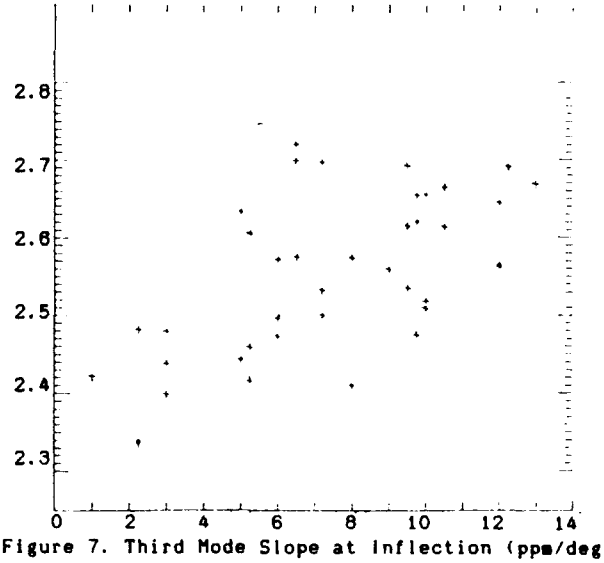


Figure 7. Third Mode Slope at Inflection (ppm/deg) va. Contour (diopters)

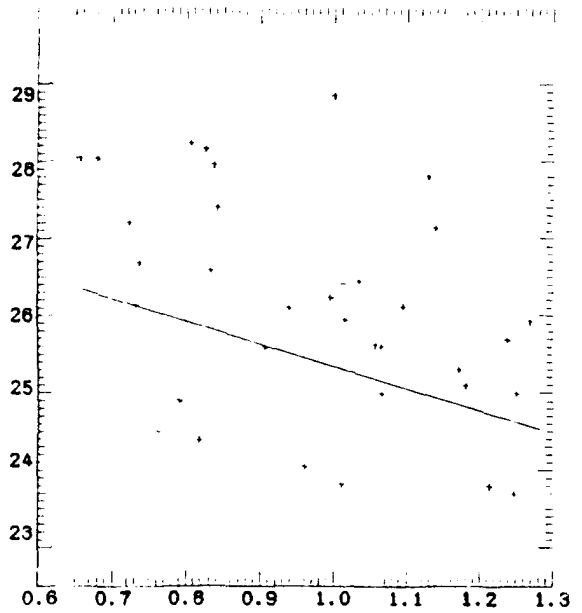


Figure 4. Fundamental Inflection Temperature (deg C) vs. Slope (ppm/deg).

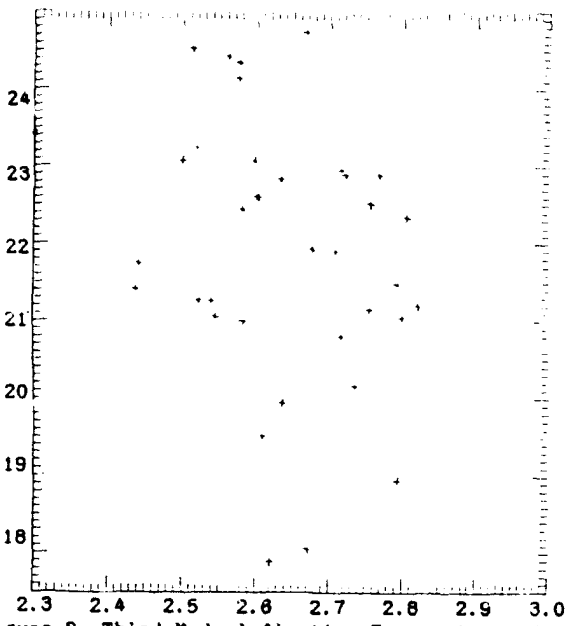


Figure 8. Third Mode Inflection Temperature (deg C) vs. Slope (ppm/deg).

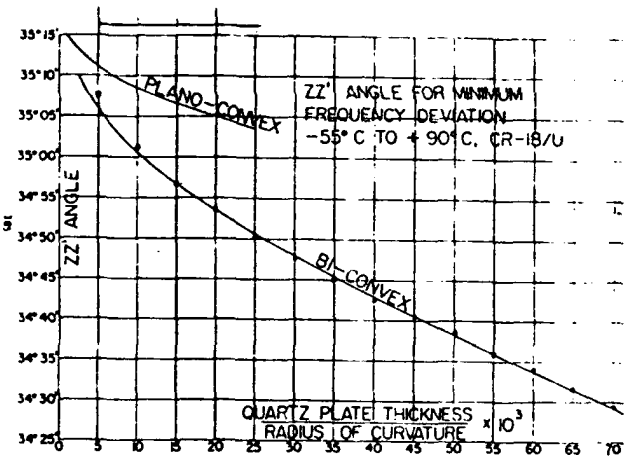


Figure 9. Tyler's Curve of Angle vs. Contour. [5]

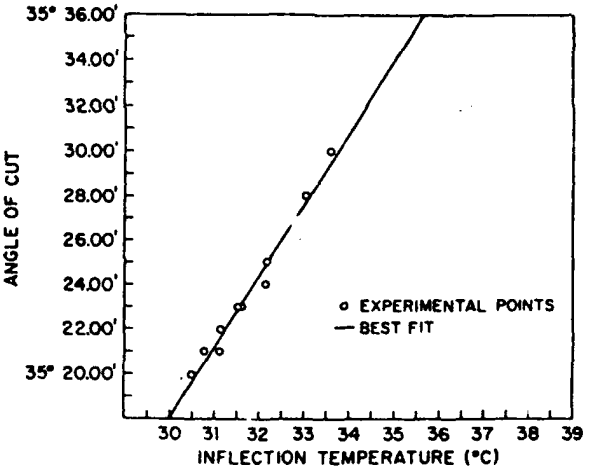


Figure 10. Quality of Results Attainable (AFTER Adams et al. [10])

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

LOAD RESONANT MEASUREMENTS OF QUARTZ CRYSTALS

Dwane Rose

Saunders and Associates, Inc.
7440 East Karen Drive, Scottsdale, Arizona 85260

ABSTRACT

The measurement of load resonant frequency can be accomplished using several measurement techniques. Load resonant frequency measurements with Transmission Systems and Crystal Impedance Meters are examined along with the measurement difficulties of each method. Potential sources of correlation differences are discussed including the impact of the crystal spurious responses on the load resonant measurement.

INTRODUCTION

A quartz crystal is used at either the series resonant frequency, determined by the motional parameters of the crystal, or at a load resonant frequency that is determined by the crystal's motional parameters and the value of an external load reactance. Normally this external reactance is a capacitor, with the resulting load frequency slightly higher than the series resonant frequency.

Load resonant frequency measurements are inherently difficult to make because of the accuracy and repeatability required for the measurement and because the crystal itself impacts the resolution of measurement. In addition, the associated strays of the measurement circuit can have a pronounced impact upon the measurement being made. These difficulties are examined for load resonant frequency measurements made on Transmission Systems and Crystal Impedance Meters. The measurement techniques presented are those typically used in the industry. The techniques described for the Transmission System are also applicable to an S Parameter Measurement System.

EQUIVALENT CRYSTAL CIRCUIT

The basic electrical equivalent circuit for a quartz crystal operated at load resonance is shown in Figure 1.

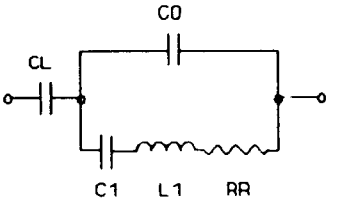


Figure 1

Where: CL = Load Capacitor (pF)
C1 = Motional Capacitance (pF)
L1 = Motional Inductance (mH)
RR = Crystal Resistance (ohms)
C0 = Static Capacitance (pF)

The load resonant frequency can be shown to be [1]:

$$FL = FR\left[\frac{C1}{2(C0 + CL)} + 1\right] \quad (1)$$

Where: FL = Load Resonant Frequency (Hz)
FR = Series Resonant Frequency (Hz)

When a crystal is measured with a physical load capacitor there are additional stray terms that are introduced into the circuit. The most prominent terms are the stray capacitance present at the junction of the crystal and the external load capacitor, the stray capacitance across the crystal, and the lead inductance of the circuit. Figure 2 shows the electrical model including these terms.

CH2965-2/91/0000-191 $1.00 © 1991 IEEE

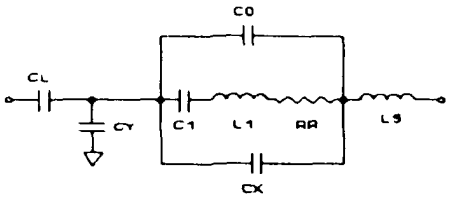


Figure 2

Where: CY = Stray Capacitance (pF)
CX = Stray Capacitance (pF)
LS = Lead Inductance (nH)

The load resonant frequency of this circuit including the circuit stray reactance is shown in Equation (2).

$$FL = FR\left[\frac{C1\,(CL + CY - CLS)}{2\,[\,(CL + CY)(CO + CO) - CLS\,(CO + CL + CX + CY)]} + 1\right] \quad (2)$$

Where: CLS = $(10^{21})/(4\ PI^2\ FR^2\ LS)$

The measurement of load resonant frequency and the impact of these strays for both Transmission Systems and Crystal Impedance Meters are discussed in the following paragraphs.

TRANSMISSION MEASUREMENT SYSTEMS

The IEC-444 [2] Standard defines a method of measuring the series resonant frequency and motional parameters of a quartz crystal using a frequency synthesizer and a vector voltmeter. A block diagram of this system is shown in Figure 3.

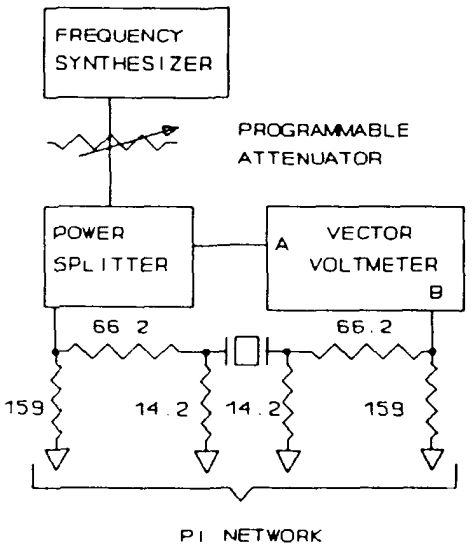


Figure 3

The Transmission System first measures the static capacitance of the crystal at a frequency that is far removed from the resonant frequency of the crystal. The frequency synthesizer is then stepped in the vicinity of resonance and impedance measurements are made. The reactance of the static capacitance is subtracted from these measurements which results in the motional arm reactance and resistance. The measurement iteration continues until the motional arm reactance equals zero. Series resonant frequency and resistance are then determined at this point. The motional capacitance is derived from the slope of the motional arm reactance as indicated by Equation (3).

$$C1 = \frac{10^{12}}{\pi\ XS\ FR^2} \quad (3)$$

Where: XS = Reactance slope (Ohms / Hz)

Load resonant frequency can be determined with a Transmission System using three distinct methods of measurements: "Calculated", "Measured", and "Physical Capacitor".

Method # 1 - "Calculated" Load Frequency

In this method of load resonant frequency measurement, the load frequency is directly calculated using Equation (1). The standard measurement configuration shown in Figure 3 is used to make the static capacitance, series resonant frequency, and motional capacitance measurements required to calculate the load resonant frequency. The series resonant frequency and motional parameters are determined from measurements made in the vicinity of series resonance.

The circuit stray reactances indicated in Figure 2 have minimal impact upon the measurement. The CY stray term is shunted out by the 14.2 ohm terminating resistor of the PI Network. The CX and LS terms are determined during system calibration and are mathematical removed from the crystal measurements.

The resolution of the "Calculated" load frequency measurement depends upon the crystal being measured. Typical resolutions are from 0.5 ppm to 3 ppm. Figure 4 shows the distribution printout of 500 measurements of a single crystal measured on an S&A 350A System. The frequency spread for this example was 2.3 ppm with a standard deviation of 0.45 ppm.

The "Calculated" technique is the simplest method of accomplishing load resonant frequency measurements with a Transmission System. The repeatability and accuracy of the series resonant frequency and motional parameters are not affected by this measurement method. Typical series resonant frequency resolution for an S&A 350A System is 0.1 ppm to 0.2 ppm.

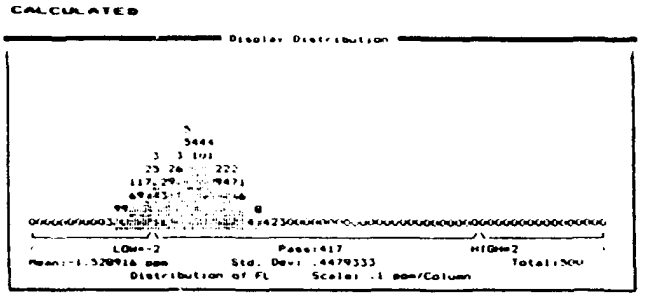


Figure 4

Another advantage of the "Calculated" method is that no additional measurements are required to determine load resonant frequency. The total measurement time required for the crystal measurement is not affected by this method.

The "Calculated" method of measurement can result in an error at load frequency if the crystal does not follow the frequency relationship indicated by Equation (1). The accuracy of the load frequency measurement is also affected by the resolution and accuracy of the static and motional capacitance measurements.

The "Calculated" method of measurement can be used on those crystals that are known to follow the relationship indicated by Equation (1) and do not require the measurement resolution offered by other measurement techniques.

Method #2 - "Measured" Load Frequency

This measurement method first measures the series resonant frequency and motional parameters in the same manner as in the "Calculated" method. Using that information the system then calculates the load frequency, sets the synthesizer to that frequency, and makes an impedance measurement. The displayed load frequency is then adjusted linearly based upon the difference between the calculated impedance and the measured impedance.

The "Measured" load frequency technique uses the standard PI Network of Figure 3. The impact of the stray terms shown in Figure 2 is the same as described for the "Calculated" method: The CY stray capacitance is shunted by the 14.2 ohm PI Network terminating resistor and the CX and LS terms are removed mathematically.

The resolution of the load frequency measurement is a function of the impedance measured at the calculated load frequency. As the impedance becomes large (greater than 1000 ohms) the measurement resolution decreases. Equation (4) shows an approximate relationship of impedance magnitude and load capacitance.

$$Z = \frac{1}{2\,\Pi\,FL\,CL} \qquad (4)$$

Where: Z = Impedance magnitude (Ohms)

Figure 5 shows the distribution of 500 repeated measurements of the same crystal used for the measurements of Figure 4. The frequency spread for this crystal was 0.6 ppm with a standard deviation of 0.1 ppm. The magnitude of the impedance of this measurement was approximately 625 ohms.

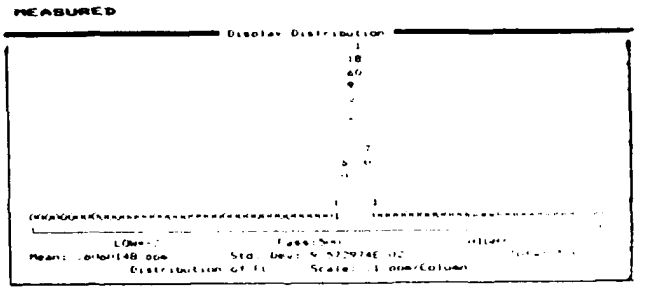


Figure 5

A significant advantage of the "Measure" method is that nonlinear crystal responses that would invalidate the frequency relationship defined by Equation (1) are included in the measurement.

This measurement method requires an additional measurement at the load resonant frequency. For an S&A 350A system this extends the total measurement time per crystal from a typical 1.2 seconds to 1.4 seconds.

The magnitude of the impedance defined in Equation (4) limits the maximum current that the system can provide to the crystal. The crystal is a current sensitive device and should be tested in the vicinity of the specified operating power. Typically magnitudes of impedance less than 1000 ohms can be measured using this technique.

Method # 3 - "Physical Capacitor"

An alternate method of measuring load frequency is to place a physical capacitor in series with the crystal. The crystal and load capacitor combination present the measurement system with a "new" crystal with a different resonant frequency and different motional parameters. This "new" crystal is measured in the same manner as described for the "Calculated" method. The frequency measured is the load frequency of the crystal at that value of load capacitance. The motional parameters measured correspond to the "new" crystal and must be mathematically corrected to obtain the correct parameters of the crystal.

When a physical capacitor is added to the PI Network, the junction between the crystal and the load capacitor becomes a point of high impedance. The stray capacitance at this junction, indicated in Figure 2 as CY, now has a significant impact upon the measurements. The total load capacitance in the circuit now includes the load capacitor, the stray to ground (CY), and the pin to pin capacitance (CX). The impact of these stray terms is shown by the following equations.

$$C0 = \frac{(C0m + CX)(CL + CY)}{(CL - C0m - CX)} - CX \qquad (5)$$

$$C1 = \frac{C1m\,(C0 + CX + CY + CL)^2}{CL\,(CL + CY)} \qquad (6)$$

$$RR = \frac{Rm\,CL\,(CL + CY)}{(C0 + CX + CL + CY)^2} - \frac{10^9}{2\,\Pi\,FL\,CL} \qquad (7)$$

Where: COm = Measured C0 (pF)
C1m = Measured C1 (pF)
Rm = Measured RR (Ohms)

These equations assume that the crystal cover is not grounded. If the cover is grounded, an unknown capacitance from the crystal electrode to the cover is in parallel with the CY stray term. Since this capacitance is indeterminate, the equations above are not applicable.

The CY term depends upon the physical layout of the test socket and is included as a constant in the measurement software. The CX and LS stray terms are determined during system calibration.

The "Physical Capacitor" technique results in very repeatable load frequency measurements. Typical resolution of load frequency measurements is from 0.1 ppm to 0.2 ppm. Figure 6 below shows the same crystal of Figure 4 measured 500 times using the "Physical Capacitor" method. The frequency spread indicated is 0.2 ppm with a standard deviation of 0.03 ppm.

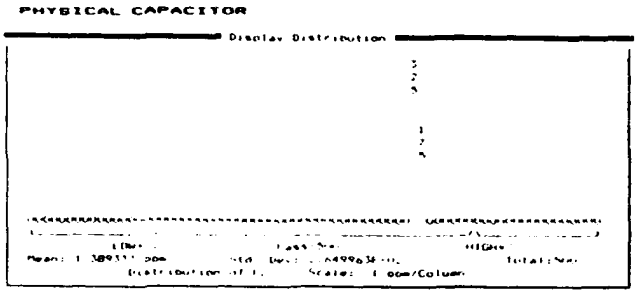


Figure 6

The "Physical Capacitor" method typically can provide the specified drive current into the crystal. The crystal is in resonance with the physical capacitor and the resulting combination presents a low impedance to the PI Network. Non-linear crystal responses at the load resonant frequency are also directly reflected in the measurement.

The measured parameters are not as accurate with this method as those measured without a physical capacitor because of the required mathematical transformations. In addition any variation in the stray CY term will result in a measurement error. The "Physical Capacitor" also requires the standard PI Network to be modified to accept a load capacitor in series with the crystal.

The series resonant frequency and motional parameters could be measured with the same repeatability of the "Calculated" method by replacing the load capacitor with a short. This can effectively be done in a 350A System by installing an additional series resonant frequency test port. The disadvantage of this technique is that the frequency measurements are made twice: once at load resonance and once at series resonance. The typical measurement time of an S&A 350A system using the "Physical Capacitor" method and an additional series resonant measurement port is extended from 1.2 seconds to 2.0 seconds.

The "Physical Capacitor" method should be used when the greatest repeatability of load resonance measurements is required. This method is also appropriate for those crystals that require a higher drive current than the "Measured" method can provide.

Each of the measurement methods of load resonant frequency for a Transmission System has unique advantages and disadvantages. The best method for a particular crystal depends upon the specific testing requirements.

Temperature Testing at Load Resonant Frequency

The impact of temperature on a crystals's frequency and resistance can be examined using the measurement techniques and equipment described previously along with a precision temperature test chamber. An S&A 2200 Temperature Test System using the S&A 350A Transmission System and an S&A 4220MR Temperature Chamber was used to measure an example crystal. A crystal was selected that does not follow the frequency relationship of Equation (1). The measurements made on this crystal illustrate the differences between the Transmission System methods of measurement. This crystal was specially selected for this example and represents a possible measurement condition but not a typical one.

A total of four measurement runs were made. The first run was at series resonant frequency, the remaining three runs were performed at load resonant frequency using the "Calculated", "Measured", and "Physical Capacitor" methods of measurement. The crystal was measured over a temperature range of 0 °C to 60 °C.

Figure 7 is the output from the measurement run made at series resonant frequency. The graph shows frequency and resistance information as a function of

temperature. The frequency data is first curve fit to a third (or fourth) order polynomial function. This curve is then plotted along with the deviation of the measured frequency from the calculated curve. It is presented in this manner so that any nonlinear characteristics in the crystal's response can be easily seen.

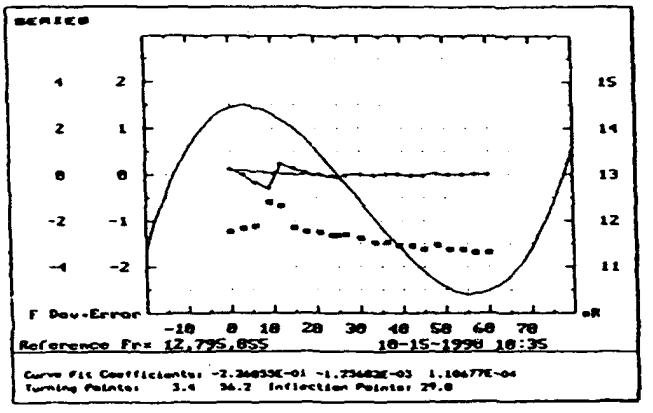


Figure 7

A nonlinear response or perturbation in frequency and resistance at 10 °C is evident in Figure 7. This perturbation is due to a coupled mode of vibration which invalidates the frequency relationship of Equation (1) .

Figure 8 shows measurements of the same crystal at a load capacitance of 18 pF using the "Calculated" method of measurement.

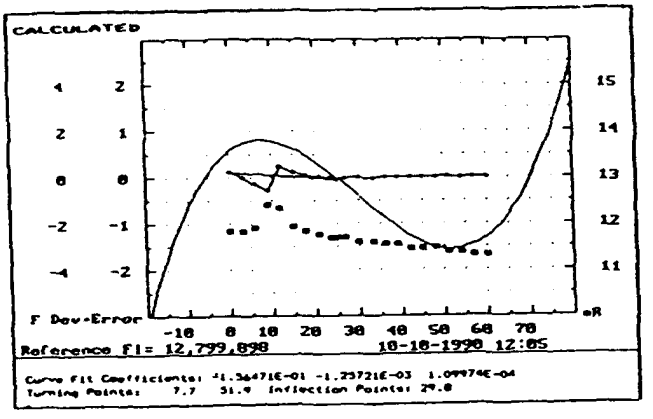


Figure 8

The coupled mode is still present in the data because the measurements made using the "Calculated" method are at series resonance. This coupled mode, however, does not exist at the load resonant frequency specified. The "Calculated" method results in the perturbation at series resonance being displayed inappropriately at the load resonant frequency.

The plot of the crystal using the "Measured" method of measurement is shown in Figure 9.

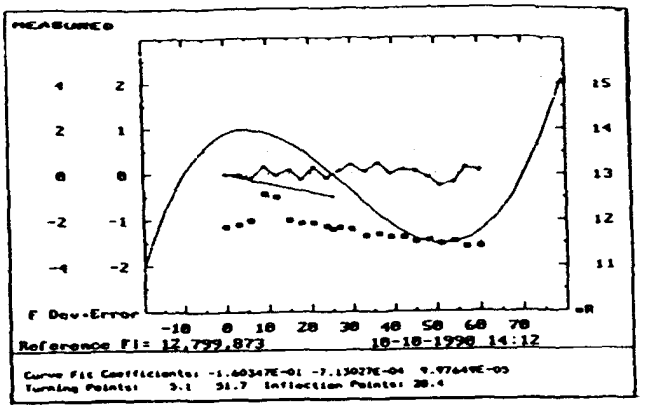


Figure 9

The coupled mode is no longer present in the frequency information because the frequency was measured at the load resonant frequency of the crystal. The motional parameters are still measured at series resonant frequency as indicated by the presence of the resistance perturbation.

Figure 10 shows the measurements made using the "Physical Capacitor" method of measurement.

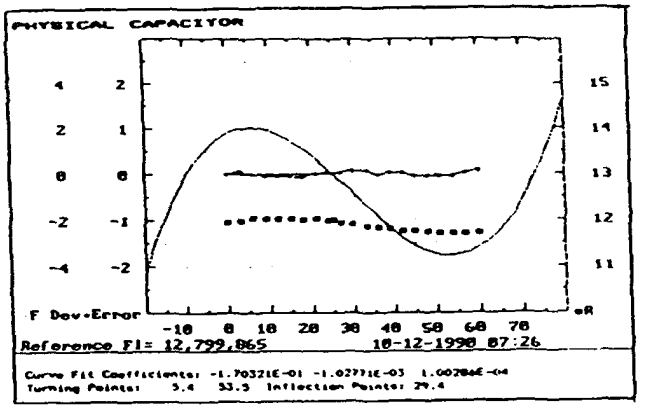


Figure 10

Since the frequency and motional parameter measurements are made at load resonant frequency, the coupled mode at series resonant frequency has no effect on the displayed data.

As these figures indicate, the presence of a nonlinear crystal response at the load resonant frequency can be detected using the "Measured" or "Physical Capacitor" method of measurement. These two methods actually measure the crystal at the specified load resonant frequency.

In comparing Figures 7, 8, 9, and 10 a difference in the first order coefficient of the curve fit between the

series resonant measurements and the load resonant measurements is evident. This difference is due to the temperature coefficient of the piezoelectric coupling factor. We have observed that this is essentially linear and can be expressed as indicated in Equation (8).

$$K = (AO(L) - AO(S))\left[\frac{FR}{FL - FR}\right] \quad (8)$$

Where: K = Piezoelectric Coupling Temperature Coefficient
AO(L) = First Order Curve Fit Coefficient at Load Resonant Frequency
AO(S) = First Order Curve Fit Coefficient at Series Resonant Frequency

The "Calculated" method shown in Figure 8 reflects a fixed piezoelectric coupling temperature coefficient used by the S&A 2200 system of 220 ppm/°C.

The "Measured" method of measurement (Figure 9) resulted in a temperature coefficient of 212 ppm/°C. This value is technically the most correct for this crystal because the stray capacitances of the test circuit do not impact the measurements.

The "Physical Capacitor" measurements shown in Figure 10 indicate a temperature coefficient of 180 ppm/°C. This value is affected by the temperature coefficient of the physical capacitor and any changes in the test circuit strays as a function of temperature.

These small differences in the temperature coefficient of the piezoelectric coupling factor are reflected in the graphs of the third order frequency curvefit. The worst case measurement error due to these differences is 1 ppm at 55 °C.

A coupled mode as indicated in Figure 7 has a negative coefficient that is typically in the range of -5 ppm/°C to -25 ppm/°C. As shown in the previous paragraphs, the presence of a coupled mode can be missed or incorrectly indicated if the test is not performed at the specified load resonant frequency. The "Measured" and "Physical Capacitor" measurement methods provide the capability of detecting these coupled modes.

The "Physical Capacitor" method can be used even if the physical capacitor is not precisely the same value as the specified load capacitor. In this case the temperature test range can be extended so that any coupled modes would be apparent. For example, a crystal that is tested at a frequency 50 ppm lower than the specified load frequency should be tested at a maximum temperature of at least 10 °C (-50/-5) greater than that specified. A crystal that is tested at a frequency 50 ppm more than the specified load frequency should be tested at a minimum temperature that is 10 °C less than that specified [4].

CRYSTAL IMPEDANCE METERS

A Crystal Impedance (CI) Meter is specifically designed for measuring the electrical parameters of a quartz crystal. These instruments contain a tuned oscillator circuit, frequency counter, and a capacitance meter. A block diagram of a CI Meter is shown in Figure 11.

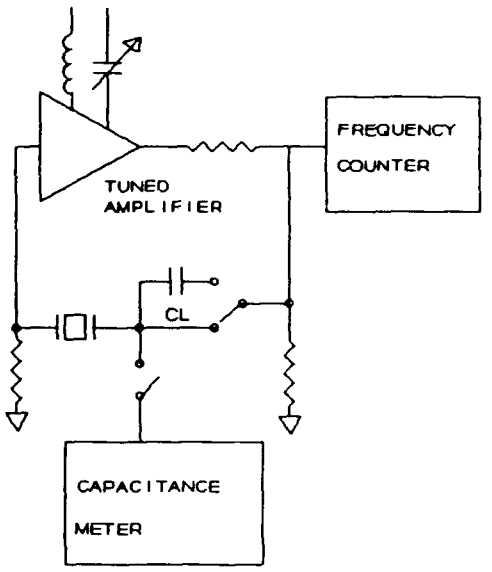


Figure 11

In order to measure a crystal, the tuned circuit is setup near the frequency of interest. The crystal is then inserted into the test socket. The CI Meter first measures the static capacitance, resistance, and series resonant frequency of the crystal. A known load capacitance is then switched in the circuit and a load resonant frequency is measured. The motional capacitance is then calculated using Equation (1). If the specified load capacitance is different than the known capacitance, the load resonant frequency is calculated using the measured motional capacitance and Equation (1).

The effects of the stray lead inductance (LS), shown in Figure 2, is minimized by tuning the oscillator's tank circuit with a resistor of approximately the same resistance as the crystal. This results in approximately zero phase shift across the test socket. If the value of the stray lead inductance is known, the frequency measurements can also be mathematically corrected using Equation (2). The stray capacitance terms (CY, CX) are included in the measurement of the known load capacitor.

A crystal was measured 500 times using an S&A 150D CI Meter. Figure 12 shows the distribution of these measurements. From this information the frequency spread for this example was 0.6 ppm with a standard deviation of 0.1 ppm.

A typical Crystal Impedance Meter has an upper frequency limit of 60 Mhz. This limit is usually inconsequential for load resonant measurements, but can be a significant performance limit for high frequency overtone crystals used at series resonance.

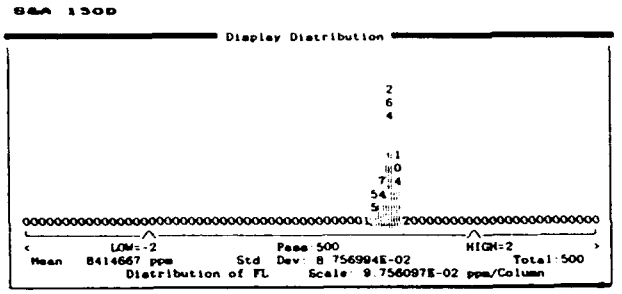


Figure 12

The significant advantage of Crystal Impedance Meters is their cost. Several Crystal Impedance Meters can be purchased for the price of a single Transmission System.

Temperature Testing at Load Resonant Frequency

An S&A 2150 Temperature Test System was used to measure the frequency and resistance deviation of an example crystal. The S&A 2150 system uses an S&A 150D Crystal Impedance Meter and an S&A 4220MR temperature test chamber. Figure 13 shows the measurements of the example crystal over a temperature range of 0 °C to 60 °C.

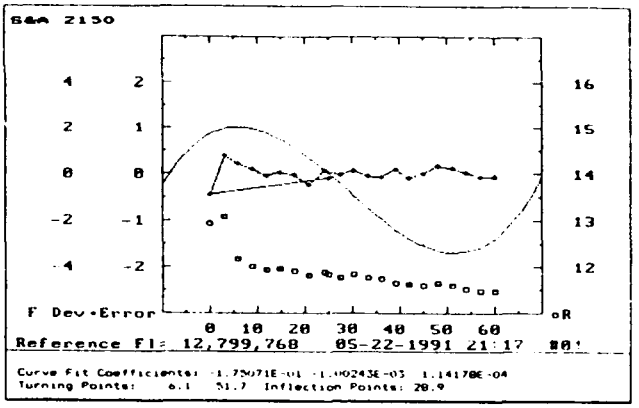


Figure 13

The S&A 2150 uses a fixed piezoelectric coupling factor temperature coefficient of 220 ppm/°C. This is the same correction as described for the "Calculated" Transmission System measurement method.

The S&A 2150 system measures the motional parameters of the crystal at the first temperature point. Subsequent measurements are made at series resonance and then converted to load resonance frequency using Equation (1). This technique can result in an error in the displayed data if a coupled mode is present. A coupled mode is not common in a crystal but should be considered if there are significant correlation errors between the S&A 2150 System and other methods of measurement.

LOAD RESONANT FREQUENCY MEASUREMENT CORRELATION

The ability of a measurement system to determine the load resonant frequency is a function of the crystal and the specified load capacitance. For example, a 12 MHz crystal with a frequency specification of +/- 10 ppm is more difficult to measure at a load capacitance of 18 pF than at 30 pF. A better method of specification is to specify the frequency as absolute and allow a tolerance on the load capacitance. The specification for the 12 MHz part could be given as 12 MHz at 18 pF +/- 0.5 pF. The specification on load capacitance can be converted to a frequency tolerance by multiplying by the trim sensitivity of the crystal at that value of load capacitance. The trim sensitivity is defined as derivative of the load frequency with respect to load capacitance. Equation (9) can be used to calculate this value from the other crystal parameters.

$$TS = \frac{500000\ C1}{(CO + CL)^2} \qquad (9)$$

Where: TS = Trim Sensitivity (ppm/pF)

For the 12 MHz crystal example, the trim sensitivity was 20 ppm/pF. This means the +/- 0.5 pF specification on the load capacitance would be mathematically equivalent to tolerance of +/- 10 ppm on the load frequency. However, it is easier to explain to a crystal user how small 0.5 pF is, than to explain a frequency variation of 10 ppm.

Ten crystals of various load resonant frequencies were measured using two different Crystal Impedance Meters and the three methods of Transmission System measurement. The temperature of the crystals was controlled during these measurements at 25.0 °C with an S&A 4220MR temperature chamber.

The first CI Meter used to make these measurements was an S&A 150C. The S&A 150C requires a setup resistor for each crystal. A variable load capacitor must also be adjusted to the value required for the crystal to be tested. The load resonant frequency is determined by switching in the load capacitor and reading the resulting frequency.

The second CI Meter used for these measurements was an S&A 150D. This CI Meter uses a fixed internal resistor to setup the tuned circuit. The frequency measurements are mathematically corrected for the stray lead inductance (LS) as indicated by Equation (2). The load resonant measurement uses a fixed load capacitor of 28 pF. The load frequency at a different load capacitance

is calculated using Equation (1).

An S&A 350A Transmission System was used to measure these crystals with the "Calculated", "Measured", and "Physical Capacitor" methods of measurement.

The crystals measured were all fundamental mode crystals. The load capacitor specification varied from 32 pF for the 1 MHz crystal to 15 pF for the 30 MHz crystal. The measurements for these crystals are listed in Table 1.

	CL	TS	1500	150C	350A-C	350A-M	350A-P
1	32	3.3	999907	1.0	1.0	3.0	-1.0
2	30	3.0	4193759	2.1	0.7	0.0	-1.0
3	30	4.5	5002841	3.6	-0.2	-0.6	-1.0
4	22	13.3	10000039	2.4	2.7	-4.6	-3.1
5	20	12.1	12799385	2.4	0.0	-0.2	-1.6
6	20	19.4	15257437	1.5	5.0	0.7	-4.1
7	15	27.2	18296994	2.1	7.8	0.8	-6.7
8	15	22.5	20009678	0.7	2.8	1.3	-2.8
9	15	26.6	21346530	2.1	4.6	-2.6	-5.3
10	15	24.8	30137215	-0.5	7.7	-4.1	4.2

Table 1

The worst case difference was 7.8 ppm for crystal #7 using the "Calculated" method of measurement. This crystal has a trim sensitivity of 27.2 ppm/pF. The 7.8 ppm difference corresponds to a load capacitance error of only 0.3 pF.

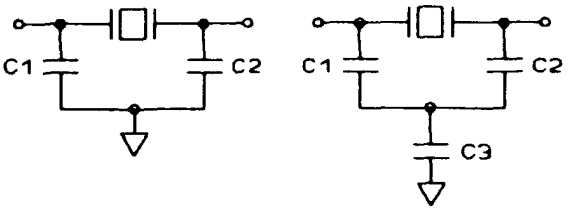
The correlation of measurements between instruments is impacted by many factors. The following paragraphs discuss some of these factors and how they affect the measurement methods previously described.

Crystal Drive Level

It is important to remember that the crystal is a current sensitive device. Crystal measurements should be performed at the specified operating power level. We recommend that a crystal be tested at a drive level of 500 microwatts or less. Higher power levels can cause the crystal to become nonlinear [5]. This nonlinearity results in errors in the frequency measurement of the crystal. The current trend in the industry is to specify even lower power levels than 500 microwatts as older specifications are rewritten to reflect current crystal products.

Crystal Cover Connection

The manner in which the crystal cover is connected will impact the load resonant measurement [3]. The crystal is actually a three terminal device as shown in Figure 14.



Cover Grounded Cover Ungrounded

Figure 14

Where: C1, C2 = Capacitance between Crystal Electrodes and Ground (pF)
C3 = Capacitance from Crystal Cover to Ground (pF)

The impact on the static capacitance measured in a Crystal Impedance Meter is shown by Equations (10) and (11) below.

Grounded Cover:

$$C0\ (MEASURED) = C0 + C1 \tag{10}$$

Ungrounded Cover:

$$C0\ (MEASURED) = C0 + \frac{C1(C2 + C3)}{(C1 + C2 + C3)}$$

$$IF\ C1 = C2\ \ AND\ 2C1 >> C3\ THEN$$

$$C0\ (MEASURED) = C0 + 0.5\ C1 \tag{11}$$

Normally the stray capacitances (C1, C2) from the crystal electrodes to the crystal cover are approximately the same value. Twice this capacitance value is usually much greater than C3 capacitance term. These approximations allow the equation for the ungrounded condition to be simplified to that shown in Equation (11).

Equations (12) and (13) show the impact of the cover connection on the static capacitance measurement made in a Transmission System.

Grounded Cover:

$$C0\ (MEASURED) = C0 \tag{12}$$

Ungrounded Cover:

$$C0\ (MEASURED) = C0 + \frac{C1\ C2}{(C1 + C2 + C3)}$$

$$IF\ C1 = C2\ \ AND\ 2C1 >> C3\ THEN$$

$$C0\ (MEASURED) = C0 + 0.5\ C1 \tag{13}$$

These equations are different from Equations (10) and (11) derived for the CI Meter. The difference in static capacitance measurements between a Crystal Impedance Meter and a Transmission System is shown by Equations (14) and (15).

Grounded Cover:

$$CO\ (CI\ METER - TRANSMISSION\ SYSTEM) = C1 \qquad (14)$$

Ungrounded Cover:

$$CO\ (CI\ METER - TRANSMISSION\ SYSTEM) = \frac{C1\ C3}{(C1 + C2 + C3)} \qquad (15)$$

These capacitance terms for Crystal #10 of Table 1 were measured as: C1 = 1.08 pF, C2 = 1.08 pF, and C3 = 0.3 pF. Using these values, the correlation difference between a Transmission System and a CI Meter is 1.08 pF if the cover is grounded. The 1.08 pF can be converted to a correlation difference in load resonant frequency by multiplying by the trim sensitivity of the crystal. The trim sensitivity for Crystal #10 is 24.8 ppm/pF at 15 pF of load capacitance. The 1.08 pF difference, therefore, corresponds to a load resonant frequency difference of 26.8 ppm.

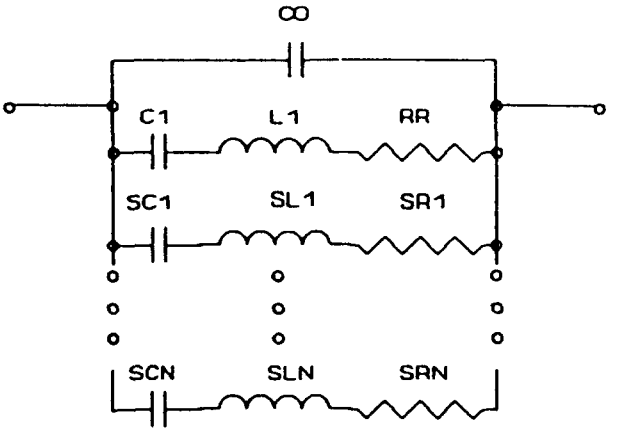
The ungrounded cover correlation difference defined by Equation (15) is only 0.13 pF (3.2 ppm). As this example illustrates, the load frequency correlation between Crystal Impedance Meters and Transmission Systems is better if the measurements are made with the crystal cover ungrounded.

Coupled Modes

A coupled mode of vibration is caused by a different vibrating mode of the crystal coupling into the standard mode of vibration. These coupled modes are normally suppressed by the design of the crystal and are not commonly present in production crystals. The presence of a coupled mode in the frequency region bounded by the series resonant frequency of the crystal and the specified load resonant frequency can have a pronounced effect on the load resonant measurement. This effect was examined in previous paragraphs.

Crystal Spurious Responses

The crystal model shown in Figure 1 assumes a perfectly linear device. Unfortunately this approximation is not always a good one. A real crystal has several spurious frequencies that are greater in frequency but usually much less in activity than the load resonant frequency. If these spurious frequencies are close to the crystal resonance and relatively large in activity they can have an effect on load resonant measurements made with the "Calculated" Transmission System measurement method. A modified crystal equivalent circuit including the spurious responses is shown in Figure 15.



Where: SC1,SL1,SR1 = Capacitance, Inductance and Resistance of the first spurious
SCN,SLN,SRN = Capacitance, Inductance and Resistance of the N'th spurious

Figure 15

The impact of a spurious frequency on the load frequency measurement made with the "Calculated" method can be approximated as a capacitance error. This offset can be calculated using Equation (16).

$$CE = \left[\frac{SC1\ FR}{2\ (SF1 - FR)} + \cdots + \frac{SCN\ FR}{2\ (SFN - FR)}\right] \qquad (16)$$

Where: CE = Capacitance Error (pF)
SF1 = Spurious Frequency #1 (Hz)
SFN = Spurious Frequency #N (Hz)

An example crystal was selected to illustrate the impact of spurious frequencies. The activity response of this crystal is shown in Figure 16.

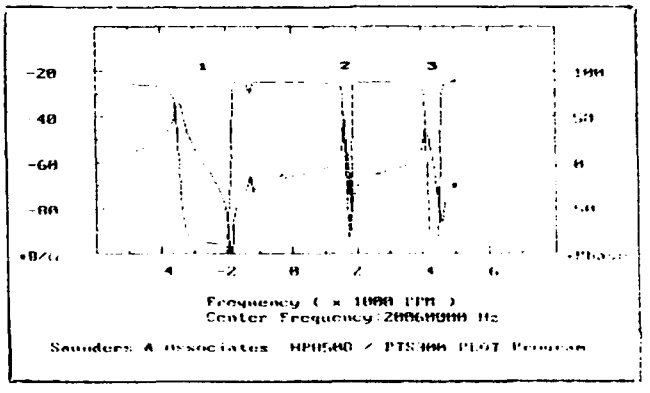


Figure 16

This graph shows the attenuation of the crystal in decibels as a function of frequency. The phase shift across the crystal is also displayed.

The frequency marked #1 is the series resonant frequency of the crystal. The frequencies indicated by #2 and #3 are the spurious frequencies whose impact on the "Calculated" load resonant frequency will be determined.

The motional parameters of each frequency were measured using a S&A 350A system. This information is contained in Table 12.

	FREQUENCY	C1	RR	CO	CL	TS
1	20000202	25.2	6.5	6.5	20	18
2	20091511	2.6	35.5			
3	20142364	4.4	27.3			

Table 12

Using Equation (16), the impact of the two spurious frequencies on the "Calculated" load frequency measurement is 0.58 pF. This corresponds to a frequency difference of 10.4 ppm between the "Calculated" technique and the "Measured", "Physical Capacitor" or Crystal Impedance measurements.

The example crystal was selected because of its large spurious responses and large trim sensitivity. The impact of typical production crystals due to the spurious frequencies is usually minimal.

CONCLUSIONS

Transmission Systems and Crystal Impedance Meters can be used to measure the load resonant frequency of a quartz crystal. The interpretation of this measurement requires an understanding of the measurement method and the characteristics of the crystal.

The specification of load resonant frequency is better represented as a tolerance specified on load capacitance. The load capacitance specification reflects the impact of the crystal and the specified load capacitance on the sensitivity of the load resonant measurement.

Crystal Impedance Meters provide a cost effective method of measuring load resonant frequency. The CI Meter uses a physical capacitor to make the load resonant measurement. If the specified load capacitance is different than the physical capacitor, a mathematical correction can be done to obtain the desired load frequency.

Transmission Systems can measure the load resonant frequency using three distinct methods; "Calculated", "Measured", and "Physical Capacitor".

The "Calculated" method measures the crystal at series resonant frequency and mathematically calculates load resonant frequency. This method can be used if the response of the crystal is well known or precise measurements are not required. The "Calculated" method is susceptible to load frequency errors due to nonlinear crystal responses.

The "Measured" method first measures the crystal in the same manner as the "Calculated" method then makes a correction based upon an impedance measurement at the calculated load resonant frequency. The "Measured method works best for those crystals with a load resonant impedance less than 1000 ohms. The measurements made include the impact of any nonlinear crystal responses.

The "Physical Capacitor" technique uses an actual capacitor inserted in the PI Network to make the load resonant measurement. This method provides a better resolution of load frequency measurement than the "Measured" and "Calculated" methods. The "Physical Capacitor" method also does not limit the drive current available to the crystal. The motional parameters must be mathematically corrected for the stray capacitances of the test socket. The value of these strays must be well known and constant in order to make accurate measurements.

The crystal spurious responses can impact the load resonant measurement made on a Transmission System using the "Calculated" method. In the example presented, the "Calculated" method resulting in a load frequency error of 0.58 pF (10.4 ppm) when compared to the measurements made with other methods.

REFERENCES

[1] Salt, D. (1987). Handbook of Quartz Devices. Van Nostrand Reinhold, UK., pp. 91-106

[2] IEC (1973). Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a pi-network. IEC Publication 444

[3] Horton, W., Payne, T., Smyth, R.,and Symonds, D., (1981). "Comparison of Methods for Measurement of Quartz Crystal Resonators with Load Capacitance", Proc. 35th Ann. Freq. Control Symposium, pp. 271-279.

[4] S&A 2100 Temperature Test System (1991) Saunders and Associates, Inc., Az., pp. A2-1 - A2-32

[5] Saunders, J., (1984). "Drive Level Effects on Transmission Measurements of AT Overtone Resonators", Proc. 6th Annual Quartz Crystal Conference.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FUNDAMENTAL PROPERTIES OF THIN FILM RESONATORS

K.M. Lakin

TFR Technologies, Inc.
701 SE Salmon
Redmond, Oregon 97756

ABSTRACT

This paper presents an overview of the physical properties of thin film resonators, composite resonators using piezoelectric thin films, and filters employing coupled resonators.

There are both similarities and substantial differences between thin film resonators and conventional single-crystal resonators used at low frequencies. For most applications the thin film resonator is fabricated with lateral dimensions in excess of 50 to 100 times the thickness and as such can be considered a one dimensional device for the purposes of wave propagation and circuit modeling. In addition, composite resonators consisting of piezoelectric films supported by membranes or bulk substrates form an important part of the thin film resonator technology. This paper will dwell somewhat on the composite resonator because properties will not be familiar to most readers and because of its potential for many applications.

1.0 INTRODUCTION

The basic resonator cross-sections are illustrated in Fig. 1. The dimensional contrast between low frequency resonators and thin film resonators is shown approximately to scale in Fig. 1b. For filter applications thin film resonators are fabricated with diameter-to-thickness ratios in excess of 50:1 as required [1]. Although low frequency resonators can be fabricated with large diameter-to-thickness ratios the resulting absolute size is impractical for most applications. TFR resonators can be fabricated with smaller resonator areas in order to raise the electrical impedance of the resonator. However, parasitic capacitance effects in electrode interconnects leading to the resonator could dominate actual resonator operation. An AlN resonator with 50 ohms reactance at 1 GHz, from the geometric capacitance, is approximately 400 micrometers square. A corresponding 10 MHz quartz resonator would be nearly 400 mm in diameter.

The smaller diameter-to-thickness ratio of LF resonators requires careful design to manage trapped energy modes in order to increase Q and minimize parasitic resonances. This is less of a problem in TFR devices designed for lower impedances because the area of the electrode is made larger to lower the electrical impedance. Over a large area the spatially periodic plate wave displacement currents sum to a small value compared to the primary thickness resonance. The effect is illustrated qualitatively in Fig. 2.

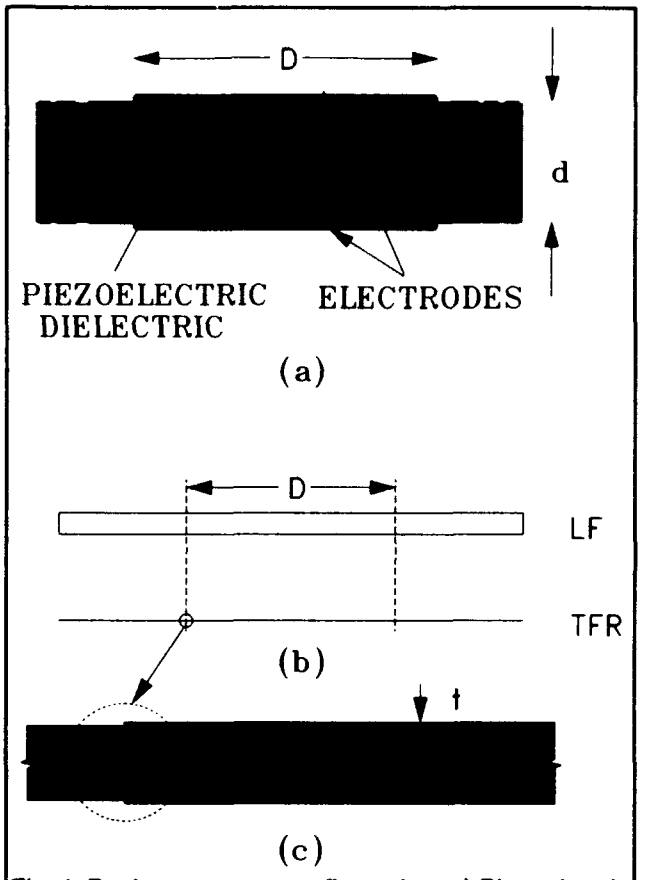


Fig. 1 Basic resonator configuration. a) Piezoelectric resonator, b) illustration of diameter-to-thickness ratios in low frequency (LF) and thin film resonators (TFR), c) electrode edge emphasizing finite electrode thicknesses employed at microwave frequencies.

Plate waves are generated at the edge of the electrodes where there is a large gradient in force in the transverse direction. Plate waves generated at the electrode edge propagate in the plate and produce a normal displacement current that is summed by the electrode area. The transverse periodic variations of displacement current from the plate waves is averaged to a relatively small value by large electrodes. In contrast, the uniform displacement current of the primary thickness mode increases in total current with electrode area. The effect of the plate waves is therefore diminished by large diameter-to-thickness ratios.

CH2965-2/91/0000-201 $1.00 © 1991 IEEE

Plate waves can appear as significant parasitic effects if the waves are trapped within the electroded region or by the edge support for the thin film. A stacked crystal filter having pass ripple due to plate waves is shown in Fig. 3 as an example.

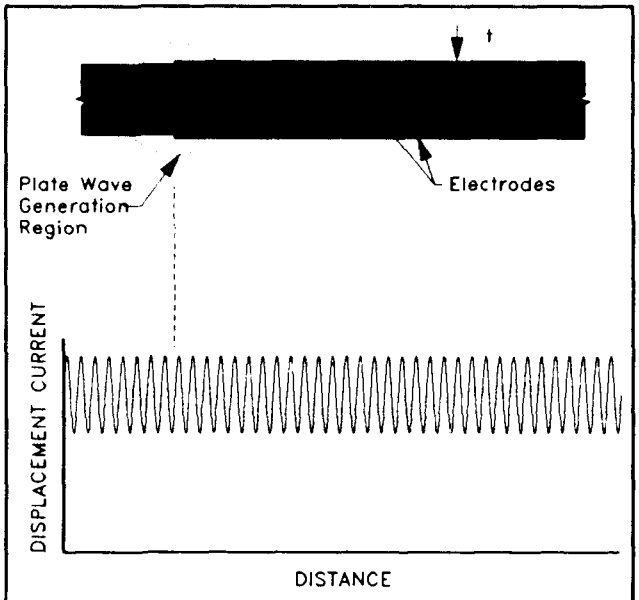


Fig. 2 Illustration of averaging effect of electrode integration on the displacement currents. The plate waves have a periodic variation in the transverse direction that is averaged to a small value whereas the thickness mode displacement current sums to a larger current without cancellation.

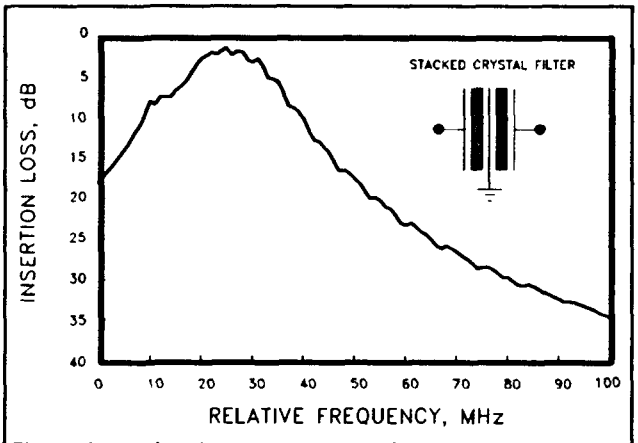


Fig. 3 Insertion loss response of a two-pole stacked crystal filter showing ripple in the passband due to plate wave effects. The center frequency of the filter is near 1 GHz and the multi-layer AlN film is approximately 10 micrometers thick and 400 micrometers square. Two single section devices are connected in series to obtain the two-pole response.

Another manifestation of large area electrodes is to minimize shape effects. Very good filters have been synthesized with resonators having square electrodes rather than circular electrodes common to low frequency resonators. The circular electrodes have an advantage in a smaller edge length for a given area than the square electrodes. Clearly square or rectangular electrodes offer advantages in packing density and pattern generation.

The most significant difference between TFR and conventional low frequency resonators is in the method of fabrication. Since in TFR the piezoelectric materials are grown in film form there is the possibility of combining materials into composite structures [2,3]. The acoustic properties of composite resonators are somewhat more complex because of the variety of modes supported by either film or substrate. The technological goal in composite resonators is to combine positive features of both film and substrate. The advantage of the film piezoelectric materials is good coupling coefficient ease of fabrication. Single crystal substrates offer low acoustic loss and, in some cases, temperature compensation. A summary of general material properties is shown in Table I

TABLE I

Summary of material properties for TFR and composite resonators.

MATERIAL	K^2, %	V, 10^5 cm/sec	TC(1), ppm/°C	REMARKS
AlN	6.5-7	11.4	-26	c-axis
AlN	4.5	6.33	-26	in-plane
ZnO	7.8	6.37	-59	c-axis
ZnO	6.3	2.83		in-plane
$LiNbO_3$	27.4	4.46	High	Y-cut, Quasi-shear
$LiNbO_3$	13.3	7.32	High	36° Rot. Y-cut, Long.
$LiTaO_3$	14.1	4.14	Low	X-cut Quasi-shear
Quartz	0.77	3.32	0	AT-cut

The film materials, ZnO and AlN, allow high frequency resonator fabrication whereas the single crystal piezoelectrics are more suited for lower frequency applications or as substrates. Other substrates for composite devices are Si, GaAs, sapphire, and a number of single crystal ceramics.

2.0 COMPOSITE RESONATOR ANALYSIS

The general analysis problem of composite resonators is in determining effective piezoelectric coupling coefficient and mode structure. The lack of precision in the crystallographic orientation of thin films, as compared to plates cut from bulk single crystals, can lead to multi-mode generation in composite or single plate resonators. A general analysis of the quasi-mode

resonators was completed and is summarized below. The analysis is for the geometry shown in Fig. 4 having a piezoelectric film on a general substrate.

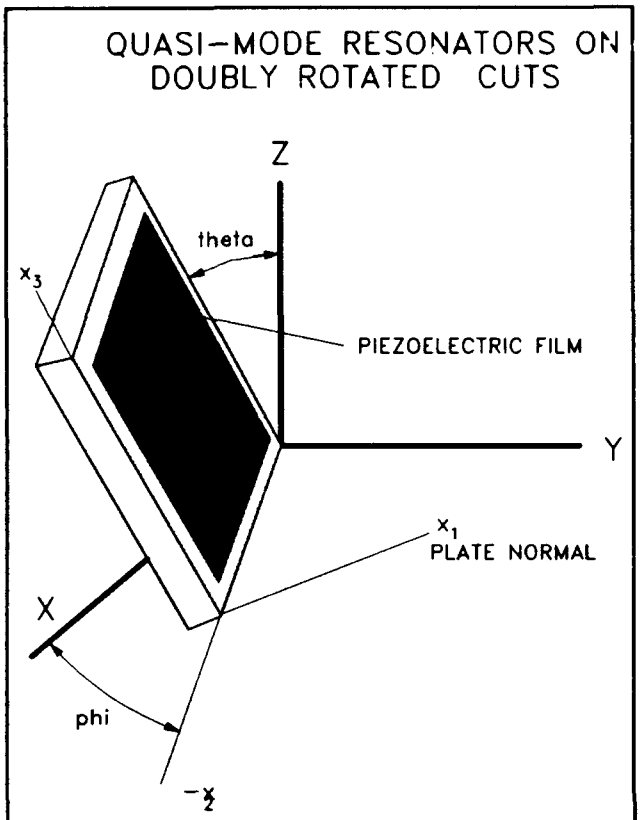


Fig. 4 Material configuration for resonator analysis. The film and substrate are of arbitrary orientation, show here as doubly rotated cuts. Complex plane-wave modes may be generated in substrate and film depending on the orientations.

The purpose of this analysis task was to derive concise equations for quasi-mode resonator analysis that could be readily programmed in FORTRAN as a series of subroutines for use in two-port network cascading similar to that used in simple pure mode resonator analysis using Mason models. Once a resonant mode is identified and given equivalent pure mode parameters, then the computationally more efficient pure mode analysis program can be used for device design.

The theory for the quasi-mode resonator is developed from the fundamental equations of linear piezoelectric elasticity,

$$\frac{\partial T_{ij}}{\partial x_i} = -\rho\omega^2 u_j \tag{1}$$

$$\frac{\partial D_i}{\partial x_i} = 0 \tag{2}$$

$$T_{ij} = c^E_{ijkl}\frac{\partial u_k}{\partial x_l} + e_{ijk}\frac{\partial \varphi}{\partial x_k} \tag{3}$$

$$D_i = e_{ikl}\frac{\partial u_k}{\partial x_l} - \epsilon^S_{ik}\frac{\partial \varphi}{\partial x_k} \tag{4}$$

where time harmonic and quasi-electrostatic conditions have been assumed. Further restrictions are made to the one dimensional plane wave case,

$$u_j = \beta_j e^{j\omega t} e^{-jkx_1} \tag{5}$$

Potential φ is eliminated as an explicit variable by solving (2) using (4) to obtain,

$$\varphi(x) = \frac{e_{11}}{\epsilon^S_{11}}u_1(x) + \frac{e_{16}}{\epsilon^S_{11}}u_2(x) + \frac{e_{15}}{\epsilon^S_{11}}u_3(x) + ax + b \tag{6}$$

where a and b are constants to be determined by boundary conditions. In an unbounded region the coefficients a and b must be zero.

The result of using (6) and (7) in (1-4) is a set of Christoffel equations for propagation in the $\pm x_1$ direction,

$$\{[\Gamma_{ij}] - \lambda[1]\}[\beta_j] = 0 \tag{7}$$

where $\lambda = \rho V^2$, V is the propagation velocity and,

$$[\Gamma] = \begin{bmatrix} C_{11} & C_{16} & C_{15} \\ C_{16} & C_{66} & C_{56} \\ C_{15} & C_{56} & C_{55} \end{bmatrix}$$

$$C_{11} = C^E_{11} + e_{11}e_{11}/\epsilon^S_{11}$$
$$C_{16} = C^E_{16} + e_{11}e_{16}/\epsilon^S_{11}$$
$$C_{15} = C^E_{15} + e_{11}e_{15}/\epsilon^S_{11}$$
$$C_{66} = C^E_{66} + e_{16}e_{16}/\epsilon^S_{11}$$
$$C_{56} = C^E_{56} + e_{15}e_{16}/\epsilon^S_{11}$$
$$C_{55} = C^E_{55} + e_{15}e_{15}/\epsilon^S_{11}$$

All material constants are referred to the wave coordinate system for propagation in the x_1 direction as illustrated in Fig. 2 for a doubly rotated crystal and attached piezoelectric film.

As is well known, the homogeneous set of equations (7) leads to the requirement that

$$\det\{[\Gamma_{ij}] - \lambda[1]\} = 0$$

The resulting third order polynomial in λ has roots m that define three modes of propagation and associated eigenvectors from

$$\{[\Gamma_{ij}] - \lambda^{(m)}[1]\}[\beta_j^{(m)}] = 0 \qquad m = 1, 2, 3$$

The general solution is a linear combination of three possible solutions, one for each mode, for each direction of propagation,

$$u_k = \sum_{m=1}^{3} \beta_k^{(m)} e^{-jk^{(m)}x_1} A_-^{(m)} + \sum_{m=1}^{3} \beta_k^{(m)} e^{+jk^{(m)}x_1} A_+^{(m)}$$

where $\beta_k^{(m)}$ are found from (7b) and $A_\pm^{(m)}$ are constants to be found from the boundary conditions.

For completeness and later use, the electric displacement current in the x_1 direction is obtained from (4) and (6) as,

$$D_1 = -\epsilon_{11}^S \alpha$$

For purposes of solving one dimensional boundary value problems it is useful to use particle velocity and traction forces as the system variables and then to reduce the complexity of the algebra and notation by using matrix techniques.

The variables can now be put in matrix notation. First for particle velocity,

$$[-v(x_1)] = [\beta][e^{-jkx_1}][A_-] + [\beta][e^{+jkx_1}][A_-]$$

Here $[\beta]$ is a 3x3 matrix for $\beta_k^{(m)}$ whose rows are indexed by k and columns by m and $[A_\pm^{(m)}]$ are similarly defined column vectors. The phase of the three modes is represented by the matrix,

$$[e^{\pm jkx_1}] = \begin{bmatrix} e^{\pm jk^{(1)}x_1} & 0 & 0 \\ 0 & e^{\pm jk^{(2)}x_1} & 0 \\ 0 & 0 & e^{\pm jk^{(3)}x_1} \end{bmatrix}$$

The electrostatic potential φ in matrix notation is,

$$\varphi(x_1) = \frac{[\tilde{e}]^T}{-j\omega\epsilon_{11}^S}[-v(x_1)] + \alpha x_1 + b, \qquad [\tilde{e}] = \begin{bmatrix} e_{11} \\ e_{16} \\ e_{15} \end{bmatrix}$$

The forces of interest are the traction forces acting on the plane perpendicular to the x_1 direction since they must be continuous across any material discontinuities occurring in the x_1 direction. The forces are,

$$[n_1 T_{1j}] = [f] = \begin{bmatrix} T_1 \\ T_6 \\ T_5 \end{bmatrix}$$

By expanding (3) and using the definition for particle velocity it can be shown that force is given by,

$$[f(x_1)] = [H][e^{-jkx_1}][A_-] - [H][e^{+jkx_1}][A_-] + [\tilde{e}]\alpha$$

where $[H] = [Z][\beta]$ and the multi-mode mechanical impedance is defined by a 3x3 matrix,

$$[Z] = [-v]^{-1}[f] = [\Gamma][\beta][V]^{-1}[\beta]^{-1}$$

and,

$$[V] = \begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix}$$

The multi-mode piezoelectric plate having electrodes is a boundary value problem well suited to the matrix approach. For the derivation of the electrical impedance of the parallel plate transducer it was convenient to give the plate a symmetric characteristic with particle velocities outward and evaluate phase at points equidistant from the center. The boundary conditions are described by mechanical impedances at each mechanical port and an applied voltage at the electrical port. Thus the problem is to find the electrical current under these conditions. From the above equations the port equations are found,

$$[f(h)] = [H][e^{-j\Phi}][A_-] - [H][e^{+j\Phi}][A_-] + [\tilde{e}]\alpha$$

$$[-v(h)] = [\beta][e^{-j\Phi}][A_-] + [\beta][e^{+j\Phi}][A_-]$$

$$[f(-h)] = [H][e^{+j\Phi}][A_-] - [H][e^{-j\Phi}][A_-] + [\tilde{e}]\alpha$$

$$[-v(-h)] = [\beta][e^{+j\Phi}][A_-] + [\beta][e^{-j\Phi}][A_-]$$

$$\varphi(h) = \frac{[\tilde{e}]^T}{-j\omega\epsilon_{11}^S}[-v(h)] + \alpha h + b$$

$$\varphi(-h) = \frac{[\tilde{e}]^T}{-j\omega\epsilon_{11}^S}[-v(-h)] - \alpha h + b$$

The boundary conditions are,

$$V = \varphi(h) - \varphi(-h)$$

$$[f(h)] = [Z_t][-v(h)]$$

$$[f(-h)] = -[Z_t][-v(-h)]$$

$$I = \frac{dq}{dt} = j\omega q = -j\omega \oint \vec{D} \cdot d\vec{S} = j\omega C_t \quad \alpha d$$

$$C_t = \frac{\epsilon_{11}^S (\text{area})}{d}$$

Now use the boundary conditions in the port equations to find the three unknown coefficients $[A_+][A_-]$ and α. The result is,

$$Z = \frac{1}{j\omega C_t}(1 - [\tilde{e}]^T[\beta][\tan\phi][\Gamma]$$

$$[\phi]^{-1}[\beta]^{-1}[\Gamma]^{-1}[\tilde{e}]/\epsilon_{11}^S)$$

$$[T] = \{[T_2]^{-1}[T_1] + [T_4]^{-1}$$

$$[T_3]\}^{-1}\{[T_2]^{-1} + [T_4]^{-1}\}$$

$$[T_1] = [1] + j[z_r][\tan\phi]$$

$$[T_2] = [z_r] + j[\tan\phi]$$

$$[T_3] = [1] + j[z_t][\tan\phi]$$

$$[T_4] = [z_t] + j[\tan\phi]$$

$$[z_r] = [\beta]^{-1}[Z]^{-1}[Z_r][\beta]$$

$$[z_t] = [\beta]^{-1}[Z]^{-1}[Z_t][\beta]$$

The above equation(s) hold for a one dimensional plane wave plate resonator of arbitrary complexity and number of layers. Complex structures are analyzed using matrix multiplication to cascade various sections. If the resonator is just a complex piezoelectric plate without attached layers or loads, then the impedance reduces to,

$$Z = \frac{1}{j\omega C_t}(1 - [\tilde{e}]^T[\beta][\tan\phi/\phi][\beta]^{-1}[\Gamma]^{-1}[\tilde{e}]/\epsilon_{11}^S)$$

If the mode structure of the plate and resonator is simple then the impedance relation is that for a single mode,

$$Z = \left(\frac{1}{j\omega C_t}\right)\{1 - K^2\frac{\tan\phi}{\phi}$$

$$\left[\frac{(z_r + z_t)\cos^2\phi + j\sin 2\phi}{(z_r + z_t)\cos 2\phi + j(z_r z_t + 1)\sin 2\phi}\right]\}$$

Finally, the electrical impedance for a simple pure mode single plate resonator is given by,

$$Z = \left(\frac{1}{j\omega C_t}\right)\left\{1 - K^2\frac{\tan\phi}{\phi}\right\}$$

The general impedance equations were programmed in FORTRAN and applied to a number of quasi-mode or composite resonator configurations.

An examination of doubly rotated crystal cuts using a transmission line model has been used by Ballato [4,5] to examine temperature coefficients and other properties of transducer plates and composite geometries such as the stacked crystal filter.

Specific cases are illustrated in Fig. 5-7 . In Fig. 5 is shown the configuration for AlN film resonator having orientations of c-axis normal and 5 degree c-axis tilt as might occur under some film deposition conditions. If a c-axis normal film of AlN is deposited on SC-cut quartz the longitudinal wave from the AlN film is coupled to both shear waves in the substrate generating the complex resonator response shown in Fig. 6. For excitation of a shear mode in AlN it is necessary to rotate the c-axis of the film 45 degrees towards the basal plane. Then the film is rotated about the plate normal in order to align the film shear wave's polarization with the desired shear wave mode in the substrate. Fig. 7 shows the resonator response for a properly aligned film and the resultant reduction in effective coupling coefficient apparent from the phase response.

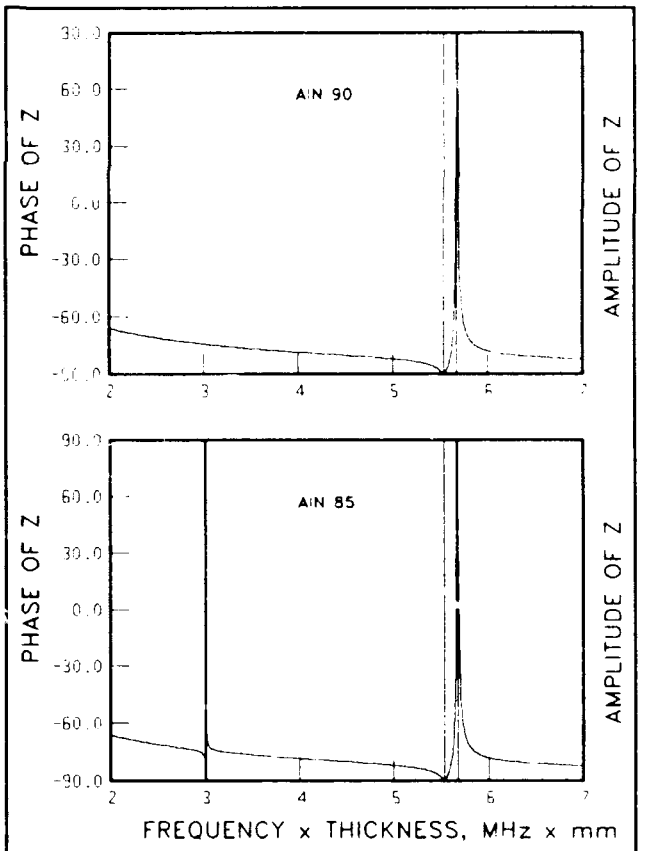


Fig. 5 Impedance response of AlN resonator. a) c-axis normal orientation, b) tilted (5 degrees) c-axis showing shear wave resonance.

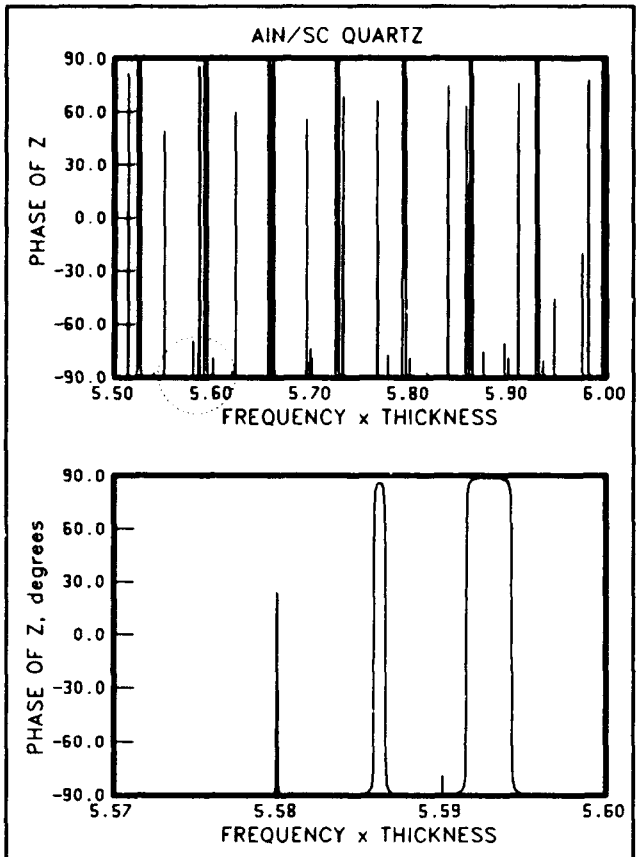


Fig. 6 Analysis of resonator impedance phase for AlN film on SC-cut quartz. The result shows the effect of mode coupling in the substrate generating both shear and longitudinal resonances. The lower plot is an expansion of the encircled portion of the upper plot.

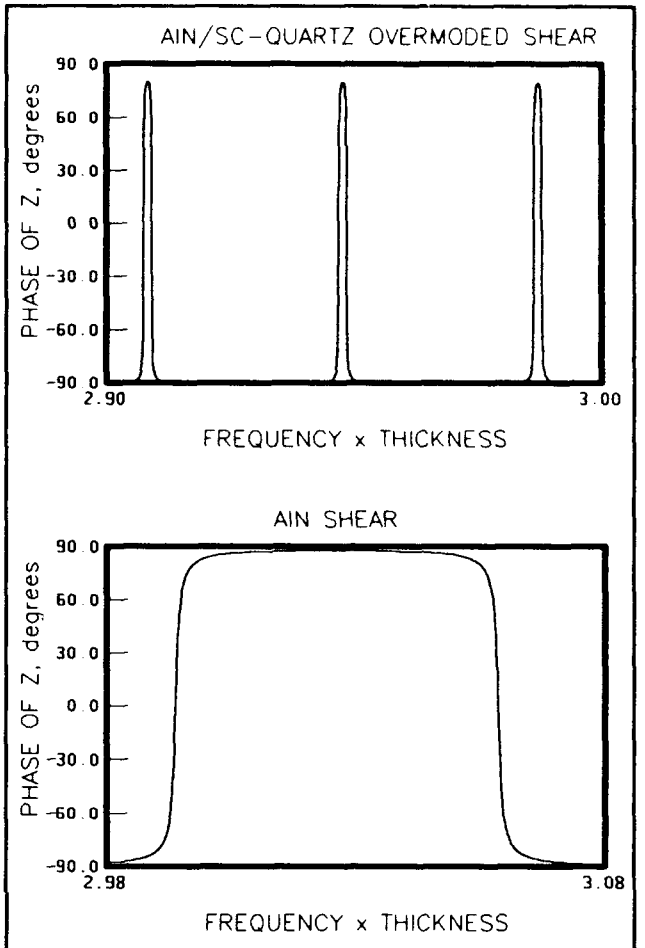


Fig. 7 AlN shear wave film on SC cut quartz substrate. The lower figure is for the AlN alone and illustrates the reduction of resonator effective coupling coefficient in the overmoded composite structure. However, the effective coupling coefficient of the composite is still as large as a fundamental mode quartz plate. The frequency-thickness product is relative to the film.

3.0 SUMMARY

For many applications thin film resonators are of such large diameter to thickness ratio to be treated to first order as one dimensional wave propagation devices as verified by previous experimental results. An analysis of the general quasi-mode composite resonator was presented and the results summarized in electrical impedance equations cast in a matrix algebra format. The formalism is being used to design composite resonators and to predict the effects of film orientation on device performance and yield.

REFERENCES

1. "Stacked Crystal Filters Implemented with Thin Films," K.M. Lakin, G.R. Kline, R.S. Ketcham, and J.T. Martin, 43rd Annual Frequency Control Symposium, Denver, CO, pp. 536-543, May 31-June 2, 1989.
2. "A Thin-Film CdS-Quartz Composite Resonator", T.R. Sliker and D.A. Roberts, J. Appl. Physics, Vol. 38 No. 5, Apr. 1967, pp. 2350-2358.
3. "Design of Ultra High Frequency Composite Piezoelectric Resonators," J.J. Rosenbaum, Proceedings of the Sixth International Symposium on Applications of Ferroelectrics 1986, pp. 296-299.
4. "Doubly Rotated Thickness Mode Plate Vibrators", A. Ballato, in Physical Acoustics: Principals and Methods Vol. XIII (W.P. Mason and R.N. Thurston, eds. Academic Press, 1977
5. "A Novel Frequency Selective Device: The Stacked Crystal Filter," A. Ballato and T. Lukaszek, Proceedings of the 27th Annual Symposium on Frequency Control, 1973, pp. 262-269.

HIGH FREQUENCY OSCILLATORS USING COINTEGRATED BAW THIN-FILM PIEZOELECTRICS WITH MICROWAVE BJTS

S. G. Burns, R. J. Weber, and S. D. Braymen
Microelectronics Research Center
Iowa State University
1925 Scholl Road
Ames, Iowa 50011

Abstract

We report on the design of UHF and L-band oscillators using of sputter-deposited, thin-film, aluminum nitride resonators cointegrated with microwave f_T = 2.5 GHz BJTs. The resonator is used as the feedback element. Design techniques and suggestions for novel circuit and system architectures using this technology are presented.

Introduction

Sputter-deposited AlN thin-film resonators (TFR) have been used as the feedback frequency control element in a variety of hybrid single mode, comb line, and voltage-controlled UHF and L-band oscillators [1-4]. Cointegration of the TFR and active devices has also been demonstrated where the piezoelectric material, AlN, was dc magnetron sputtered onto a p^+ doped silicon membrane whose function was to serve as an etch stop for the EDPW anisotropic etch used to remove substrate silicon [5,6]. Although the process of using a p+ etch stop is well defined and yields high-Q TFR devices, it is not as compatible with the process flow for the active devices as would be desired and these process compatibility issues have limited the scope of cointegrated circuits and systems. In this paper, we present an overview of a novel process flow designed to address these compatibility issues and the oscillator design methodology. This technology uses RIE trench-isolated 2.5 GHz BJTs cointegrated with high-Q AlN resonators synthesized with an anisotropic etch along the <111> crystal axes with the first level Al metal serving as the etch stop. Dielectrically assisted liftoff is used for AlN definition. A typical oscillator design includes one gain stage with the TFR connected so as to obtain either a Colpitts or Pierce configuration. The TFR is modeled by a Butterworth Van-Dyke equivalent circuit where discrete circuit parameters are obtained from a TOUCHSTONE™ optimization fit from S-parameter measurements. The resultant design is then incorporated into the top metallization layer mask. BJT S-parameter data, AlN resonator data, and oscillator design methodology for a single-mode cointegrated 1 GHz oscillator using these fabrication and design techniques will be presented along with suggestions for additional advanced applications.

Cointegrated Process Discussion

Cointegrated circuits, illustrated conceptually in Figure 1, requires a compatible semiconductor process that permits fabrication of high-performance BJTs with high f_T on the same substrate as high-Q thin-film AlN resonators.

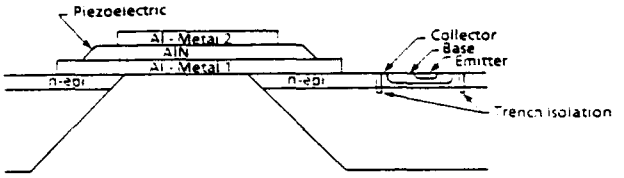


Figure 1. Cointegration concept. Schematic representation showing integration of a TFR and Si BJT on the same substrate.

The UHF and L-band oscillators currently under evaluation are design using the ASIC shown in the ICED layout, Figure 2. This ASIC includes trench-isolated npn transistors and resistors and digitally-scaled MOS capacitors with fusible links for tuning if required. Table 1 summarizes the component content in the ASIC chip. A series of test patterns are also incorporated in the design.

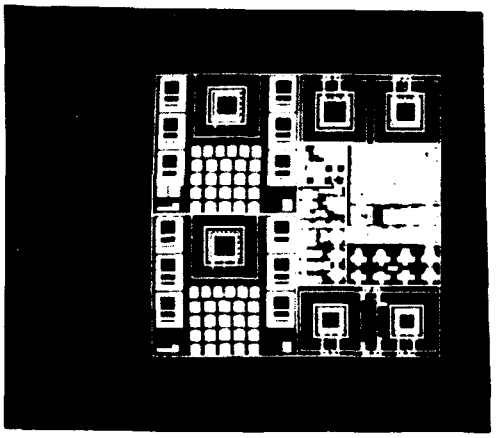


Figure 2. ICED ASIC layout.

CH2965-2/91/0000-207 $1.00 © 1991 IEEE

Table 1. ASIC COMPONENT DESCRIPTION

COMPONENT	NUMBER	DESCRIPTION
npn type 1	8	4 μm, f_T = 2.5 GHz
npn type 2	12	4 μm, f_T = 1.5 GHz
npn type 3	12	2 μm × 25 μm, five fingers
pnp type 1	1	lateral, low β
pnp type 2	1	lateral, β < 25
resistor(s)	120	10 Ω - 2 kΩ, guard ring and active base diffusions
capacitors	48	1.25 pF - 10 pF, MOS
inductors	3	5 T, 275 μm square, 20 μm pitch, 5 nH.
resonators	6	400 μm × 400 μm and 300 μm × 300 μm

The ten-mask process, using 2 μm design rules, is shown in Figure 3. The substrate material is a p-type 100 mm silicon wafer, both sides polished. An n^+ spreader layer and an n-type drift layer are epitaxially grown on the substrate. *Mask 1* is used to define the trench isolation for all transistors, resistors, and capacitors. Trenching is subsequently performed using RIE. *Mask 2* is used to define the sinker oxide etch and diffusion with trench refill using either SiO_2 or polysilicon. This refill is used to obtain surface planarity for subsequent photoresist deposition and metallization. *Mask 3* is used to define the high-resistivity base guard-ring diffusion. The low-resistivity active-base diffusion is defined by *Mask 4*. This diffusion is also used for the resistors. A shallow emitter diffusion is obtained using *Mask 5* yielding an active base region thickness under 0.25 μm. *Mask* 6 is used to define the via etch for all the device contacts. Silicon-doped aluminum (1-2%) is used for the first metal layer. Metal 1, defined by *Mask 7*, is used for the bottom TFR electrode as well as for test structure device contacts. *Mask 8* is used for the nominal 400 μm × 400 μm AlN piezoelectrically-active area definition. A dielectric film is then deposited on the wafer for lift-off. Oriented AlN, nominally 5 μm thick at 1 GHz, is deposited on the entire 100 mm wafer using dc magnetron sputtering of a 99.999% purity aluminum target in a nitrogen plasma. Experimental results have shown that this lift-off procedure provides significantly higher coupling coefficients then the previously reported aperture masks. The top metal layer, metal 2, defined by *Mask 9*, is used for the top TFR electrode definition as well as the circuit interconnects. A cleaved cross-section of the c-axis vertical AlN is shown in Figure 4 along with a top view of the TFR showing the multilayer defined structure. The AlN is 5 μm thick with a 0.4 μm Al top electrode. By not etching a backside via, one obtains an overmoded TFR response. To obtain single-mode and overtone operation, *Mask 10* is used to define the backside via. Anisotropic etching occurs along the <111> crystal axes with the first metal Al serving as the etch stop. The diffusion profiles required for high f_T transistors are considerably easier to obtain because a p^+ layer is not used for the resonator etch-stop membrane. As observed from the process flow, virtually all device processing is completed except for the topside circuit-defining metallization before the TFR processing is initiated. Other electrode configurations are used for spurious mode control.

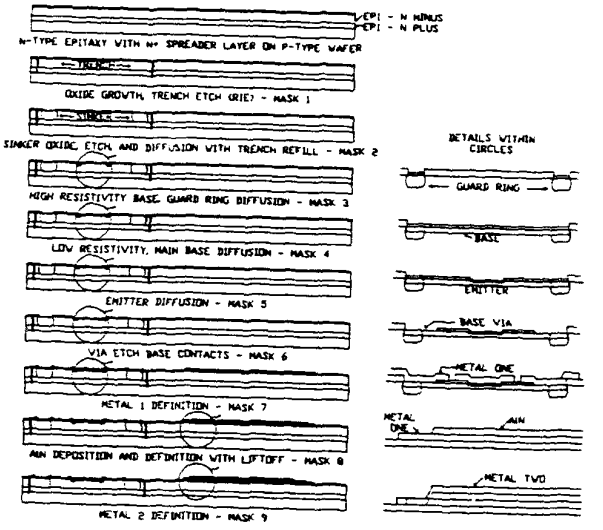
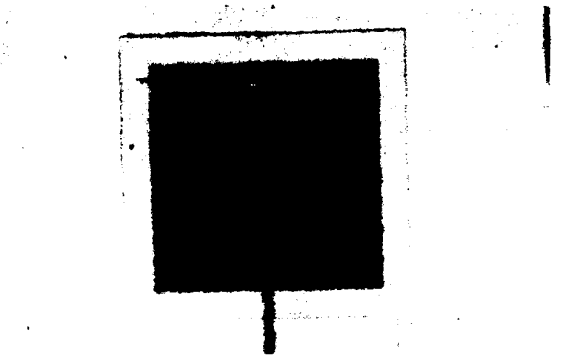


Figure 3. Cointegrated process flow using nine masks. Mask 10 is for the TFR via from the backside of the die.



Figure 4. Cleaved AlN used for TFR and top side view.



Oscillator Design

The oscillator was designed, using the three-port paradigm shown in Figure 5, with the measured S-parameters of the BJT devices and TFRs as input to design algorithm given by

$$| [S][S'] - [1] | = 0 \qquad (1)$$

where [S] is the three-port scattering matrix of the active device portions of the network and [S'] is the three-port scattering matrix of the embedding and interconnecting network which includes the load. This closed form solution technique is described by Weber [7]. The TFR S-parameters, lumped constant R, L, and C circuit elements including parasitics, and the BJT S-parameters were input to EEsof's TOUCHSTONE™ circuit analysis program to synthesize the oscillator topology. An oscillator design may include one or more one- or two-port TFRs or TFR-based filters and one or more oscillating, amplifying, and sustaining active stages. S_{11} for a cointegrated 1.18 GHz TFR is shown in Figure 6(a) along with its Butterworth VanDyke equivalent circuit obtained from a TOUCHSTONE optimization routine, Figure 6(b). [S] for a typical f_T = 2.5 GHz BJT is shown in Figure 7.

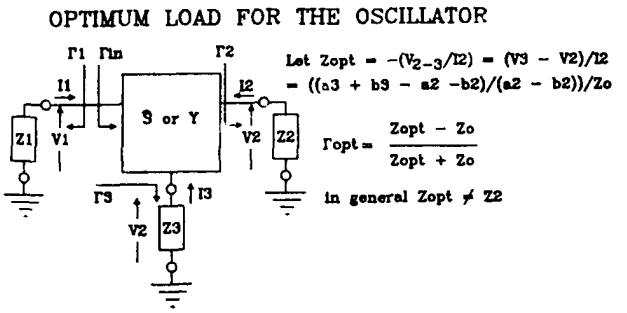


Figure 5. Three port oscillator configuration.

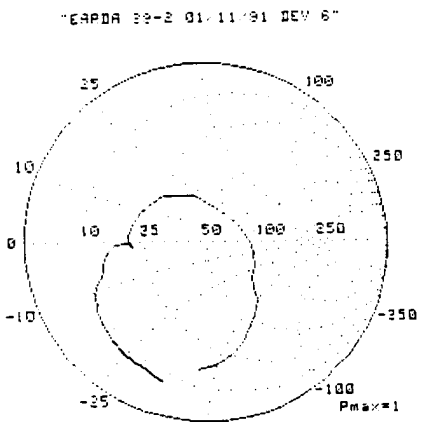


Figure 6(a). S_{11} of TFR cointegration on wafer 39-2 f = 1.18 GHz.

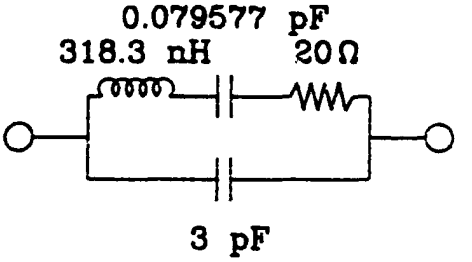


Figure 6(b). Butterworth VanDyke model for a 1 GHz TFR component values determined by a TOUCHSTONE™ optimization.

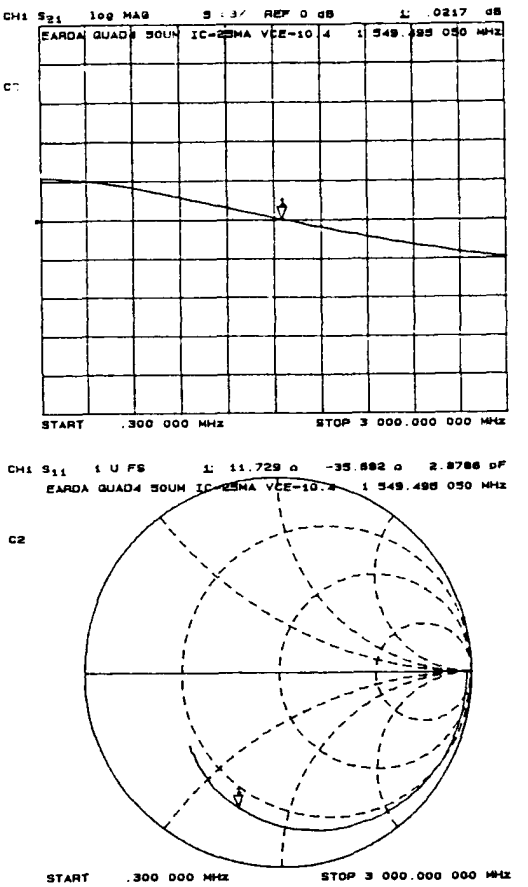


Figure 7. S_{21} and S_{11} for a cointegrated 2.5 GHz BJT.

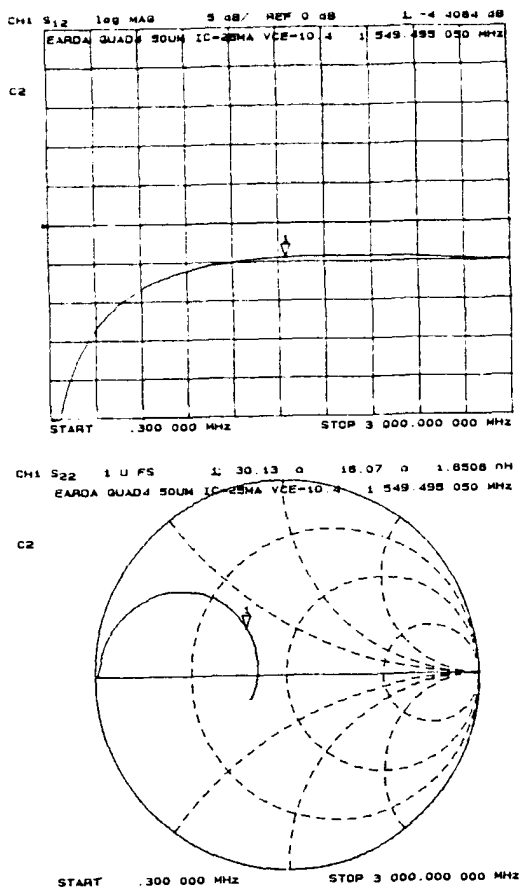


Figure 7. Continued. S_{12} and S_{22} for a cointegrated 2.5 GHz BJT.

As shown in Figure 8, a 1 GHz oscillator design, we employ a Darlington gain stage to accommodate the losses in the circuit elements including the 20 Ω TFR series resistance. As shown in Figure 9, a topside partial view of the ASIC which includes the Figure 8 design, an onboard spiral inductor is used to provide additional phase delay and is not required for designs where $f_o < 25\ \%\ f_T$.

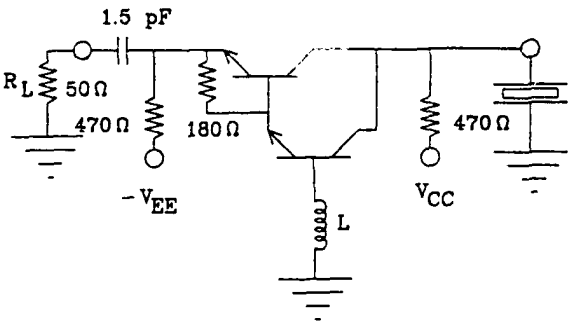


Figure 8. Darlington common-base amplifier used to obtain an oscillator.

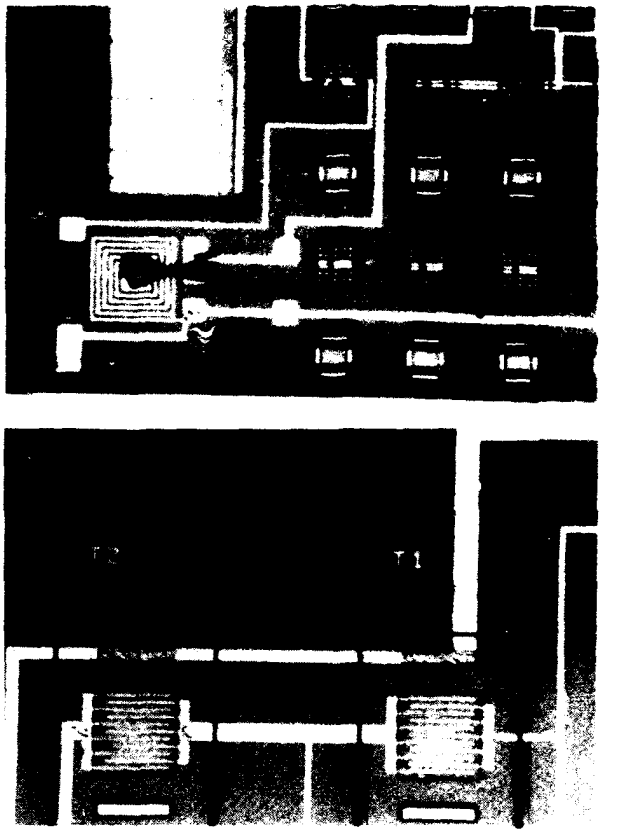


Figure 9. ASIC design showing scaled capacitors, matching inductor and Darlington connected BJTs.

Conclusions and Directions for Future Work

Performance of the oscillator design shown in Figure 8 is currently being evaluated with a view to design and implement cointegrated versions of the previously reported hybrid single mode, overmoded, VCO, and mode-selectable configurations. Near term work will focus on the designs and applications of these cointegrated TFR oscillators and bandpass amplifiers. In addition, to the applications in communications, navigation, avionics, and telemetry systems, work is also proceeding on using the TFR as a chemical sensing element [8]. This concept is illustrated in Figure 10 where the outputs of two identical TFR-based cointegrated oscillators (one exposed to the sensing environment and the other TFR passivated) are mixed to obtain a low-frequency baseband output proportional to nanogram mass sensitivity.

Acknowledgements

This work was supported primarily by the Iowa Department of Economic Development Education and Research Development Program and the Department of Electrical Engineering and Computer Engineering at Iowa State University. The work has been conducted at the Iowa State University Microelectronics Research Center.

References

[1] S. G. Burns and P. H. Thompson, "Design, Analysis, and Performance of UHF Oscillators Using Thin-Film Resonator-Based Devices as the Feedback Element," *Proceedings of the IEEE 32nd Midwest Symposium on Circuits and Systems*, 1989, pp. 1005-1010.

[2] M. M. Driscoll, S. V. Krishnaswany, R. A. Moore, and J. R. Szedon, "UHF Film Resonator Evaluation and Resonator-Controlled Oscillator and Filter Design Using Computer-Aided Design Techniques," 1985 IEEE MTT-S International Symposium.

[3] S. G. Burns, G. R. Kline, and K. M. Lakin, "Design and Performance of Oscillators Using Semiconductor Delay Lines," *Proceedings of the IEEE Ultrasonics Symposium*, 1987, pp. 369-373.

[4] P. H. Thompson, S. G. Burns, G. R. Kline, and R. J. Weber, "Design and Performance of Voltage-Controlled Oscillators Using TFR Stacked Crystal Filters," *Proceedings of the 43rd Annual IEEE Symposium on Frequency Control*, 1989, pp. 58-62.

[5] A. Burkland, A. R. Landin, G. R. Kline, and R. S. Ketcham, "A Thin-Film Bulk-Acoustic Wave Resonator Controlled Oscillator on Silicon," *IEEE Electron Device Letters*, Vol. EDL-8 (11), pp. 531-533, November 1987.

[6] H. Satoh, H. Suzuki, C. Takahashi, C. Narahara, and Y. Ebata, "A 400 MHz One-Chip Oscillator Using an Air-Gap Type Thin Film Resonator," *Proceedings of the Ultrasonics Symposium*, 1987, pp. 363-368.

[7] R. J. Weber, "Oscillator Design Using S Parameters and a Predetermined Source or Load," *45th Annual IEEE Symposium on Frequency Control*, 1991.

[8] R. O'Toole, S. G. Burns, H. R. Shanks, A. D'Silva, R. J. Weber, and M. D. Porter, "Microwave Frequency Oscillators Based on AlN: Toward Integrated Chemical Sensors," Optical and Piezoelectric Sensors Symposium, of the Electrochemical Society, October 1990.

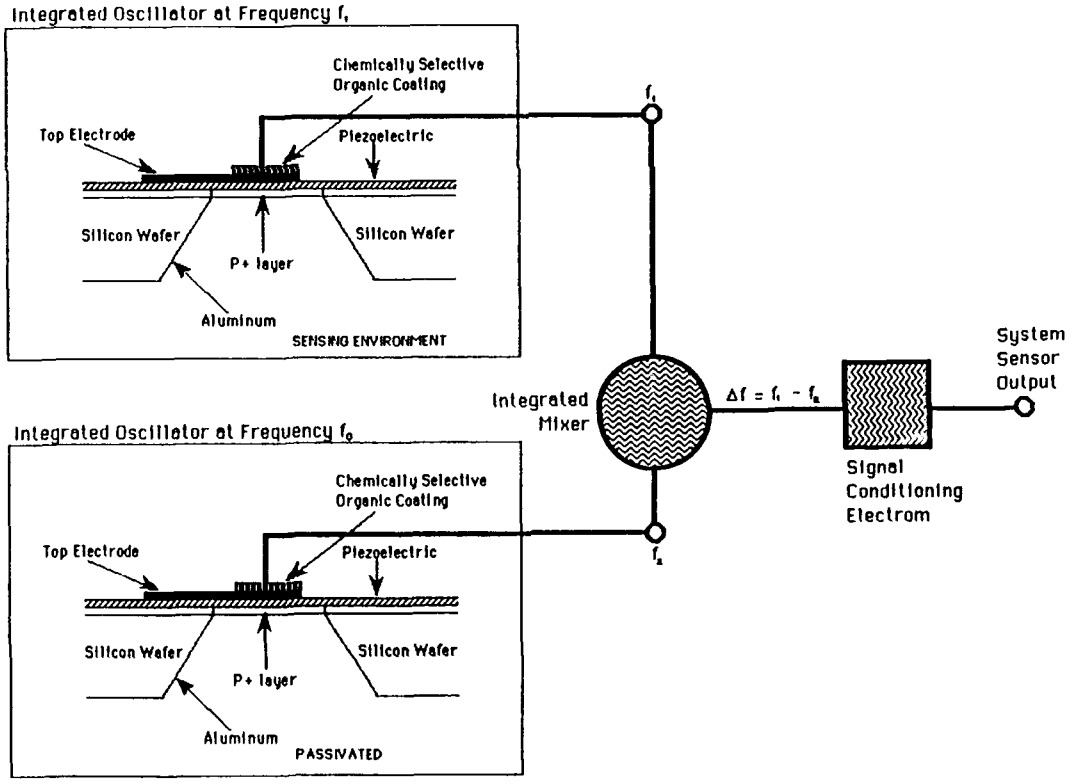


Figure 10. Proposed thin-film piezoelectric resonator for sensor applications.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PIEZOELECTRIC RESONATORS BASED ON SINGLE CRYSTALS OF STRONG PIEZOELECTRICS

V.B.Grouzinenko, V.V.Bezdelkin

Scientific-Research Institute "Phonon"
Moscow, U.S.S.R.

Abstract

The investigation results of piezoelectric BAW resonators using single crystals of"strong" piezoelectrics are given. The possibility of face shear mode excitation in lithium tantalate piezoelectric elements of single-rotated cuts is reported which enables manufacturing resonators with zero frequency-temperature coefficient. The information regarding resonance spacing value of the resonator versus operating mode of vibration and the frequency range used is included. The realizable parameters of new resonator types using lithium tantalate and $La_3Ga_5SiO_{14}$ (LGS) are also given.

Introduction

The search and application of new piezoelectric materials, investigation of physical-chemical properties of known piezoelectric crystals are the main trends of improvement of piezoelectric BAW resonators: their miniaturization, enlargement of resonance spacing, temperature and time frequency stability, decreasing capacitance ratio, the inductance etc.

For many fields of application (oscillators with frequency pulling, synthetizers for TV sets, microprocessors and communication channels of high density, etc.) resonators with a wide resonance spacing and small motional inductance are required. The quality of quartz single crystals attained nowadays at the expense of improving their growing up technology and subsequent processing (sweeping, crystal elements etching in the form of inverted mesas, etc.) made it possible to enlarge the frequency range of crystal units operating at the fundamental frequency from 30 MHz up to 500 MHz, providing their resonance spacing of order of 0.1%. However, for many types of radioelectronic equipment this is insufficient.

The most wide application in the world practice find at present BAW resonators with piezoelectric element made of lithium tantalate and lithium niobate ($LiTaO_3$, $LiNbO_3$) single crystals, having sufficiently high electromechanical coupling factor. Resonators based on these materials have a wide, up to 11%, resonance spacing, small capacitance ratio and small inductance. Moreover, lithium tantalate resonators have temperature compensated cuts with zero frequency-temperature coefficient of the first order (Tables 1 to 3).

Predicted (1978) and experimentally found (1985) in the U.S.S.R. [1, 2], new cuts with zero FTC under special field excitation enable to design miniature lithium tantalate resonators with piezoelectric elements operating in face shear modes over the frequency range 300-3000 kHz. Such resonators reveal frequency-temperature characteristic in the form of "inverted" symmetrical, 2nd order parabola with a positive steepness coefficient $c = (0.7 - 1.5) \times 10^{-8} °C^{-2}$. The inductance of such resonators can change from 0.03 H up to 1.5 H and the resonance spacing from 0.8% up to 1.9% depending on the frequency value and on the design of the resonator. The motional parameters and resonance spacing of these resonators depend on the angle of cut, geometrical size of a crystal element, configuration and dimensions of an electrode and also on their ratios. As a result of these investigations the optimum dimensional ratios and those angles of cut were established ensuring improved parameters and characteristics of contour mode resonators.

Optimization of shape and angle of cut of lithium tantalate crystal elements with xyls/$\pm\alpha$/$-\beta$ - orientation made it possible to miniaturize thickness

CH2965-2/91/0000-212 $1.00 © 1991 IEEE

shear mode resonators, to reduce up to 3-4 times the expenditure of costly raw materials, to improve their temperature stability over the frequency range 3 to 30 MHz. The resonance spacing of such resonators attains 6-8%, steepness coefficient of FTC parabola attains 9×10^{-8} °C^{-2} and their motional inductance $(0.1-1) \times 10^{-3}$ H [3, 4].

The investigations [5] made it possible to determine the optimum shape, electrode and crystal element dimensions of resonators operating in longitudinal mode over the frequency range 60 - 800 kHz. The maximum value of resonance spacing attainable by using "entire" electrodes with crystal elements having zyb/+45°-orientation is 2.5%. Reducing electrode length by 0.2-0.3 mm from the initial value enables to increase the resonance spacing of resonators approximately 1.5 times. The minimum inductance value of lithium tantalate resonators vibrating with longitudinal mode can attain 0.02-0.05 H.

Miniature resonators operating in flexural mode in the length-width plane (XY-plane) can be manufactured for operation over the frequency range of 25-250 kHz. Analogously to longitudinal mode resonators, flexural mode resonators, having crystal elements of zyb/+45°-orientation with electrodes along all the length have maximum resonance spacing of 1.3%. It can be enlarged approximately 1.2 times by corresponding shortening of driving electrodes. This is approximately 20-30 times superior to the analogous quartz crystal resonators (for example, of NT-cut). Minimum motional inductance values of such resonators achievable are equal to 0.1-0.2 H, and their frequency-temperature characteristics have the shape close to the quadratic parabola with the steepness coefficient $(7 - 9) \times 10^{-8}$ °C^{2} [6].

Lithium niobate resonators with strip piezoelectric elements of yzb/-17°-cut operating in thickness shear mode are of considerable practical interest. Their temperature instability over the temperature range from -60 to +85°C is relatively large and is approximately equal to $\pm 4000 \times 10^{-6}$. Compared to the resonance spacing value equal to 11% this instability constitutes 0.8% only. This enables a wide application of such resonators in oscillators of clock signals.

The parameters of realized miniature lithium tantalate and lithium niobate resonators are given in Tables 1 and 2.

Table 1. Parameters of manufactured miniature lithium tantalate resonators operating in different modes

Parameter designation	Modes of vibration			
	Flexural	Longitudinal	Contour	Thickness shear
Frequency range, kHz	30-250	60-600	450-900	3000-30000
Stepness coefficient of FTC, 10^{-8}°C^{-2}	7-9	7-9	6-9	9-11
Resonance spacing, %	0.5-1.5	1.0-3.1	0.8-1.8	6-8
Motional inductance, H	2-80	0.2-1.0	0.1-0.7	$(0.1-1.0)\times10^{-3}$
Motional resistance, Ohm	5-150	1-150	20-60	5-30
Quality factor Q, 10^3	30-100	50-200	30-120	0.7-2.5

Table 2. Parameters of a lithium niobate resonator manufactured

Frequency, kHz	Frequency-temperature coefficient,	Resonance spacing, %	Motional inductance, H	Motional resistance, Ohm	Quality factor, Q
4000	96-100	10-11	$(2-3)\times10^{-3}$	30-70	2-3

The values of recommended frequency ranges of lithium tantalate resonators with piezoelectric elements vibrating in different modes, and realizable values of resonance spacing of such resonators are given in Figure 1.

Among new promising piezoelectrics it is necessary to point out gallium silicate of lanthanum ($La_3Ga_5SiO_{14}$)(LGS) which had been first grown up and investigated in the U.S.S.R. [7].

LGS-crystals belong to the same symmetry group as α -quartz crystals, i.e. to the group 3:2. These crystals density is equal to $\rho = 5743$ kg/m^3, their acoustic quality, Q, is several times superior to that of a quartz crystal and their independent piezoelectric modules d_{11} and d_{14} are superior to those of quartz 2.7 and 6.3 times, respectively. These crystals exhibit orientations of temperature compensated cuts, ensuring the existence

of zero FTCs for flexural, longitudinal, contour, and thickness modes. Frequency-temperature characteristics of resonators have the form of quadratic parabolas (Fig. 2), the resonance spacing of such resonators is 3-5 times larger, and their inductance is several times smaller than of analogous quartz crystal units [8 , 9] .

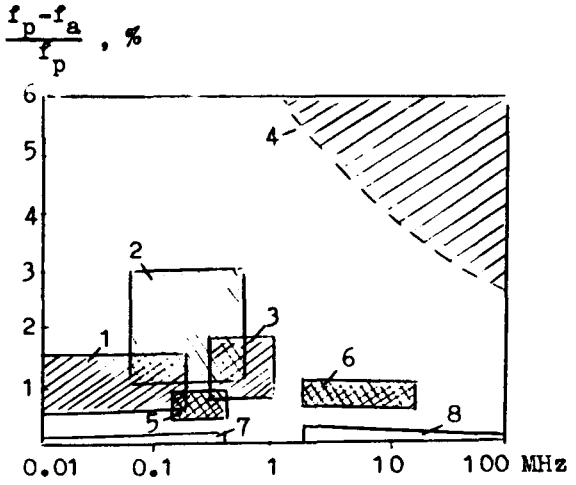


Fig. 1: Recommended frequency ranges of the resonator application: 1,2,3,4 - lithium-tantalate resonators operating in flexural, longitudinal, face-shear, thickness-shear modes, respectively 5,6 - $La_3Ga_5SiO_{14}$-resonators operating in longitudinal and thickness-shear modes, respectively 7,8 - quartz crystal resonators

LGS-crystals are simple in technological processing. They exhibit small anisotropy of linear expansion, do not have phase transitions of α - β quartz type, which makes it possible to use severe temperature and velocity-regimes during their processing, and to use resonators based on these crystals under high environmental temperatures, for example, as temperature gauges, because the melting temperature of LGS-crystals is approximately equal to 1470°C.

LGS-crystals are easily dissoluble in weak acid concentrations which significantly facilitates the process of planar photolithographic formbuilding of miniature piezoelectric elements and monolithic vibrators based on them.

The velocity of acoustic wave propagation in LGS-crystals is by 25-30% less than in quartz. This factor and also smaller inductance value compared to quartz crystal resonators provide smaller (4-5 times) LGS-resonator dimensions within low frequency range.

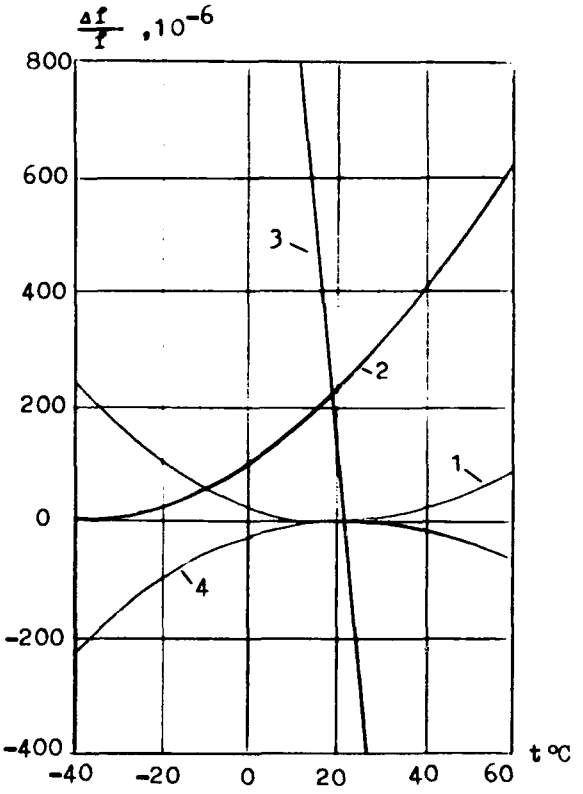


Fig. 2: Typical frequency-temperature characteristics of thickness-shear mode resonators:
1,2 - lithium tantalate; 3 - lithium niobate; 4 - $La_3Ga_5SiO_{14}$

An important feature of LGS-crystals is the possibility of isomorphic substitution of Ca, Nd, Ge and a number of other elements for La, Ga, Si [10] which enables to grow up crystals with different electrophysical characteristics and to manufacture of them resonators with necessary parameters.

All the above serves an indication of promising use of the new "strong" piezoelectric LGS in the production of piezoelectric resonators and of miniature ones especially.

Parameters of LGS-resonators manufactured are given in Table 3.

Table 3. Parameters of miniature resonators manufactures of LGS-crystals

Parameter designation	Mode of vibration	
	Longitudinal	Thickness shear
Frequency range, kHz	250-500	5000-20000
Steepness coefficient of FTS, $10^{-8}°C^{-2}$	5 - 8	4 - 6
Resonance spacing, %	0.5 - 0.7	0.7 - 1.0
Motional inductance, H	3 - 6	$(3\text{-}6)\times10^{-3}$
Motional resistance, Ohm	50 - 250	5 - 50
Quality factor Q, 10^3	20 - 30	10 - 30

Conclusion

The paper lists parameters of miniature resonators with strip piezoelectric elements made of lithium tantalate (thickness-shear mode) having high frequency stability and small dimensions of the crystal element.

The resonator operating in face-shear mode enable to improve parameters of lithium-tantalate resonators and to raise their economy within the frequency range of 500-2000 kHz.

The evidence of a worth-while wide production application of piezoelectric resonators made of langasite, a new piezoelectric material, seems to be proven.

The $La_3Ga_5SiO_{14}$-resonators possess satisfactory temperature versus frequency stability, have ecologically clean production technology and are promising for the application within the frequency range from 10 kHz to 50 MHz.

References

[1] Patent No.1216835 (U.S.S.R.), Applied 04.05.84 (In Russian).

[2] Kolesnik V.V., Kochetkov Y.A. et al. Characteristics of countour mode of lithium tantalate piezoelectric resonators, Kristallographia, v.31, issue 4, p.710-714, 1986.(In Russian).

[3] Grouzinenko V.B., Bezdelkin V.V. at al. Microminiature high-frequency lithium tantalate resonators, III All-Union Conference: Actual problem of receiving and application of ferroelectric and piezoelectric materials, Moscow, p.126, 1987. (In Russian).

[4] Dsiba S.I., Grouzinenko V.B. et al. Miniature lithium tantalate piezoelectric resonator, Electronic Engineering Series Radiodetails and Components, issue 4(69), p.57-58, 1987. (In Russian).

[5] Kochetkov Y.A., Jaroslavsky M.I. et al. Piezoelectric resonators with crystal elements of lithium tantalate operating in longitudinal modes, Kristallographia, v.30, issue 5, p.1039, 1985. (In Russian).

[6] Grouzinenko V.B., Kochetkov Y.A. et al. Lithium tantalate resonators with flexural modes in XY-plane, Acoustical Journal, v.XXXI, No.3, p.393-395,1985.(In Russian).

[7] Kaminsky A.A., Mill B.V. et al. Nonlinear-active material $(La_{1-x}Nd_x)_3Ga_5SiO_{14}$, Izv.Akad.Nauk. SSSR, Ser. Physical, v.47, No.10, p.1903-1909, 1983. (In Russian).

[8] Bezdelkin V.V., Grouzinenko V.B. et al. Application of galling silicate of lanthan crystals in piezoelectric BAW resonators, XIV-Union Conference on Acoustic Electronics and of Solid-State Physical Acoustics, Kishinev, part 2, p.85-86, 1989. (In Russian).

[9] Andreev I.A., Doubovik M.F. New piezoelectric "langasite" $La_3Ga_5SiO_{14}$-material with zero frequency-temperature coefficient of elastic vibrations. Letters to Journal of Theoretical Physics, v.10, No.8, p.1385, 1984.

[10] Mill B.V., Butashina A.V. et al. Modified gallates with the structure of $Ca_3Ga_2GeO_{14}$, *Reports of the USSR Academy of Science*, v.264, No.6, p.1385, 1982. (In Russian).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A COMPARISON OF PREDICTED AND MEASURED PROPERTIES OF DOUBLY-ROTATED LITHIUM TETRABORATE RESONATORS

A. Ballato, J. Kosinski*, S. Mallikarjun**, and T. Lukaszek*

US ARMY LABCOM, ATTN: SLCET-DS, Fort Monmouth, NJ 07703-5601

*US ARMY LABCOM, ATTN: SLCET-MA, Fort Monmouth, NJ 07703-5601

**Department of Chemistry and Physics, Monmouth College, West Long Branch, NJ 07764

Abstract

Recently, the authors have presented calculations of the properties expected for doubly-rotated lithium tetraborate resonators. The loci of cuts possessing both first and second order temperature compensation for the simple thickness modes of plate resonators driven by thickness-directed electric fields were identified. Lithium tetraborate plates cut to the nominal orientations of these loci have since been obtained and measurements have been undertaken to verify the predictions. The mode spectra, equivalent circuit parameters, and frequency-temperature behavior have been examined for the first, third, and fifth harmonics for all three simple thickness modes of the resonators. The observed values of frequency constants and piezocoupling are in good agreement with the predicted values. The measured temperature coefficients are less well in agreement with the predicted values.

Introduction

In a recent series of papers [1-5], the authors have presented calculations of the properties expected for doubly-rotated lithium tetraborate resonators. The calculations employed the one-dimensional approximation [6] and examined both thickness excitation (TE) and lateral excitation (LE) of thickness modes (TM) in plate resonators. Of primary interest were the wave velocities, piezocoupling, and temperature coefficients of frequency. The data of Shiosaki et al. [7] as listed in Table 1 were used in the calculations with one exception; the thermal expansion coefficients were used to calculate the temperature coefficients of density, rather than using the published values.

In carrying out the calculations over the entire primitive region of the material, the loci of cuts possessing both first and second order temperature compensation for TETM plate resonators were identified. One cut, employing the extensional or 'a'-mode, with nominal angles of cut (YXwl)$40^{\circ}/33^{\circ}$ has been given the designation of TA-cut. This cut is proposed for applications when frequency stability with temperature over an extended range is a requirement, jointly with moderate piezoelectric coupling. The other cut, employing the slow shear or 'c'-mode, with nominal angles of cut (YXwl)$19^{\circ}/56^{\circ}$ has been designated as the TC-cut. This cut is proposed for applications demanding superior frequency stability with temperature, and piezocoupling values larger than available with quartz. In both cases, first- and second-order temperature compensation is predicted only for the fundamental harmonic when driven by thickness excitation.

Experimental Samples

In order to verify the predicted behaviors, a set of lithium tetraborate plates cut to the nominal orientations of the TA- and TC-cuts was obtained from a

TABLE 1. $Li_2B_4O_7$ MATERIAL CONSTANTS @ 25°C

Item	Value	$T^{(1)}$ $10^{-6}/K$	$T^{(2)}$ $10^{-9}/K^2$
ϵ^S_{11}	78.8×10^{-12} F/m	97.1	2800
ϵ^S_{33}	71.5×10^{-12} F/m	545	2900
e_{15}	0.472 C/m^2	-1050	-900
e_{31}	0.290 C/m^2	-573	-6910
e_{33}	0.928 C/m^2	-600	-6500
c^E_{11}	$135.\times10^9$ N/m^2	-81	-440
c^E_{12}	3.57×10^9 N/m^2	3370	-17400
c^E_{13}	33.5×10^9 N/m^2	465	-2300
c^E_{33}	56.8×10^9 N/m^2	364	-1800
c^E_{44}	58.5×10^9 N/m^2	-18.1	500
c^E_{66}	46.7×10^9 N/m^2	-272	-450
α_{11}	----------	11.1	5.6
α_{33}	----------	-3.74	21
ρ	2439 kg/m^3	-2.01	-31.8

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

material grower [8]. The plates are of plano-plano goemetry, 14mm diameter, 168.4 μm thick, with both sides optically polished. The plates were cleaned using organic solvents and a UV/ozone cleaner, then electroded, mounted, and sealed in two-point mount HC-6 type cold-weld enclosures. A nominal mass loading of 0.48% was obtained with 300Å Cr/1200Å Au electrodes of 5.34 mm diameter.

Mode Spectra

The mode spectra of the resonators were examined using an HP8753B Network Analyzer/HP85046A S-Parameter Test Set in conjunction with HP85160A Measurement Automation Software [9]. Figure 1 shows the observed mode spectrum of the TA-cut and Figure 2 shows the mode spectrum of the TC-cut. The observed frequencies of the various modes and their harmonics are listed in Tables 2 and 3 for the TA- and TC-cut respectively, along with certain other quantities of interest. The observed frequencies are in good agreement with the predictions, as are the observed values of piezocoupling. Note that for the TA-cut b-mode, a Q·f product of $2\text{-}3\text{x}10^{13}$ is observed. This high value is consistent with other lithium-based

TABLE 2. TA-CUT OBSERVED FREQUENCIES

Mode	M	f_R [MHz]	Q [10^3]	τ_1 [fs]	k [%]	b
a	1	21.041	1	6912	13	0.42
a	3	63.860	21	118	10	0.24
a	5	106.490	32	47	9.5	0.22
b	1	15.154	350	30	1.4	0.88
b	3	45.419	450	8	1.2	0.68
b	5	76.017	147	14	1.7	--
c	1	10.718	5	2884	17	0.69
c	3	32.537	21	234	15	0.47
c	5	54.284	37	82	13	0.41

TABLE 3. TC-CUT OBSERVED FREQUENCIES

Mode	M	f_R [MHz]	Q [10^3]	τ_1 [fs]	k [%]	b
a	1	20.095	2	4449	19	0.49
a	3	61.346	11	246	15	0.33
a	5	102.365	23	68	12	0.22
b	1	14.275	180	62	2.3	0.86
b	3	42.805	163	23	1.9	0.58
b	5	71.298	140	19	0.5	0.11
c	1	9.408	3	5524	25.5	0.89
c	3	29.017	12	456	20.3	0.50
c	5	48.400	50	64	8.4	0.09

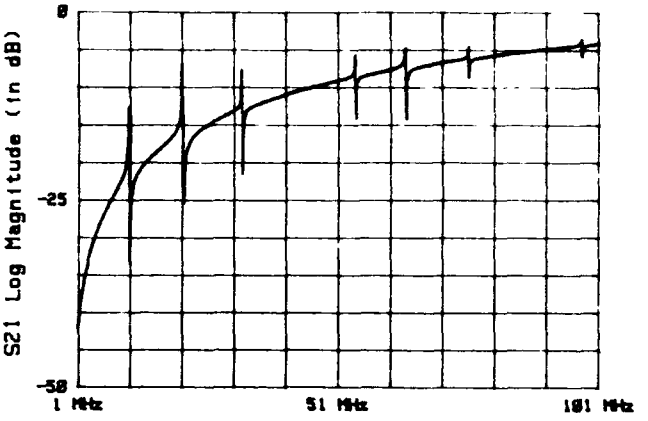


Figure 1. Mode spectrum of the $Li_2B_4O_7$ TA-cut resonator driven by thickness excitation.

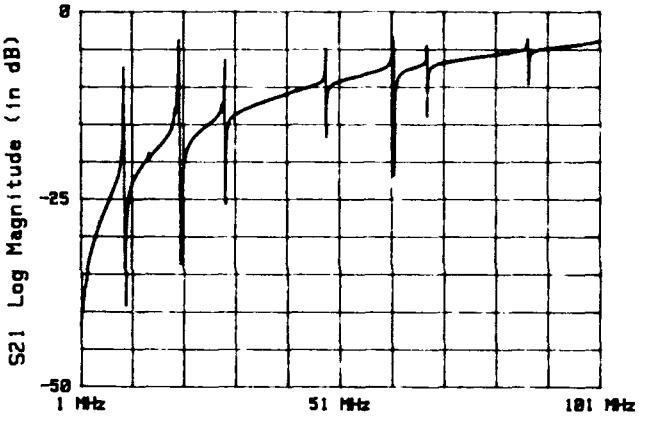


Figure 2. Mode spectrum of the $Li_2B_4O_7$ TC-cut resonator driven by thickness excitation.

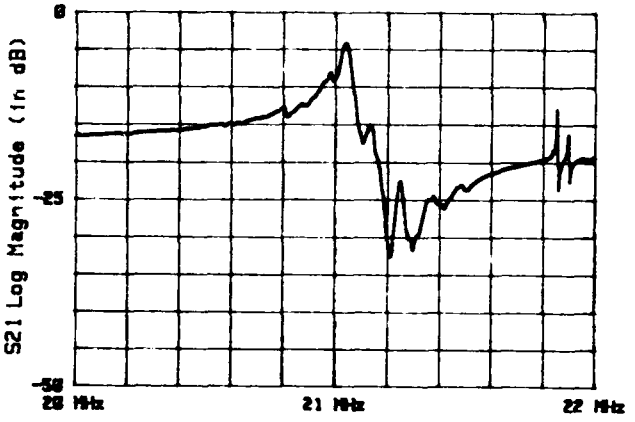


Figure 3. Measured response of the fundamental harmonic of the $Li_2B_4O_7$ TA-cut 'a'-mode.

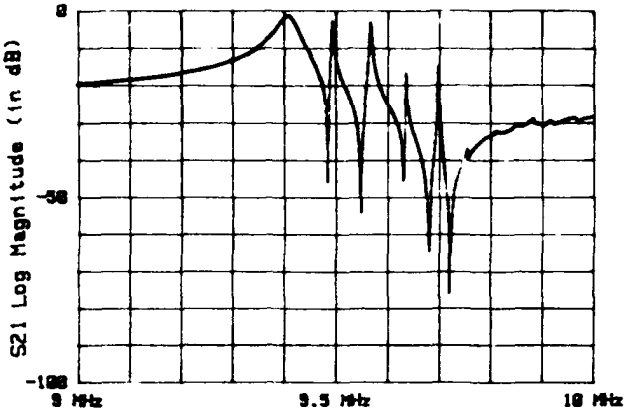


Figure 4. Measured response of the fundamental harmonic of the $Li_2B_4O_7$ TC-cut 'c'-mode.

piezoelectrics such as $LiNbO_3$ and $LiTaO_3$. The motional time constant τ_1 is equal to $1/(2\pi f_R Q)$.

Figures 3 and 4 show respective views of the TA-cut fundamental 'a'-mode and TC-cut fundamental 'c'-mode in the vicinity of resonance. The presence of unwanted modes is noted in both cases. For the units tested here, the fundamental harmonic of any thickness mode was least afflicted by unwanted modes, while the harmonics suffered progressively increasing interference from unwanted modes. The implied loss of energy trapping with increasing harmonic is reflected in the one-dimensional electroded fraction of the active area denoted as b in Tables 2 and 3 [10-12].

Equivalent Circuit Parameters

The equivalent electrical circuit parameters of the resonators have been examined for the fundamental, third, and fifth harmonics for all three TE modes of the resonators. The measurements were made using two techniques:

1) HP4191A RF Impedance Analyzer/ FLUKE 6060A Synthesizer with PTI developed software [13],

2) HP8753B Network Analyzer/HP85046A S-Parameter Test Set with HP85165A Resonator Measurement Software (implements EIA-512) [14].

Some difficulty was experienced in the measurement process due to the previously noted interference from unwanted modes.

The results of the equivalent circuit parameter measurements are summarized in Tables 4 and 5 for the TA- and TC-cuts, respectively.

TABLE 4. TA-CUT EQUIVALENT ELECTRICAL CIRCUIT PARAMETERS

Mode	M	$R_1[\Omega]$	L_1[mH]	C_1[fF]	Co[pF]
a	1	39.5	0.325	175	12.07
a	3	10.7	0.565	11.0	11.87
a	5	13.0	0.625	3.6	12.27
b	1	16	60	1.85	12.41
b	3	50	78	0.16	11.80
b	5	124	38	0.11	11.62
c	1	10.3	0.80	280	11.76
c	3	10.9	1.1	21.5	11.21
c	5	12	1.27	6.8	11.93

TABLE 5. TC-CUT EQUIVALENT ELECTRICAL CIRCUIT PARAMETERS

Mode	M	$R_1[\Omega]$	L_1[mH]	C_1[fF]	Co[pF]
a	1	13.4	0.19	332	11.30
a	3	9.9	0.27	24.8	12.08
a	5	11.3	0.40	6.05	12.08
b	1	12.0	24	5.2	11.73
b	3	58.6	35.4	0.39	11.61
b	5	615	190	0.03	11.29
c	1	8.7	0.46	635	12.07
c	3	11.4	0.74	40	10.93
c	5	24	4.0	2.65	11.49

Temperature Behavior

The frequency-temperature behavior of the zero-temperature coefficient cuts was measured using an IEC-444 pi-network transmission measurement system which has been described previously [15], and is currently being upgraded. The upgrade includes the incorporation of an HP8508A Vector Voltmeter which was used in conjunction with a modified version of the Stimulus-Response Software listed in HP Product Note 8508-2 [16]. Due to the presence of the unwanted modes, it was necessary to measure the frequency spectrum as a function of temperature and extract the data for each particular mode from the measured spectra.

Figures 5 and 6 show respectively the frequency-temperature behavior of the predicted zero-temperature coefficient modes of the TA- and TC-cuts. Although the measured first-order temperature coefficients show reasonable agreement with the predicted values, the second-order temperature coefficients are less well in agreement with the predicted values. The predicted and measured temperature coefficients of frequency of all the modes and harmonics examined here are summarized in Tables 6 and 7 for the TA- and TC-cuts respectively.

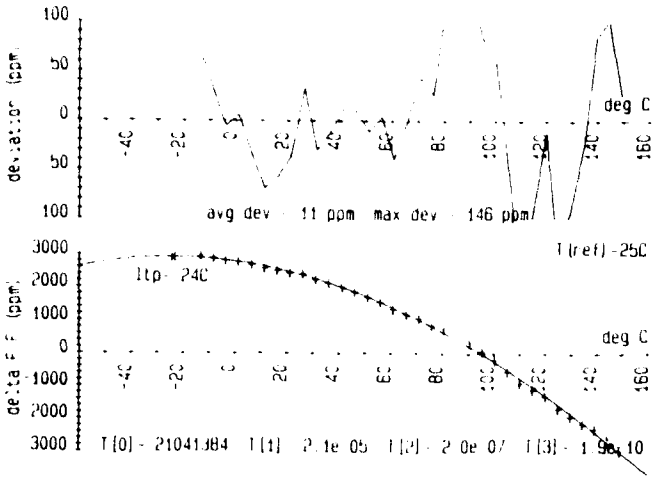


Figure 5. TA-cut 'a'-mode fundamental harmonic frequency-temperature behavior.

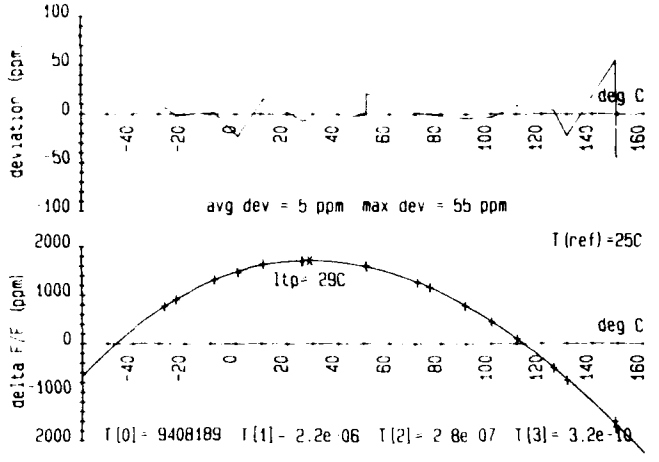


Figure 6. TC-cut 'c'-mode fundamental harmonic frequency-temperature behavior.

TABLE 6. TA-CUT PREDICTED AND MEASURED FREQUENCY-TEMPERATURE BEHAVIOR

		Predicted		Measured			
Mode	M	$T_f^{(1)}$	$T_f^{(2)}$	$T_f^{(1)}$	$T_f^{(2)}$	$T_f^{(3)}$	T_i
a	1	+1	-26	-21	-202	+186	+387
a	3	-26	-151	-28	-67	+48	+501
a	5	-30	-166	-29	-66	+45	+511
b	1	-64	+472	-76	-49	--	---
b	3	-64	+475	-8	-5	--	---
b	5	-64	+476	-14	+5	--	---
c	1	-80	+219	-11	-4	--	---
c	3	-107	+113	-14	+4	-26	+83
c	5	-111	+100	-14	+5	-10	+199

Units: $T_f^{(1)}[\times 10^{-6}/K]$, $T_f^{(2)}[\times 10^{-9}/K^2]$, $T_f^{(3)}[\times 10^{-12}/K^3]$, $T_i[^\circ C]$

TABLE 7. TC-CUT PREDICTED AND MEASURED FREQUENCY-TEMPERATURE BEHAVIOR

		Predicted		Measured			
Mode	M	$T_f^{(1)}$	$T_f^{(2)}$	$T_f^{(1)}$	$T_f^{(2)}$	$T_f^{(3)}$	T_i
a	1	+39	-35	+46	-234	+461	+195
a	3	-10	-337	-2	-140	+221	+236
a	5	-15	-372	-5	-129	+175	+270
b	1	-31	+217	-47	-86	+216	+157
b	3	-31	+219	-47	-83	+109	+277
b	5	-31	+219	-47	-85	+100	+307
c	1	-3	+25	+2	-277	+324	+309
c	3	-47	-242	-58	-167	+235	+262
c	5	-52	-272	-63	-172	+256	+249

Units: $T_f^{(1)}[\times 10^{-6}/K]$, $T_f^{(2)}[\times 10^{-9}/K^2]$, $T_f^{(3)}[\times 10^{-12}/K^3]$, $T_i[^\circ C]$

Conclusions

The predicted values of the first-order temperature coefficients of frequency $T_f^{(1)}$ show reasonable agreement with the measured values, but the predicted values of the second-order temperature coefficients of frequency $T_f^{(2)}$ do not agree well with the measured values. Although $Li_2B_4O_7$ shows promise as a piezoelectric material exhibiting the desireable properties of high Q, high coupling, and zero-temperature behavior, the design of resonant devices will require a substantially improved knowledge of the basic material constants.

Acknowledgment

The authors wish to thank INRAD, Inc., for generously supplying the accurately oriented and optically polished samples, and the Frequency Control and Timing Branch of ETDL for processing the plates into resonators.

References

[01] A. Ballato, J. Kosinski, M. Mizan, and T. Lukaszek, "Lateral- and Thickness-Field Coupling in Lithium Tetraborate," in Proceedings of the 43rd Annual Symposium on Frequency Control, May-June 1989, pp. 557-562.

[02] A. Ballato, J. Kosinski, and T. Lukaszek, "Lateral-Field Temperature Behavior of Lithium Tetraborate," IEEE Ultrasonics Symposium Proceedings, Montréal, Canada, October 1989, pp. 441-444.

[03] A. Ballato, J. Kosinski, M. Mizan,

and T. Lukaszek, "Temperature and Coupling Behavior of Resonators and Transducers of Lithium Tetraborate Driven by Lateral- and Thickness Fields," Technical Report SLCET-TR-90-1, US Army Laboratory Command, Fort Monmouth, NJ, January 1990, 88 pp.

[04] A. Ballato, J. Kosinski, and T. Lukaszek, "Lithium Tetraborate Transducer Cuts," Technical Report SLCET-TR-90-4, US Army Laboratory Command, Fort Monmouth, NJ, March 1990, 18 pp.

[05] A. Ballato, J. Kosinski, and T. Lukaszek, "Lateral-Field Temperature Behavior of Lithium Tetraborate," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 38, No. 1, January 1991, pp. 62-66.

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

[07] T. Shiosaki, M. Adachi, and A. Kawabata, "Growth and Properties of Piezoelectric Lithium Tetraborate Crystal for BAW and SAW Devices," IEEE International Symposium on Applications of Ferroelectrics (ISAF) Proceedings, Lehigh University, Bethlehem, PA, June 1986, pp. 455-464.

[08] INRAD, Inc., 181 Legrand Avenue, Northvale, NJ 07647, (908) 767-1910.

[09] HP 85160A Measurement Automation Software Rev. A.02.00, Hewlett-Packard Company, 3000 Hanover Street, Palo Alto, CA 94304, 1989.

[10] J. Kosinski, S. Mallikarjun, and A. Ballato, "Mass Loading Measurements of Quartz Crystal Plates," in Proceedings of the 43rd Annual Symposium on Frequency Control, May-June 1989, pp. 365-371.

[11] J. Kosinski, A. Ballato, and S. Mallikarjun, "Quartz Crystal Resonator Mass Loading," Technical Report SLCET-TR-89-7, US Army Laboratory Command, Fort Monmouth, NJ, August 1989, 24 pp.

[12] M. Schmid, E. Benes, W. Burger, and V. Kravchenko, "Motional Capacitance of Layered Piezoelectric Thickness-Mode Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 38, No. 3, May 1991, pp. 199-206.

[13] R. C. Smythe, "Crystal Parameter Measurement Methods," Technical Report DELET-TR-82-0390-F, US Army Electronics Research and Development Command, Fort Monmouth, NJ, January 1984, 130 pp.

[14] HP 85165A Resonator Measurement Software Rev. A.02.02, Hewlett-Packard Company, 1400 Fountaingrove Parkway, Santa Rosa, CA 95403, 1990.

[15] J. Kosinski, A. Ballato, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Temperature Derivatives of the Dynamic Permittivity and Permeability of the Simple Thickness Modes of Quartz Plates," in Proceedings of the 42nd Annual Symposium on Frequency Control, June 1988, pp. 53-64.

[16] "HP8508A Vector Voltmeter Automatic Stimulus Response Measurements" (Product Note 8508-2), Hewlett-Packard Company, South Queensferry, West Lothian, Scotland EH30 9TG , March 1988.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN ANALYSIS OF THE INFLUENCE OF DESIGN PARAMETERS ON THE RESONANT FREQUENCY AND Q-FACTOR OF SURFACE TRANSVERSE WAVE (STW) RESONATORS

E. Bigler*, B.A. Auld**, E. Ritz*** and E. Sang***

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32,avenue de l'Observatoire - 25000 Besançon - France
** Ginzton Laboratory, Stanford University, Stanford CA 94305 - USA
*** Tektronix, SAW Technology, MS 50-240, P.O. Box 500, Beaverton, OR 97077 - USA

Abstract

Surface transverse wave (STW) resonators offer several advantages over conventional Rayleigh wave (SAW) resonators (higher operating frequencies, better temperature stability) but they are more sensitive to variations of design parameters. Geometrical dimensions, metal thickness or groove depth have to be carefully selected and controlled to ensure low loss and high Q operation of a STW resonator.

A model is presented for predicting the resonant frequency of STW resonators operating in the fundamental mode, as a function of different design parameters. The model is based on the near-stopband approximation and the formalism of cascaded transfer matrices. Results are obtained in the case of a 3-grating configuration (mirror-cavity-mirror) for both grooved and metal strip gratings. Based on simple analytical expressions, a design procedure is used and the influence of various parameters is demonstrated and compared to already published experimental data.

New experimental results obtained on metal strip quartz STW resonators are presented and discussed.

1. Introduction

Since Surface Transverse Waves (STW) were introduced [1], several examples of STW devices (delay lines and resonators) have been fabricated and tested [2-4]. Potential advantages of STW resonators over conventional SAW resonators have been pointed out for high stability oscillator applications : higher operating frequencies (+ 60 %) for same geometries, better temperature stability, better performances in terms of phase noise and aging [4].

Theory of STW devices can now be considered as well established [2, 5-9] but since only few experimental results have been reported on fabrication and test of STW resonators, our knowledge of the influence of design parameters on fundamental properties (resonant frequency, Q-factor, insertion losses, ...) is not as good as for conventional SAW resonators. Only recently was presented in an experimental work by Avramov [10] a systematic investigation of the influence of metal thickness on fundamental properties of metal strip STW resonators.

The scope of this paper is to present a model predicting the influence of various design parameters : geometrical dimensions, metal thickness or groove depth, on the resonant frequency of STW resonators. The model is based on the theory [2, 5, 6] of STW propagation at the surface of an anisotropic substrate covered with a periodic array of grooves or metal strips. The aim is here to keep analytical expressions as far as possible in the analysis, in order to be able to understand as directly as possible the influence of different factors on the resonant frequency.

A non-piezoelectric model will be applied to singly-rotated quartz plates, together with the near stopband approximation of coupled mode theory [2, 5-7, 11] describing wave propagation under a grating. This approximation is valid for small values of h/Λ up to a few percent (where h denotes grooved depth and Λ is the periodicity of the grating)corresponding to most experimental cases already published.

Another advantage of analytical expressions is to simplify the computer program used to find the resonant frequency, then allows an easy investigation the influence of all design parameters by varying the input data. Results will be presented in the case of the 3-grating configuration (mirror-cavity-mirror) for both cases of metal strips and grooved gratings. The influence of cavity length, metal thickness, grating periodicities is demonstrated, and predicted results are compared with already published experimental data [4, 10]. New experimental results obtained on (37°-Y-rotated) quartz STW metal strip resonators are presented and show the importance of the relative position of the resonant frequency within the mirror stopband to ensure low loss and high Q.

2. Modelling STW propagation for both grooves and metal strip gratings

a) mode shape and coupled wave equations

Following [2, 5-6], Floquet expansion of a STW mode, expressed in a set of singly-rotated (X,Y,Z) axes (see Fig. 1) is given by

$$u_x = \sum_n a_n e^{-\chi_n y} e^{-i\beta_n z} e^{i\omega t} \qquad (1)$$

CH2965-2/91/0000-222 $1.00 © 1991 IEEE

where u_x is the transverse component of mechanical displacement
a_n, χ_n, β_n are respectively the amplitude, penetration factor and propagation constant of the n-th Floquet mode
β_n are related by

$$\beta_n = \beta_o + \frac{2n\pi}{\Lambda} \tag{2}$$

where Λ is the periodicity of the array.

Following convention, determining signs and phase factors in all derivations should be noted on Fig. 1

- the origin of Z-axis is taken on the left-hand side of a strip
- +Y direction points towards the inside of the crystal.

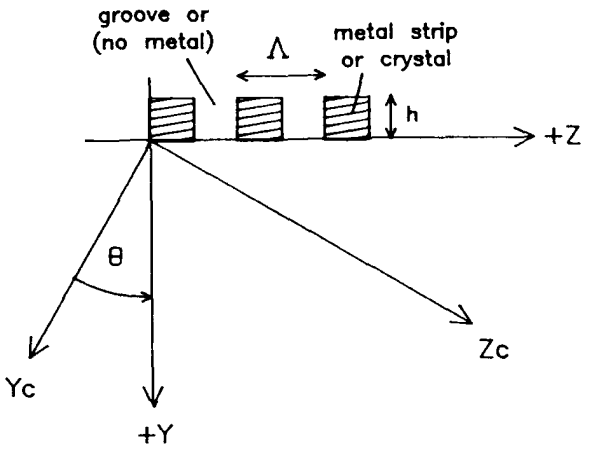


Fig. 1 : STW propagation under an array of grooves or metal strips
h = grooved depth / strip height
Λ = periodicity of the array
$X_c\, Y_c\, Z_c$ crystallographic axes
X = X_c, Y, Z singly-rotated axes.

In the near-stopband approximation, only two Floquet modes $n = 0$ and $n = -1$ are considered in the expansion (1). Substituting (1) into the propagation equation and using 1st-order Taylor expansions of ω and β_0, β_{-1} yields [6, 12]

$$\delta\beta = \beta_o - \frac{\pi}{\Lambda} = \beta_{-1} + \frac{\pi}{\Lambda} \tag{3}$$

$$\delta\omega = \omega - \frac{\pi V_s}{\Lambda} \tag{4}$$

$$\chi_o = -i\,\frac{C_{56}}{C_{66}}\left(\frac{\pi}{\Lambda} + \delta\beta\right) + \frac{2\pi\rho V_s^2}{C_{66}}\left(\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2} \tag{5}$$

$$\chi_{-1} = -i\,\frac{C_{56}}{C_{66}}\left(-\frac{\pi}{\Lambda} + \delta\beta\right) + \frac{2\pi\rho V_s^2}{C_{66}}\left(-\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2} \tag{6}$$

where C_{55} and C_{66} = rotated elastic constants of the crystal expressed in (XYZ) axes
ρ = mass density of the crystal
V_s = velocity of SSBW mode.

In the nonpiezoelectric approximation, V_s is given by

$$V_s = \frac{C_{eff}}{\rho} \tag{7}$$

where C_{eff} is a combination of elastic constants given by

$$C_{eff} = \frac{C_{55}C_{66} - C_{56}^2}{C_{66}} \tag{8}$$

In Eqs. (5) and (6), a physically acceptable solution should correspond to a decay in the +Y direction ; for this reason when square roots become complex, they should be taken with a positive real part, to ensure a positive real part of both χ_0 and χ_{-1}.

The problem is then completely solved by the determination of mode amplitudes a_0 and a_{-1} and dispersion relation connecting $\delta\omega$ with $\delta\beta$; this is achieved using 1st order Datta-Hunsinger surface boundary conditions [13], and the following coupled modes equations are obtained [2, 5, 12] valid for both cases of metal strips or grooves.

First Floquet mode a

$$D_1 a_o = iq\sqrt{K}\, a_{-1} \tag{9}$$

$$D_2 a_{-1} = -iq\sqrt{K}\, a_o \tag{10}$$

Second Floquet mode b

$$D_2 b_o = iq\sqrt{K}\, b_{-1} \tag{11}$$

$$D_1 b_{-1} = -iq\sqrt{K}\, b_o \tag{12}$$

where

$$D_1 = \left(\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2} - q_o \tag{13}$$

$$D_2 = \left(-\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2} - q_o \tag{14}$$

$$q_o = \frac{1}{\sqrt{2}}\,\frac{\pi}{2}\,\frac{h}{\Lambda}\left(\frac{\pi}{\Lambda}\,\frac{C_{eff}}{C_{66}}\right)^{1/2}\frac{\rho'}{\rho}\left(1 - \left(\frac{V_m}{V_s}\right)^2\right) \tag{15}$$

$$q = \frac{1}{2}\,\frac{\rho'}{\rho}\left(1 - \left(\frac{V_m}{V_s}\right)^2\right) \tag{16}$$

$$K = \frac{\pi}{\Lambda}\left(\frac{h}{\Lambda}\right)^2\frac{C_{eff}}{C_{66}} \tag{17}$$

In Eqs. (15) and (16) appear the only constants dependent on the nature of the elements in the array, i.e.

h = groove depth / strip height
ρ' = mass density of metal ($\rho' = \rho$ for grooves)
V_m = velocity of shear waves inside the metal ($V_m = V_s$ for grooves).

Equations (9-17) are simplified expressions, valid near the stopband, assuming that $\delta\omega/nV_s \ll 1$ and $\delta\beta\Lambda/n \ll 1$.

Within these approximations, the case of grooved gratings appears to be a limit case of metal strip gratings with $q_0 = 0$ and $q = 1$.

b) Analytical solution of dispersion equation

From eqs. (9-12) the dispersion relation can be derived as

$$D_1 D_2 = K q^2 \tag{18}$$

The limit case of grooved gratings is obtained with $q_0 = 0, q = 1$, i.e.

$$\left(\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2}\left(-\delta\beta - \frac{\delta\omega}{V_s}\right)^{1/2} = K \tag{19}$$

Eq. (19) is nothing but the well-known dispersion equation of STW under grooved gratings as obtained in [2].

Eq. (18) may be solved analytically introducing dimensionless variables r, s, r_0

$$r = \frac{\delta\beta}{Kq^2} \tag{20}$$

$$s = \frac{\delta\omega / V_s}{Kq^2} \tag{21}$$

$$r_o = \frac{q_o}{q\sqrt{K}} \tag{22}$$

Combining (18) with (20-22) yields a dimensionless dispersion relation

$$\left((r-s)^{1/2} - r_o\right)\left((-r-s)^{1/2} - r_o\right) = 1 \tag{23}$$

Introducing a branch cut on the negative part of the real axis to define complex square roots, Eq. (23) may be expressed in parametric form

$$(r-s)^{1/2} - r_o = \rho\, e^{i\theta} \tag{24}$$

$$(-r-s)^{1/2} - r_o = \frac{1}{\rho}\, e^{-i\theta} \tag{25}$$

where $\rho\, e^{i\theta}$ is a complex number.

Squaring both sides of eqs. (24) and (25) yield the following equation

$$\sin\theta\left(\rho - \frac{1}{\rho}\right)(r_o + \cos\theta\left(\rho + \frac{1}{\rho}\right) = 0 \tag{26}$$

Different solutions of (26) correspond to different branches of the dispersion relation
1) $\theta = 0$ corresponds to a passband with $\delta\beta$ real < 0
2) $\theta = n$ is excluded by our choice of branch-cut
3) $\rho = 1$ corresponds to a stopband
4) $r_0 + \cos\theta\,(\rho + 1/\rho) = 0$ corresponds to the upper part of passband

Case 1 - passband with $\delta\beta < 0$ corresponds to the case of the cavity of a STW resonator, with un-attenuated propagation along the surface (z+), and energy trapping. Dispersion curve has an hyperbolic shape, but an explicit relation between $\delta\omega$ and $\delta\beta$ has to be numerically computed, except in the case of grooved gratings.

Case 3 - corresponds to stopband and may be completely solved ; with $\rho = 1$, after squaring eqs. (24) and (25) yield

$$\frac{\delta\beta}{Kq^2} = r = i\,(2\,r_o \sin\theta + 2\sin\theta\cos\theta) \tag{27}$$

$$\frac{\delta\omega}{V_s\,Kq^2} = s = -\,(r_o^2 + 2\,r_o\sin\theta + \cos 2\theta) \tag{28}$$

Eq. (27) shows that inside stopband, $\delta\beta$ is purely imaginary, and complex square roots of eqs. (24-25) will have both positive real parts in the lower half of stopband only.

Assuming that

$$\delta\beta = -i\sigma \tag{29}$$

where σ is real and positive, the analytical form of dispersion equation takes the unexpected form of a classical "limaçon" curve

$$\left|\left(\frac{\sigma}{Kq^2}\right)^2 + \left(\frac{\delta\omega}{V_s\,Kq^2} + r_o^2 - 1\right)^2\right|^{1/2} = 2\,(r_o + \cos\theta) \tag{30}$$

In the case of grooved gratings, $q = 1$, $r_0 = 0$, the limaçon curve degenerates as expected into the limit shape of an ellipse [2].

Case 4 - A discussion not reproduced here shows that complex square roots $(\delta\beta\text{-}\delta\omega/V_s)^{1/2}$ and $(-\delta\beta\text{-}\delta\omega/V_s)^{1/2}$ cannot exhibit a real part both at the same time ; thus one of the penetration constants χ_0 and χ_1 of Eqs. (5-6) has a negative real part, corresponding to energy radiation inside the substrate.

Figure 2 summarizes the shapes of different branches of the dispersion relation obtained here.

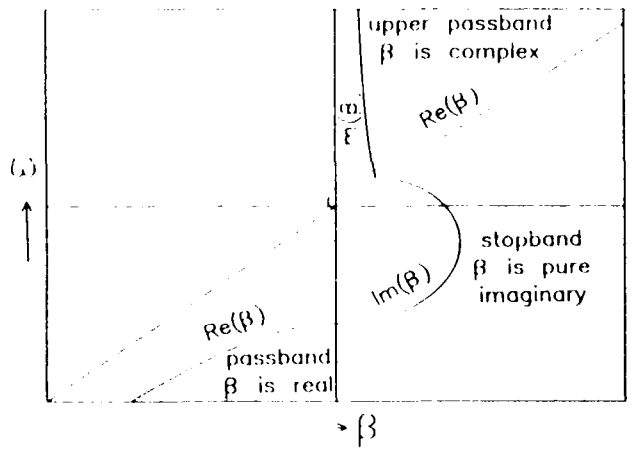


Fig. 2 : Dispersion curves for metal strip gratings in the near stopband approximation

c) Transfer matrices for a stack of grooves or metal strips

From coupled equations (9-12), following [2, 11, 14] the transfer matrix connecting the amplitudes of forward and backward travelling waves (Fig. 3a) can be obtained for a given stack of grooves, or metal strips, starting at a distance d of the origin, with a length L (Fig. 3b).

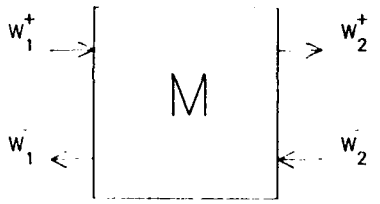


Fig. 3a : Definition of transfer matrix M in coupled wave theory with two Floquet modes

$$\begin{pmatrix} w_1^+ \\ w_1^- \end{pmatrix} = (M) \begin{pmatrix} w_2^+ \\ w_2^- \end{pmatrix} \tag{31}$$

with

$$(M) = \begin{vmatrix} e^{+iLw/\Lambda}\, a' & e^{-iLw/\Lambda}\, c' \\ e^{+iLw/\Lambda}\, b' & e^{-iLw/\Lambda}\, d' \end{vmatrix} \tag{32}$$

where

$$a' = \frac{1}{D_2 - D_1} \left(D_2\, e^{i\delta\beta L} - D_1\, e^{-i\delta\beta L} \right) \tag{33}$$

$$b' = -\exp(2i\pi d/\Lambda) \left(e^{i\delta\beta L} - e^{-i\delta\beta L} \right) \frac{i\sqrt{K}}{D_2 - D_1} \tag{34}$$

$$c' = b'^* \tag{35}$$

$$d' = a'^* \tag{36}$$

Symmetry relations (35) and (36) are imposed by energy conservation.

Particular case : expressions inside stopband

From eqs. (27-30) coefficients a' and b' inside stopband can be expressed in a parametric form, easily programmable

$$\cos\theta = -\frac{q_o}{2q\sqrt{K}} + \left(\frac{1 - \frac{\delta\omega}{V_s} / Kq^2}{2} - \frac{q_o^2}{4Kq^2} \right) \tag{37}$$

$$a' = \cosh(\sigma L) + i \frac{\sinh \sigma L}{\operatorname{tg}\theta} \tag{38}$$

$$b' = e^{2i\pi d/\Lambda} \frac{\sinh \sigma L}{\sin\theta} \tag{39}$$

$$\delta\beta = -i\sigma \; ; \; \sigma = -2Kq^2 \sin\theta \left(\frac{q_o^2}{q\sqrt{K}} + \cos\theta \right) \tag{40}$$

Eqs. (38) and (39) are the extension to the case of metal strip gratings of similar expressions obtained for grooves in [2]. Note a change of sign between this work and [2], due to a change in the convention of orientation for + Y direction.

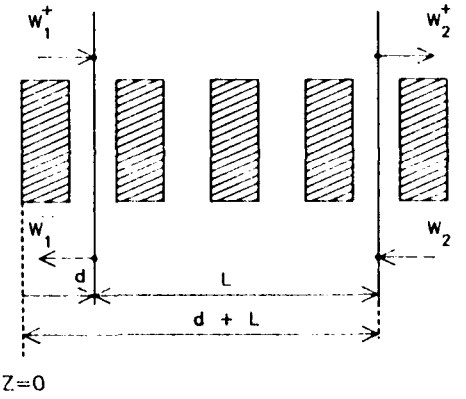


Fig. 3b : Stack of grooves or metal strips starting at a distance d from the origin, of length L

3. Condition of resonance of a 3-grating structure

a) General expression

Consider a STW resonator with a symmetric structure mirror-cavity-mirror (Fig. 4a).

The cavity structure may be any combination of grooves, or metal strips represented by a matrix M_r

$$(M_r) = \begin{vmatrix} a_r & b_r^* \\ b_r & a_r^* \end{vmatrix} \tag{41}$$

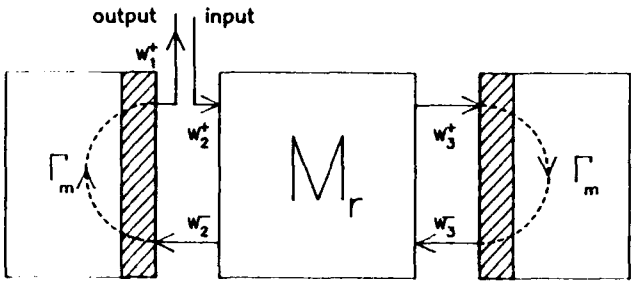


Fig. 4a : Resonant condition of a 3-grating structure

Mirrors M_m and M'_m are supposed to be an infinite stack of grooves, operating inside stopband. Such an infinite array is characterized by a single complex reflection coefficient Γ_m, since no energy is supposed to flow at infinity. Γ_m is defined by

$$\Gamma_m = \frac{w_3^-}{w_3^+} = \frac{w_2^-}{w_1^+} = \frac{b_m}{a_m} \tag{42}$$

where a_m and b_m are coefficients of the mirror matrix M_m.

Resonant condition is obtained when input w_2^+ and output w_1^+ are equal, and combining matrix M_r with Γ_m yields

$$1 = \Gamma_m \frac{b_r + a_r^* \Gamma_m}{a_r + b_r^* \Gamma_m} \tag{43}$$

b) Application to a 3-grating resonator

Fig. 4b shows a model of metal strip resonator or grooved resonator, not taking into account the presence of IDT transducers.

Reflection coefficient Γ_m is obtained from (38, 39) in the limit of inifinite array length, using

$$\lim_{L \to \infty} \frac{\cosh \sigma L}{\sinh \sigma L}$$

$$\Gamma_m = \sin\theta - i\cos\theta = e^{i(\theta - \pi/2)} \tag{44}$$

Consider the case of a cavity made with v strips and (v+1) gaps, of periodicity Λ_r (Fig. 4b). Cavity length will be given by

$$L_r = (2v + 1)\,\Lambda_r/2 \tag{45}$$

Cavity operating inside passband will be represented by a transfer matrix with $d = -\Lambda_r/2$ and $L = L_r$ in Eqs. (32-34). This expression is valid only near stopband. Far from stopband, another approximation can be made [6, 12] due to the fact that coupling between forward and reverse Floquet modes vanishes, with as a consequence that off-diagonal terms of transfer matrices will also vanish. Transfer matrix in passband, far from stopband will take the simpler form

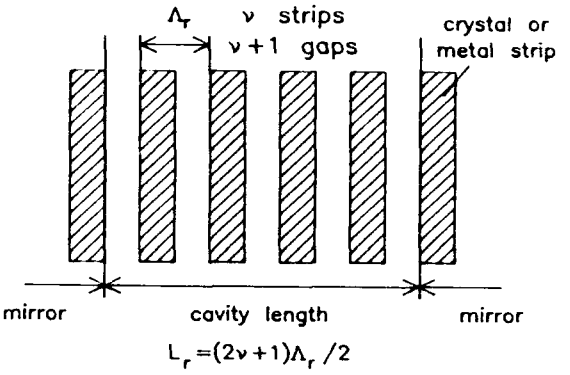


Fig. 4b : Model of a 3-grating resonator

$$\left(M_r\right) = \begin{vmatrix} e^{i\beta_r L_r} & 0 \\ 0 & e^{-i\beta_r L_r} \end{vmatrix} \tag{46}$$

where the propagation constant β_r is given by the limit form of dispersion equation of a medium without any periodicity.

Case of grooves : limit case is SSBW curve

$$\beta_r = \frac{\omega}{V_s} \tag{47}$$

Case of metal strips : limit case is Love wave propagation, corresponding to propagation under a continuous layer of effective thickness $h_{eff} = h/2$.

Love wave dispersion curve can be recalculated from fundamental equations given in [12] and takes the following form

$$\beta_r^2 = \left(\frac{\omega}{V_s}\right)^2 \left(1 + \left(\frac{\omega}{V_s}\right)^2 \left(B\frac{h}{2}\right)^2 \left|\left(\frac{V_s}{V_m}\right)^2 - 1\right|^2\right) \tag{48}$$

where Love wave constant B is given by

$$B = \frac{C_{eff}}{C_{66}} \frac{\rho'}{\rho} \left(\frac{V_m}{V_s}\right)^2 \tag{49}$$

Combining the expressions of mirror reflection coefficient Γ_m with transfer matrix of the cavity M_r, the condition of resonance takes the final form

$$\Gamma_m^2 = e^{2i\beta_r L_r} \tag{50}$$

Following [9], and introducing the phase $(-\pi/2 + \theta)$ of the reflection coefficient Γ_m, eq. (50) can be rewritten as

$$\beta_r L_r = -\frac{\pi}{2} + \theta + N\pi = \frac{\pi}{2} + \theta + N'\pi \tag{51}$$

The algorithm used to find the resonant frequency can now be summarized as follows, illustrated on Fig. 5.

* Let angular frequency ω scan the lower half of mirror stopband

$$\omega : \left(\frac{\pi V_s}{\Lambda_m} - K_m M V_s \right) \rightarrow \left(\frac{\pi V_s}{\Lambda_m} \right) \tag{52}$$

where

$$K_m = 2 \frac{\pi}{\Lambda_m} \left(\frac{h}{\Lambda_m} \right)^2 \frac{C_{eff}}{C_{66}} \tag{53}$$

and M = modifying factor for metal strip gratings [5, 6, 12]

$$M = q^2 \left(1 + \frac{\pi}{2} \frac{1 - (V_m / V_s)^2}{1 + (V_m / V_s)^2} \right) \tag{54}$$

($M = 1$ for grooved gratings).

* Plot curve β_r - $[(\pi/2 + \theta)/L_r]$ as a function of $\delta\omega$, where θ is given by eq. (37).

* Plot equidistant lines separated by π/L_r and check for intersection (Fig. 5).

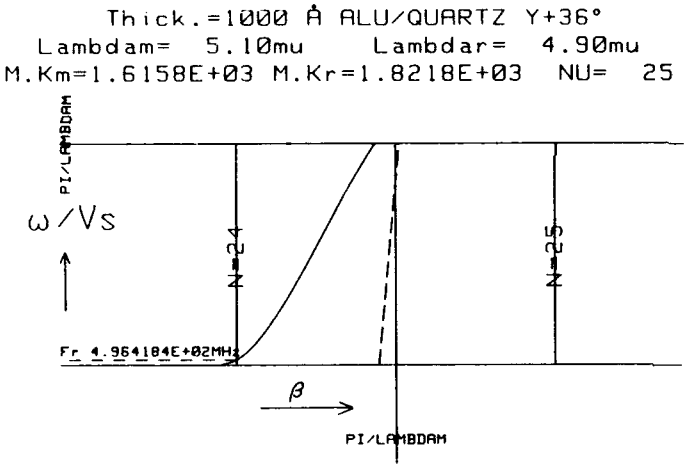


Fig. 5 :
Graphical determination of resonant frequency

4. Simulation results

Simulations using the algorithm described in Sec. 3 will be presented here, obtained with the following nominal parameters

- nonpiezoelectric approximation for quartz, Y + 36° cut (convention on cut angles according to [15])
- operating frequency close to 500 MHz
- metal thickness or groove depth : 1 000 Å
- periodicity of mirrors : 5.1 µm
- periodicity of cavity : 4.9 µm

constants of Y + 36° quartz : ρ = 2 650 kg/m^3
C_{66} = 2,895 × 10^{10} Pa C_{eff} = 6.830 × 10^{10} Pa
V_s = 5 070 m/s
metal Aℓ + 4%Cu ρ' = 2 695 kg/m3 Vm = 3 046 m/s

a) Influence of number of elements inside cavity

Fig. 6 shows the dependence of the resonant frequency vs number of strips v inside cavity (see Fig. 4b) in the case of a grooved resonator (Fig. 6a) and metal strip resonator (Fig. 6b). Since maximum of energy trapping effects are obtained at the bottom of the stopband, best results are expected in both cases in the vicinity of 25 and 48 strips inside cavity. Based on these simulations, metal strip resonators have been fabricated and tested with 25, 31 and 37 strips. Results are presented below in sec. 5.

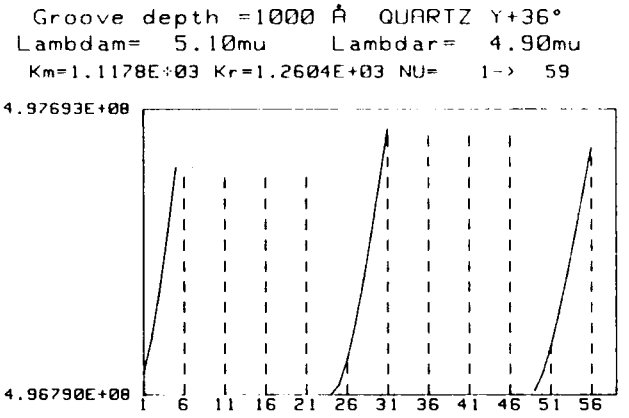


Fig. 6a : Grooved resonator

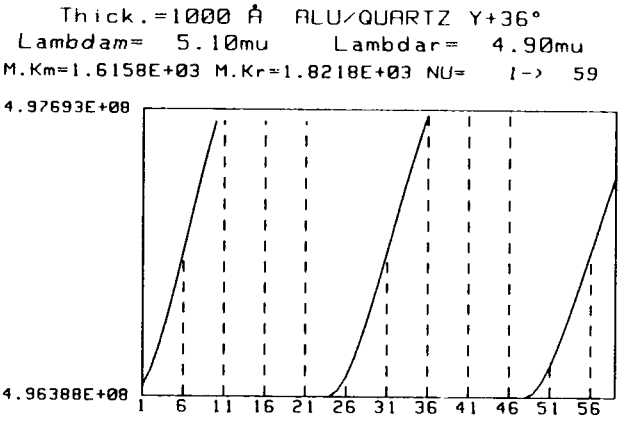


Fig. 6b : Metal strip resonator

Influence of number of elements inside cavity on the resonant frequency. Resonant frequencies in the upper part of stopband are not of practical use due to radiation losses of the device.

b) Influence of different design parameters

Based on nominal parameters exposed before, a metal strip resonator design with 25 grooves inside cavity exhibits sensitivity to different parameters.

* a scaling factor applied to both geometrical dimensions Λr and Λm will scale resonant frequency by the same factor.
* metal thickness changing from 1 000 Å to 800 Å will increase resonant frequency by 900 ppm.
* changing cut angle from 36° to 37° decreases resonant frequency by 2 000 ppm, but as mentioned in [4, 6] the most important effect of changing cut angle is changing turnover temperature of the device.

5. Comparison with already published data. New experimental results

a) Corrections to the nonpiezoelectric approximation

Comparison of frequency ranges predicted by the model (Fig. 6) for a 500 MHz device with already published data [4] shows that predicted operating frequencies are lower 'han observed experimentally. A reason to this may be found in the nonpiezoelectric approximation used here, predicting SSBW velocity V_s = 5 070 m/s for Y+36° quartz cut, while experimental value given by Lewis [16] is 5 099 m/s, very close to the value given by classical piezoelectric bulk mode analysis. To take into account this increased value (by 0.5 %) for V_s, an attempt was made by simply changing value of V_s and $C_{eff} = \rho V_s^2$ in all formulae used in the simulation, and a better agreement between theory and experiment was found for new metal strip resonators (see below sec. 5c).

b) Influence of metal thickness : comparison with Avramov's results [10]

A systematic investigation of the influence of metal thickness is presented by Avramov [10] on 620 MHz metal strip STW resonators on (Y+36°) quartz. To simulate the behavior or such a device, the following inputs were used in our simulation program.
mirror periodicity Λ_m = 4.1 µm
acoustic wavelength λ_a = 8.2 µm
resonator periodicity Λ_r = 4.05 µm
number of strips inside cavity : 11
quartz SSBW velocity : 5 099 m/s

Table 1 compares experimental values of frequency shifts with predicted results.

Metal thickness h/λ_a	Experimental [10] frequency shift $(F_o-F)/F_o$ (ppm)	Predicted frequency shift $(F_o-F)/F_o$ (ppm)
1 %	2 000	1 850
1.3 %	3 000	3 200
1.6 %	4 000	4 800
2. %	6 000	7 600
2.5 %	12 000	12 000

Table 1 : Influence of metal thickness on a 620 MHz metal strip resonator.
Reference frequency Fo corresponds to h/λ_a = 0.4 %

The main effect of changing metal thickness is to reduce the frequency range between the lower edge and middle of stopband, and does not affect significantly the relative position of the resonance inside stopband. This will explain why STW resonators can be trimmed by metal etching without changing their characteristics. However important changes of metal thickness will significantly affect the Q-factor value [10], and this is not taken into account in the present model.

c) Results obtained on 500 MHz metal strip resonators

Metal strip resonators were fabricated and tested according to the design curve of Fig. 6b. Three sets of devices were fabricated on Y+37° quartz with respectively 25, 31 and 37 strips inside cavity. Device with 37 strips did not exhibit any usable resonance peak, and preliminary results obtained on the first prototypes with 25 and 31 strips are listed on Table 2. Aluminum thickness was 1 100 ± 50 Å, Λ_m = 5.1 µm Λ_r = 4.9 µm.

Number of strips inside cavity	25	31
Predicted frequency (non piezo) Vs = 5 066 Vm = 3 046 (MHz)	495.386	496.010
Corrected value of predicted frequency V_s = 5 092 V_m = 3 046 (MHz)	497.679	498.449
Experimental frequency (MHz)	497.737 ± 0.014	498.037 ± 0.010
Unloaded Q factor (experimental)	6.90 ± 0.61 × 10^3	2.98 ± 0.08 × 10^3
Insertion losses (experimental) (dB)	9.1 ± 0.4	11.4 ± 0.2

Table 2 : Comparison of predicted values of frequency with experimental results on metal strip STW resonators
Mean value and standard deviation of experimental data are obtained from a set of four devices

These results clearly show the influence of a correct position of the resonant frequency inside stopband, best values of Q factor and insertion losses being obtained at lower edge of stopband. Corrections to the nonpiezoelectric model by simply changing values of V_s give a better agreement between theory and experiment ; but the question is to know a correct value for V_m, metal properties depending on the deposition process.

Frequency response of a resonator with cavity length of 25 strips is shown on Fig. 7.

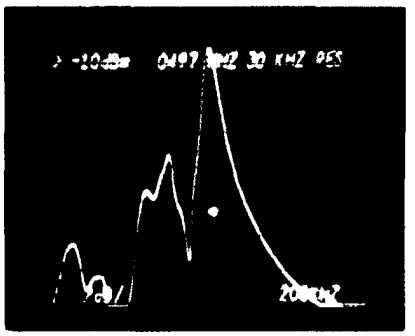


Fig. 7 : Measured frequency response
for STW resonator with cavity length of 25 strips.

Conclusion

The model presented here allows a systematic investigation of the influence of design parameters on the resonant frequency of STW resonators. Operating at lower edge of stopband to ensure high Q and low loss has been pointed out as one of the most important requirements for STW resonators.

The reason is that energy-trapping effects vanish when resonant frequency falls in the upper part of stopband. Careful choice of geometrical parameters (groove pitch, cavity length) are, a priori, required in the design, but resonant frequency can be trimmed by changing metal thickness without altering other characteristics of the device. The importance of taking piezoelectricity into account was recognized for better precision. In its present form the model can be extended to any combination of groove/metal strip gratings that could be required, for example, to control precisely temperature behavior of a STW resonator.

Acknowledgements

The authors wish to thank William S. Drummond and Robert A. Mueller fortheir invaluable assistance with device fabrication and measurement.

References

[1] B.A. Auld, J.J. Gagnepain, M. Tan, "Horizontal shear waves on corrugated surfaces", Elect. Lett., vol. 12, n° 24, pp. 650-651 (1976).

[2] A. Renard, J. Henaff, B.A. Auld, "SH surface wave propagation on corrugated surfaces of rotated Y-cut quartz and berlinite crystals", 1981 Ultrasonics Symp. Proc. IIEE, cat. 86CH2375-4, pp. 257-280.

[3] A. Ronnekleiv, "High Q resonators based on surface transverse waves", Proc. 1986 Ultrasonics Symp., IEEE cat. n° 86CH2375-4, pp. 257-280.

[4] T.L. Bagwell and R.C. Bray, "Novel surface transverse wave resonators with low loss and high Q", Proc. 1987 Ultrasonics Symp., IEEE cat. n° 87CH2492-7, pp. 319-324.

[5] D.F. Thompson, B.A. Auld, "Surface Transverse Wave propagation under metal strip gratings", Proc. 1986 Ultrasonics Symp., IEEE cat. n° 86CH2375-4, pp. 261-266.

[6] B.A. Auld and D.F. Thompson, "Temperature compensation of Surface Transverse Waves for stable oscillator applications", Proc. 1987 Ultrasonics Symp., IEEE cat. n° 87CH2492-7, pp. 305-312.

[7] C.A. Flory and R.L. Baer, "Surface transverse wave mode analysis and coupling to interdigital transducers", Proc. 1987 Ultrasonics Symp., IEEE cat.n° 87CH2492-7, pp. 313-318.

[8] A. Ronnekleiv, "Anisotropy in surface transverse waves resonators", Proc. 1987 Ultrasonics Symp., IEEE cat.n° 87CH2492-7, pp. 325-329.

[9] B.A. Auld, "The STW resonator, a new control element for high stability oscillators", Proc. 3rd European Frequency and Time Forum, Besançon, France, pp. 152-158 (march 1989).

[10] I.D. Avramov, "High Q metal strip SSBW resonators using a SAW design", IEEE Trans. Ultr. Ferro. Freq. Contr., vol. UFFC-37, n° 6, pp. 530-534 (nov. 1990).

[11] P.S. Cross and R.V. Schmidt, "Coupled surface acoustic wave resonators", Bell System Technical Journal, vol. 56, pp. 1447-1482 (1977).

[12] D.F. Thompson, "Temperature compensation of microwave acoustic resonators", Ph. D. Dissertation, Ginzton Laboratory, Stanford University, (june 1986).

[13] S. Datta and B.J. Hunsinger, J. Appl. Phys., vol. 50, pp. 5561-5565 (1979).

[14] L.A. Coldren and R.L. Rosenberg, Proc. 1976 Ultrasonics Symp., IEEE cat. 76CH1120-5SU, pp. 266-271.

[15] IEEE Standard on Piezoelectricity 176-1949, Proc. IRE, vol. 37 (1949), pp. 1378-1395.

[16] M. Lewis, "Surface skimming bulk waves, SSBW", Proc. 1977 Ultrasonics Symp., IEEE cat. 77CH1264-1SU, pp. 744-752.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

1 GHz VOLTAGE CONTROLLED OSCILLATOR USING A SSBW/BGW COMBINED MODE RESONATOR FILTER

IVAN D. AVRAMOV

Institute of Solid State Physics
72 Lenin Boulevard
1784 Sofia, Republic of Bulgaria

Abstract

Results from a 1 GHz voltage controlled oscillator (VCO) stabilized with a novel 3-pole combined mode resonator filter (CMRF) using surface skimming bulk waves (SSBW) and Bleustein-Gulyaev waves (BGW) are presented. Fabricated on rotated Y-cut quartz the CMRF can handle more than 2W of power on a substrate area of 3 mm^2, has a typical insertion loss of 5-7 dB, a 1 dB bandwidth of 0.17% and a bandpass ripple within 1 dB. The smooth phase response around its center frequency allows linear tuning of the VCO. Compared to a surface acoustic wave coupled resonator filter (SAWCRF) the CMRF has a 3-4 times larger fractional bandwidth and about 5 times lower group delay which makes it ideal for wideband frequency modulation (FM) applications. The CMRF stabilized VCO can transmit more than 200 KBits/s when square wave modulated. When sine wave modulated the 3 dB frequency deviation drop occurs at 100 KHz. Design details are described and possible applications are discussed.

Introduction

Shear horizontal surface waves were first postulated by A.E.H. Love to explain seismic data [1]. Guiding of shear horizontal waves on a substrate surface by surface corrugation and by metallic gratings was demonstrated by B.A. Auld et al. in 1976 [2]. Coupling to bulk shear transverse waves on piezoelectric substartes with interdigital transducers (IDT) for device applications was investigated by Yen et al. [3] and by Browning and Lewis [4] in 1977. These waves were referred to as shallow bulk acoustic waves (SBAW) or surface skimming bulk waves (SSBW). Since that time the interest in these waves has been increasing because, compared to surface acoustic waves (SAW), they have some significant advantages for device applications. These are: 1.6 times higher propagation velocity, increased power handling capability, decreased surface sensitivity and superior aging, higher intrinsic material Q and lower device phase noise.

Propagation of SSBW under metal strip gratings has been investigated in earlier work [5], [6], [7]. It is well known that periodic transducer and grating metalization on rotated Y-cut quartz can drastically change the nature of the wave, converting it from a bulk wave to a guided Bleustein-Gulyaev wave (also called surface transverse wave (STW) [6]), which is a surface wave in nature [5]. At the Brag frequency of the grating the coupling of the SSBW and the BGW and the difference between their propagation velocities can be predicted by carefully selecting the effective metal thickness and the number of transducer finger pairs [5]. Below the cutoff frequency of the grating the BGW is closely trapped to the surface and propagates without or with negligible diffraction into the bulk [6]. In 1987 this phenomenon was successfully implemented in two-port resonators by Bagwell and Bray who demonstrated 500 MHz STW resonators with unloaded Q of about 12000 and insertion loss of 5.5 dB [8].

This paper presents design details and results from a novel combined mode resonator filter recently reported in [9]. Its design and configuration are very similar to a two-port synchronous IDT SAW resonator [10] but the frequency response in the filter passband is formed by two BGW cavity modes and one SSBW resonant mode which are excited with equal magnitudes and a predicted frequency displacement. The result is a three pole Chebychev type resonator filter with high power handling ability, low insertion loss, increased fractional bandwidth, and extreemely low group delay which makes it very suitable for VCO with wideband FM capabilities, low-noise oscillators operating at high loop power levels and transceiver front-end applications .

Modelling and Design of the CMRF

As shown by Lee in 1980 the difference between the velocities of the BGW and SSBW propagating under a periodic structure (IDT or reflector grating) is caused by mass loading and depends on the crystal orientation [5]. It was found that this effect is sufficiently strong for 36° rotated Y-cut quartz, a cut which is very well suited for device applications for its superior temperature stability and high SSBW propagation velocity. The normalized deviation in the BGW velocity ($\Delta v/v$) from the SSBW velocity at cutoff for this cut depends on the mass loading as follows:

$$(\Delta v/v)/(h_{eff}/\lambda)^2 = 16 \qquad (1)$$

CH2965-2/91/0000-230 $1.00 © 1991 IEEE

where λ is the acoustic wavelength and h_{eff} is the effective metal thickness which is the real thickness h times the metallization fill factor m.

The acoustic radiation from the IDT consists of two components, one related to the SSBW radiation and the second one associated with the BGW. As shown in [5] the values of both radiation resistances R_{SSBW} and R_{BGW} depend on the number of finger pairs N in the IDT and $(\Delta v/v)$, which in turn is a function of the mass loading as shown by equation (1). The relationship between both radiation resistances and the parameter $N(\Delta v/v)$, derived in [5], is shown in Fig. 1.

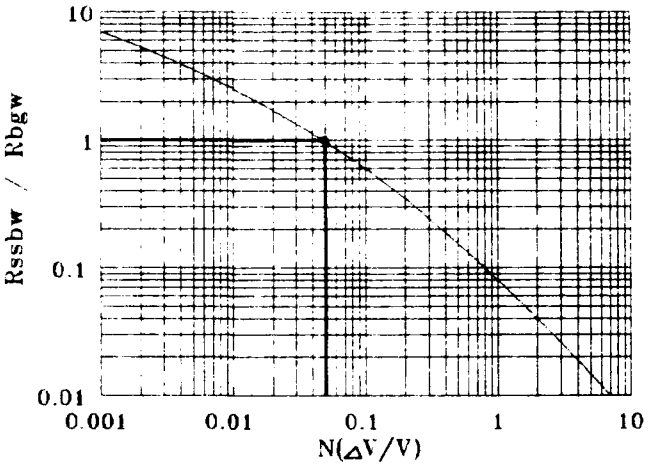


Fig. 1. Radiation resistance dependence on mass loading and IDT length.

It is evident that R_{SSBW} decreases with increased mass loading and number of finger pairs while R_{BGW} behaves in the opposite manner. Mass loading and increased number of IDT finger pairs is to favor excitation of BGW at the expense of SSBW. As evident from Fig. 1. both radiation resistances become equal for:

$$N(\Delta v/v) = 0.045 \qquad (2)$$

According to equation (1), this is valid for a large number of N and h_{eff}/λ combinations which offers a great design flexibility.

The basic idea of this work is, using a simple two-port resonator configuration, to obtain the frequency response of a three-pole Chebychev type resonator filter using SSBW and BGW cavity modes with equal magnitudes and predicted difference in frequency. It is well known that, by altering the effective cavity length of a two-port resonator, it can be designed to support one, two or more cavity resonant modes with a defined distance in frequency between them [19]. Here such a configuration was designed to support two BGW cavity modes with equal magnitudes which determine the bandwidth of the filter. There is always a large dip between these modes. It has to be compensated for in order to obtain a flat passband response. This is accomplished by exciting an SSBW mode with the same magnitude as the BGW modes right in the middle between them. As discussed above, this is possible by properly selecting the number of IDT finger pairs and the effective metallization.

The CMRF schematic layout considered here is shown in Fig. 2.

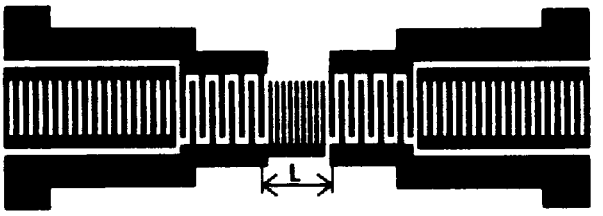


Fig. 2. CMRF schematic layout

It is a typical two-port synchronous IDT resonator configuration. The resonant peak in such a design occurs below the center frequency of the grating stopband [10] which ensures proper guiding of the BGW [6]. A grating between both IDT with a period slightly lower than the transducer/reflector one, keeps the BGW tightly bound to the surface when *propagating* between the IDT [7], [8].

Since BGW are surface waves in nature, the BGW response of the device in Fig. 2. can readily be analyzed using Coupling-of-Modes (COM) theory applied to SAW devices. This analysis was performed by means of a COM software that has been initially developed for modelling SAW unidirectional transducers and resonators [11]. Here it was modified to analyze BGW two-port resonator configurations. The parameters used in the COM algorithm were altered till very good agreement between theoretical and experimentally obtained results from test devices was achieved. This modified COM analysis software was used to synthesize the 1 GHz CMRF.

The design goal was a CMRF with a center frequency f_o = 1000 ± 1 MHz, a 1 dB bandwidth of 1.6 MHz (0.16%) and an insertion loss of less than 8 dB. In order to achieve the desired bandwidth two BGW cavity modes have to be excited with a difference Δf_{BGW} of about 1.2 MHz (0.12%) between each other. The next step is to determine the number of IDT finger pairs so that one SSBW mode occurs about 600 KHz (0.06%) higher than the left BGW peak. From (2) it follows that for $\Delta v/v$ = 0.06% the necessary N value is 75. According to (1) the corresponding h_{eff}/λ value is 0.61% which can be realized with h/λ = 1.5% and m = 0.4. The design procedure for this case is illustrated in Fig. 3. It shows theoretically predicted BGW frequency responses for different values of the IDT distance

L which determines the cavity length [19]. First L is roughly selected to meet the requirement for the BGW mode distance. Then it is altered by small parts of λ till both modes achieve the same magnitude. As evident from Fig. 3 this is obtained for L = 6.445λ.

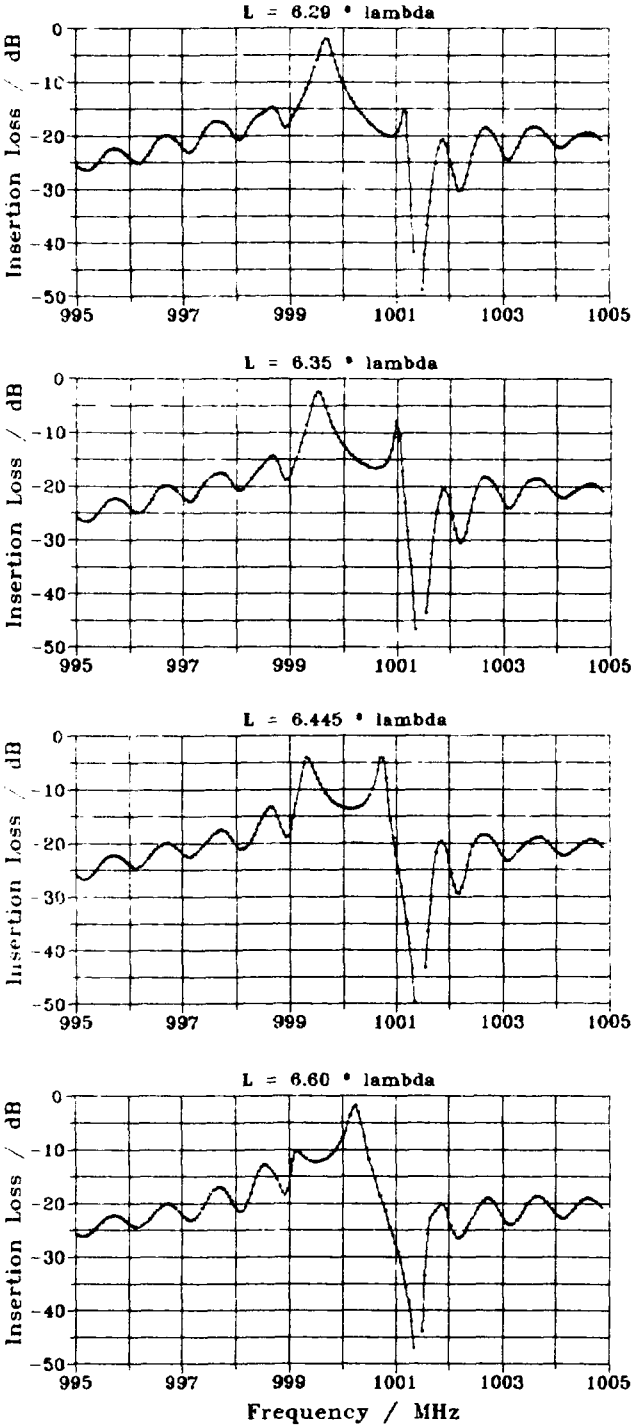


Fig. 3. Theoretically predicted BGW responses for different L-values.

If a high Q single mode resonator is desired an L value should be chosen so that the resonant peak is placed maximally to the left of the stopband center frequency while other cavity modes are well enough suppressed [6], [8]. Obviously here this is the case for L = 6.29λ.

Two dozen devices were fabricated on 36° rotated Y-cut quartz in a standard single step photolithographic process. Test devices were first realized for a random center frequency. Their characteristics were measured in order to determine the equivalent propagation velocities of the BGW and SSBW modes under the CMRF configuration. Then a mask correction was performed to achieve the desired center frequency. The design parameters are summarized in Table 1.

Fig. 4a shows the theoretically predicted position of the SSBW resonant mode that is expected to flatten the dip between both BGW modes forming a Chebychev type passband response.

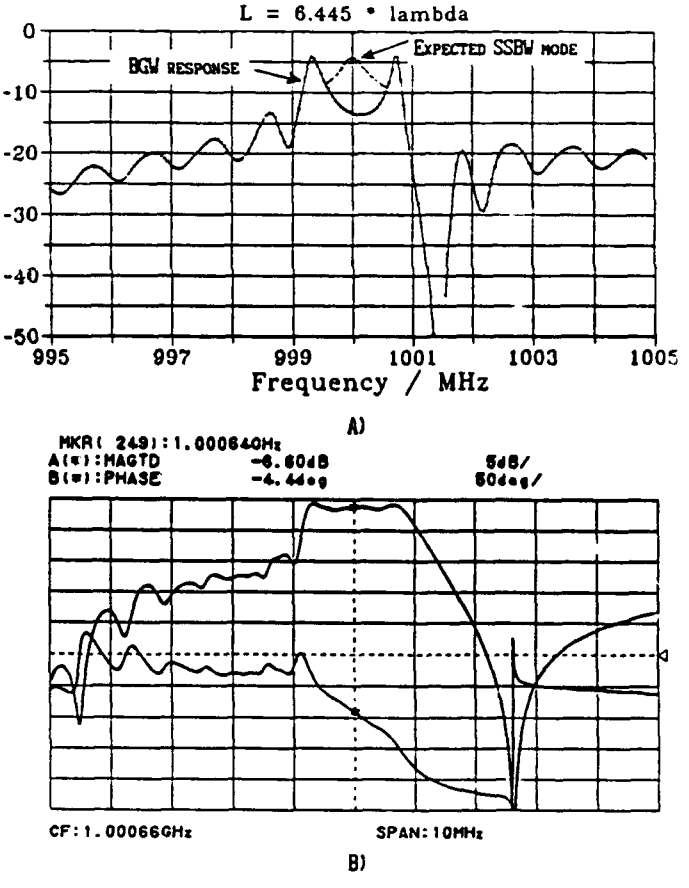


Fig. 4. Comparison between theoretically predicted and experimental CMRF characteristics.
a) Predicted position of SSBW mode. vertical axis: Insertion Loss / dB
b) Frequency response (upper curve) and phase response (lower curve) of a typical experimented CMRF.

The comparison with the experimental CMRF frequency response (upper curve in Fig. 4b) indicates a very good agreement between theory and experiment.

Table 1. CMRF design parameters

Design Parameter	1 GHz CMRF
Number of BGW cavity modes	2
Number of strips per reflector	900
Number of finger pairs per IDT	75
Number of center grating strips	12
Reflector/transducer period	2.532 μm
Wavelength	5.064 μm
Period of center grating	2.4 μm
Acoustic apperture	100 λ
Metallization	760Å (Al+4%Cu)
Equivalent SSBW velocity	5064 m/s

Actually a second SSBW cavity mode should be expected to appear about 0.06% higher than the right BGW mode. However this SSBW mode will strongly diffract into the bulk since it is on the steep right slope of the grating stop band characteristic. This explains its absence on the the experimentally obtained plot in Fig. 4b.

Fig 4b. shows the frequency and phase response of a typical CMRF. The Chebychev type passband characteristic is clearly visible. The phase slope over the entire 1 dB bandwidth is about 180°, a fact of great practical importance for low-cost oscillators. Unlike STW resonator oscillators that have only 90° phase shift over the 3 dB resonant bandwidth, an oscillator stabilized with a CMRF does not need additional phase shift to set the phase condition for oscillation. If necessary, reversing the CMRF bond leads is sufficient to start oscillation. In this respect the CMRF behaves exactly as a two-pole SAW coupled resonator filter (CRF) [12] but its 1 dB bandwidth is typically 3-4 times larger while the group delay is about 5 times lower which is due to the third Chebychev pole. This is evident from Fig. 5 which shows the CMRF passband and group delay behavior in detail. The group delay at the center frequency is only 180 ns. A 1 GHz STW delay line with this group delay would have at least 20 dB insertion loss. The CMRF in Fig. 5 has an insertion loss of only 5.9 dB while the 1 dB bandpass ripple does not exceed 1 dB. As a matter of fact the lowest loss achieved with a CMRF device was 2.9 dB [9] which, to the best of the author's knowledge, is the lowest loss accomplished with an STW or SSBW device to date. In this device the low loss was obtained using thinner metallization and at the expense of smaller device bandwidth and higher out of band ripple (see Fig. 4 in [9]). This low loss is very essential for receiver front end applications.

Other Useful CMRF Properties

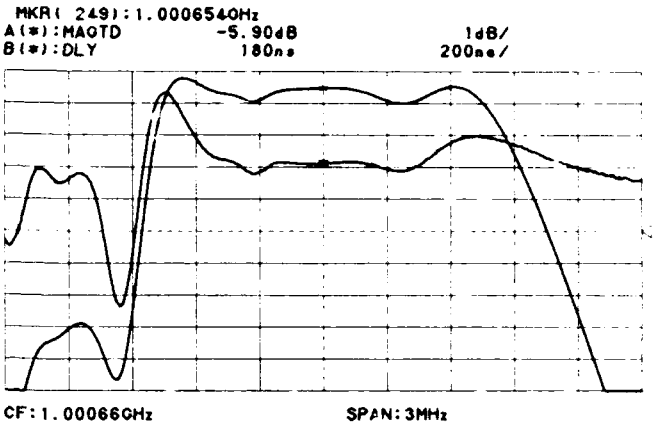


Fig. 5. Detailed frequency response and group delay CMRF characteristics.

Fig. 6 is the wide band frequency response of the CMRF over a frequency span of 2 GHz. It is evident that the filter suppresses all upper UHF channels (below 860 MHz) by ≥20 dB which is another useful characteristic for receiver front ends.

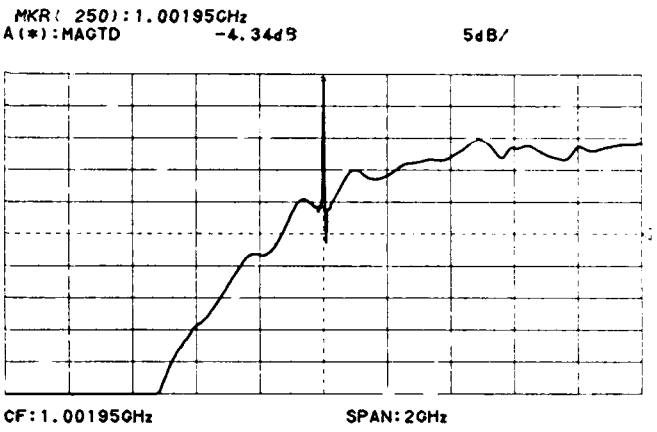


Fig. 6. Wideband CMRF frequency response.

Another respectable feature of the CMRF is its high power handling capability. According to Bagwell and Bray the highest peak stress component in an STW resonator is the shear stress T_{12}, where 1 is the propagation direction (normal to x) and 2 is normal to 1 in the substrate plane [8]. For STW resonators on 36° rotated Y-cut quartz with a propagation normal to x T_{12} can be calculated using following equation:

$$|T_{12}|_p = 2050 \sqrt{\frac{Q P_d \lambda}{L D W}} \qquad (3)$$

where Q is the device unloaded quality factor, P_d is the power dissipated in the device, λ is the acoustic wave length, and L are the acoustic width and effective cavity length respectively and D is the effective penetration dept into the substrate. With Q = 960, λ = 5.064 μm, W = 100λ, L = $(f_o/2\Delta f_{BGW})\lambda$ = 385λ and assuming that D/λ = 0.75 [8], for P_d = +33 dBm (2W) a peak stress value $|T_{12}|_p = 1.044 \times 10^8$ N/m^2 is obtained. This stress is much lower than the degradation peak stress value of 2.8×10^8 N/m^2 for SAW resonators with Al-Cu metallization. In fact the device in Fig. 5 which occupies a substrate area of about 3 mm^2, was operated with an incident power of +33 dBm for a period of 5 days. No measurable performance degradation was observed. This high power handling capability of the CMRF could be of great practical importance in the design of oscillators with high loop power and extremely low noise floor. As shown by Parker the noise floor of stable feedback-loop oscillators is expressed by following term [13]:

$$S_{\phi\varepsilon}(f) = 10 \log (2GFKT / P_o) \qquad (4)$$

where G is the compressed power gain of the loop amplifier, F is the amplifier's noise figure, T is the temperature in °K, K = 1.38054×10^{-23} Ws/K is the Boltzmann's constant and P_o is the carrier power level at the amplifier output. With G = 4 (6 dB), F = 2.5 (4 dB), T = 300K and P_o = 2W the oscillator would have a thermal noise floor of 194 dBc/Hz which would be an improvement of 10 dB compared to the best SAW oscillator reported to date [14].

1 GHz CMRF Based Voltage Controlled Oscillator

The CMRF was designed to meet the requirements to a low-noise temperature stable VCO with a wide linear tuning range and good FM capabilities. The FM bandwidth of SAW resonator oscillators is limitted to a few KHz because of the very high resonator Q. SAW delay line oscillators offer better FM capabilities but at frequencies around 1 GHz and higher, low loss and small group delays are difficult to achieve with delay lines on temperature stable substrate materials. Single phase unidirectional transducers (SPUDT) are an alternative since very low insertion loss can be achieved [20], but if these devices are fabricated on quartz, the transducers become too long to ensure small group delays. As it will be shown below the CMRF presented here seems to be a very good solution since it offers low loss, low group delay, sufficient bandwidth, smooth phase response and good temperature stability.

To test the advantages of CMRF in VCO applications one of the devices was implemented in the oscillator circuit shown in Fig. 7. It consists of two broadband single stage bipolar transistor amplifiers A1 and A2, a CLC-type varactor phase shifter and a buffer amplifier A3. The phase shifter of the type in Fig.7 is known for its tuning linearity when used in a SAW resonator stabilized VCO [14]. Here it is placed between A1 and A2 to ensure sufficient isolation from the CMRF in the loop. With the varactors BBY 39 from Philips this phase shifter provides a linear phase shift of about 60° which is sufficient to provide tuning over the desired frequency range of ± 400 ppm. The gain of A1 and A2 can be adjusted in wide limits by means of series and parallel feedback resistors. Here the total loop gain was adjusted so that the gain compression did not exceed 6 dB over the entire tuning range. A3 stabilizes the circuit against load changes. The whole oscillator circuit is realized in surface mounted devices (SMD) technology on a miniature printed circuit board and sealed in a standard hermetic dual-inline metal package. The design and construction of the oscillator circuit are described in detail in [15].

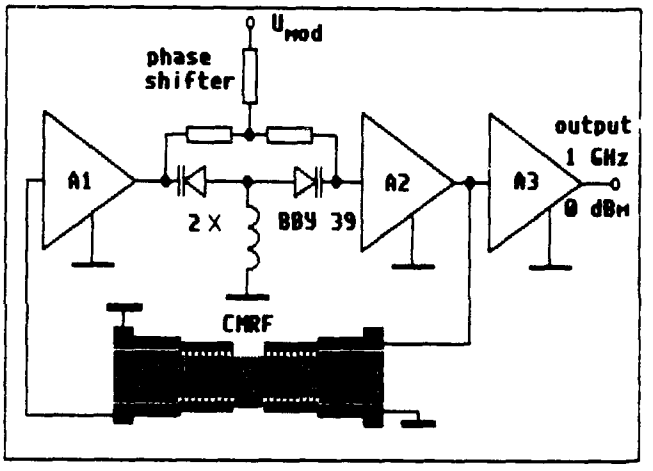


Fig. 7. VCO schematic.

Fig. 8 is the volt/frequency characteristic of the VCO over the whole tuning range.

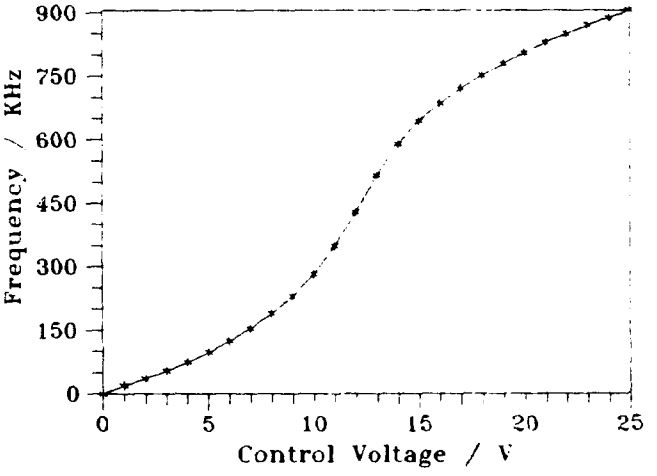


Fig. 8. Volt/frequency VCO characteristic.

It is S-shaped which means that it follows the phase slope change of the CMRF phase response. As shown in Fig. 9 this characteristic has a linear part of about 350 KHz in which extremely linear FM data transmission is possible. In fact with a sine wave modulation and a frequency deviation of ± 100 KHz the nonlinear distortion does not exceed a few tenths of a percent. This value was measured with modulation frequencies of 1...20 KHz.

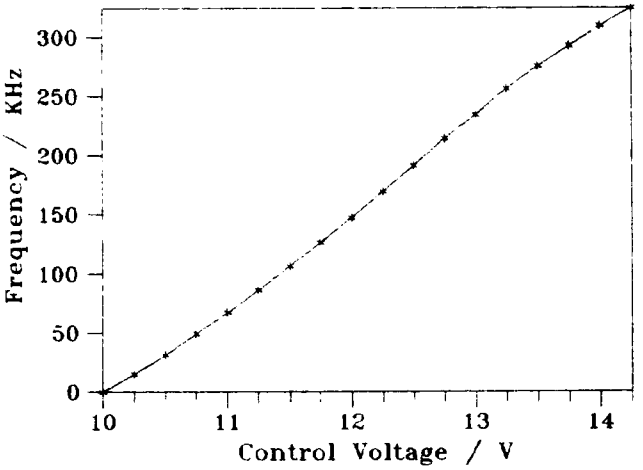


Fig. 9. Linear part of the volt/frequency VCO characteristic.

To test the FM capabilities of the VCO a simple but very linear and sensitive detector circuit was built. Its schematic is shown in Fig. 10.

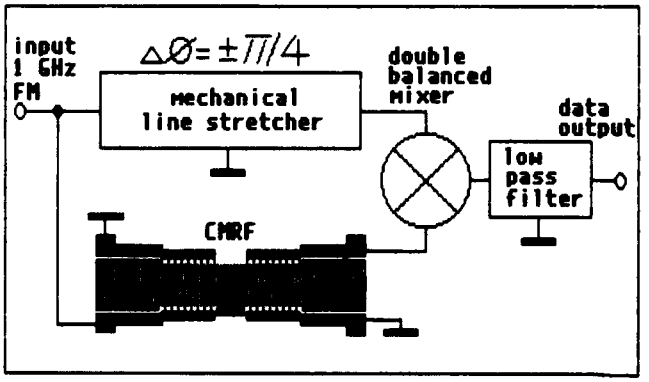


Fig. 10. 1 GHz FM-detector circuit.

The 1 GHz FM signal derived from the frequency modulated VCO is applied through a mechanical line stretcher to the reference arm of a double ballanced mixer and then, delayed through a CMRF, to the other arm of the mixer. The CMRF which has the same center frequency and is identical to the one in the VCO, has a very steep phase response and converts the FM signal into a phase modulated (PM) signal. The latter is detected by the mixer operating as a phase detector, all higher frequency mixer products are filtered by a low pass-filter and the output data is compared with the sine wave or square wave signal modulating the VCO. The mechanical line stretcher is used to adjust quadrature between the mixer arms for best linearity and highest sensitivity. The detector was tested using an FM source with a high rate FM capability and was found to be at least an order of magnitude faster than the VCO. For frequency deviations of up to ± 100 KHz the nonlinear distortion caused by the detector was found to be negligible.

The VCO was first modulated with a sine wave signal from a low-distortion audio frequency source. A deviation dF_o of ± 100 KHz was adjusted at a very low modulating frequency (f_m = 500 Hz). Then the modulating frequency was increased and the amplitude of the detected sine wave at the FM-detector output, which is proportional to the frequency deviation, was evaluated. The results are recorded in Fig. 11.

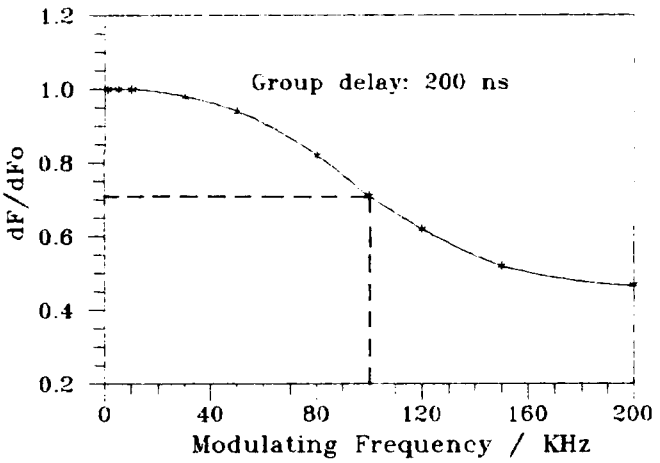


Fig. 11. VCO frequency deviation dependence on modulating frequency.

It is evident that the VCO ability to track the modulating signal decreases with modulating frequency. This is due to energy storage effects in the CMRF. The 3 dB drop in deviation (dF/dF_o = 0.71) occurs at a frequency of 100 KHz which corresponds to about $50\tau_g$, where τ_g is the CMRF group delay value at the VCO carrier frequency.

A similar experiment was performed with square wave modulation, however here the rise and fall time of the detected pulses was evaluated. Fig. 12 is a comparison between the modulating and the detected pulses. It is clearly seen that the VCO integrates the modulating signal in a very similar manner to an RC-integrator. When the VCO is pulsed

the frequency deviation increases and decreases exponentially with time according to following equations:

For the rise time:

$$dF = dF_o \, [1 - \exp(-t/T)] \qquad (5)$$

and for the fall time:

$$dF = dF_o \exp(-t/T) \qquad (6)$$

where the time constant T depends on the CMRF group delay at the VCO carrier frequency. For the VCO experimented here it was found that $T = 18\tau_g = 3.6\mu s$. Assuming that this value is the limit for reliable frequency shift keying (FSK) it turns out that the VCO can transmit FSK signals at a maximum rate of 278 KBits/s.

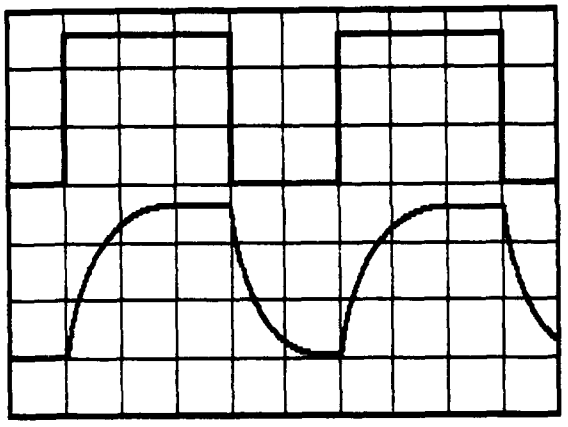


Fig. 12. Square wave modulation results
horizontal axis: 2 μs/div.
upper trace: modulating signal, 500 mV/div.
lower trace: detected signal, 50 mV/div.

The FM experiments described above were performed also with a 767 MHz VCO that was stabilized with a high Q STW resonator presented in [16]. Here τ_g was about 6 times higher than in the CMRF-VCO. The results are compared in Table 2. This table contains also results from phase noise measurements on both oscillators. It is evident that in respect to FM capabilities the CMRF-VCO is at least 10 times faster than the STWR-VCO although the latter has 6 and not 10 times higher group delay. The difference of about 15 dB in flicker FM noise levels between the two oscillators is due to the group delay difference between the STWR and the CMRF and as shown in [14] can be calculated using the equation:

$$\Delta S'_{\phi\varepsilon}(f) = 10 \log (\tau_{g1} / \tau_{g2})^2 \qquad (7)$$

where τ_{g1} and τ_{g2} are the group delays of the STWR and the CMRF respectively.

Table 2. Comparison between the 767 MHz high Q STWR-VCO and the CMRF-VCO.

Parameter	767 MHz STWR-VCO	1 GHz CMRF-VCO
Group delay at f_o	1.16 μs	200 ns
Loaded Q ($Q_L = \pi f_o \tau_g$)	2800	630
Sine wave f_m at 3 dB deviation drop	9.2 KHz $(94\,\tau_g)^{-1}$	100 KHz $(51\,\tau_g)^{-1}$
Deviation rise and fall time	36 μs $(31\,\tau_g)$	3.6 μs $(18\,\tau_g)$
Phase noise suppression at 100 Hz at 1 KHz at 10 KHz away from carrier	83 dBc/Hz 111 dBc/Hz >118 dBc/Hz	67 dBc/Hz 97 dBc/Hz >118 dBc/Hz

Fig. 13 shows the phase noise performance of both oscillators. In the flicker frequency region both graphs are parallel and show the trade off between VCO phase noise performance and FM capability. The system noise floor is 118 dBc/Hz. Therefore the actual phase noise suppression value of the oscillators for 10 KHz offset should be higher than indicated in Fig. 13.

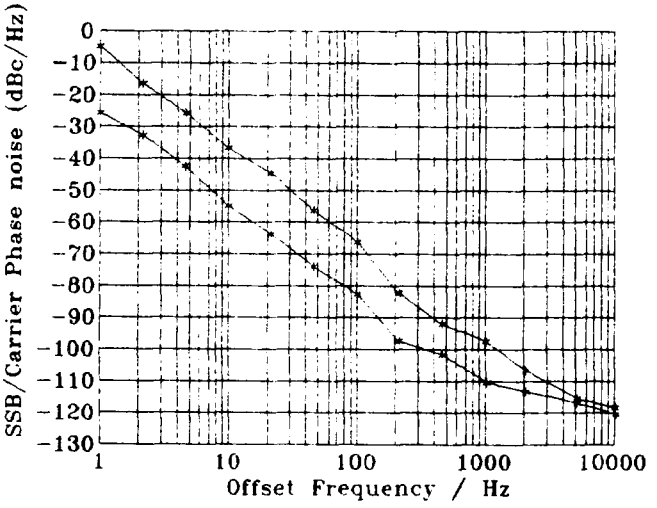


Fig. 13. Phase noise VCO performance.
upper curve: 1 GHz CMRF-VCO
lower curve: 767 MHz high Q STWR-VCO

Fig 14. shows the frequency instability of the 1 GHz VCO with temperature.

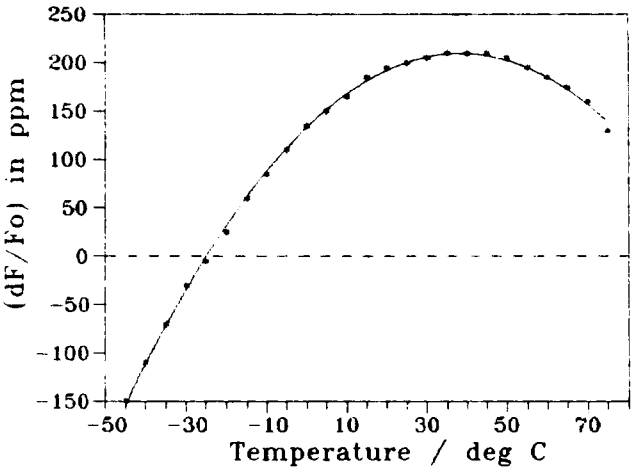


Fig. 14. Temperature dependence of the 1 GHz CMRF-VCO.

As shown in [7] and [8] the turn over temperature depends not only on the quartz cut but also on the metal thickness. Since the latter is very important for the whole CMRF performance, first the device has to be optimized to meet the electrical parameter requirements and than the turn over temperature has to be adjusted by properly selecting the quartz cut. Here this was done by fabricating test devices on four different cuts: 36° 00′, 36° 20′, 36° 40′ and 37° 00′. The desired turn over temperature of 40° C was obtained with the 36° 40′ cut. From Fig. 14 it is evident that the total temperature induced frequency shift does not exceed ± 180 ppm in the temperature range (-45...+75°C) which is much less than the tuning range of the VCO. This fact is of great practical importance for wide range temperature compensated oscillators.

Summary and conclusions

It has been shown that the novel combined mode resonator filter presented here is very well suited to high performance voltage controlled oscillator designs with excellent FM capabilities and respectable phase noise performance, offering a large area of novel applications.

At the Institute of Solid State Physics in Sofia, Bulgaria the VCO presented here was successfully implemented in a local area transceiver for transmitting HiFi sound in buildings using 1 GHz FM channels. Low-noise multiple channel reception over a bandwidth of 2 MHz was performed with a low-cost double conversion superheterodine receiver using a high Q 1015 MHz STW resonator in its first local oscillator circuit. With a transmitted power of -16 dBm (25 μW), coupled from the VCO output, reliable reception of a HiFi sound channel with a standard 200 KHz FM bandwidth through four floors in a concrete building was realized [17].

For another application the CMRF based VCO was implemented in a temperature compensated oscillator with zero warm up time and extremely low weight and power consumption [18]. The method for combined analogue and digital temperature compensation used in this oscillator was realized with low-cost CMOS circuits keeping the power consumption of the compensating circuitry within 5 mA. A total temperature instability of ± 1.4 ppm over the temperature range (-45...+75°C) was achieved which, to the best of the authors' knowledge, is the best result reported with a temperature compensated SAW oscillator so far. Detailed results from these two applications of the CMRF based VCO are being prepared for publication and will be reported at the 5th Conference with International Participation "Acoustoelectronics'91" to be held in Varna, Bulgaria in September this year.

Several 1 GHz CMRF-VCO were fabricated and tested. Their typical performance can be summarized as follows:

Characteristics of 1 GHz CMRF stabilized VCOs.

Center frequency:	1000 MHz
Frequency setting:	± 1000 ppm
Tuning range:	± 450 ppm
Output power:	0 dBm
Spectral purity at:	
10 Hz	≥ 30 dBc/Hz
100 Hz	≥ 60 dBc/Hz
1 KHz	≥ 90 dBc/Hz
10 KHz	≥ 120 dBc/Hz
Harmonic suppression:	≥ 40 dBc
Supply voltage:	15 V
Supply current:	20 mA
Tuning voltage:	0...30 V
Dimensions:	35×28×6 mm

FM capabilities:

Group delay of the CMRF:	≤ 300 ns
Frequency offset:	0...100 KHz
Modulation rate at modulation index = 1:	
Square wave:	≥ 200 KBits/s
Sine wave:	≥ 65 KHz

Using conventional photolithography such devices can be fabricated reproducibly and cost effectively for frequencies up to 1.6 GHz. It is the author's belief that they could find a variety of applications in local area radio networks (LARN), portable and vehicular communications. Further work is required to improve the side lobe suppression and to determine the sensitivity of the CMRF parameters to fabrication tolerances.

Acknowledgements

The author wishes to gratefully acknowledge Dr. P. M. Smith from McMaster University and Mr. P. J. Edmonson from Mohawk College, Hamilton, Ontario, Canada for their permission to use the COM-analysis software. He wishes to thank also Prof. I. B. Yakovkin and Dr. A. S. Kozlov from the Institute of Semiconductor Physics, USSR Academy of Sciences, Novossibirsk, USSR for the fruitfull discussions and for having the CMRF devices fabricated in their laboratories. The author greatly appreciates the support of Mr. T. Thienel from the Technical University in Dresden, Germany who realized the printed circuit boards for the VCO design. He would like to thank also Mrs. A. Stratieva and Mr. M. Mihailov from the Institute of Solid State Physics in Sofia, Bulgaria for expert crystal orientation and preparation of the quartz substrates.

References

[1]. A.E.H. Love, "Some Problems of Geodinamics", Cambridge, University Press, 1911.

[2]. B.A. Auld, J.J. Gagnepain, and M. Tan, "Horizontal Shear Waves on Corrugated Surfaces", Electronics Letters, 12, (24), November 1976, pp. 650-651.

[3]. K.H. Yen, K.L. Wang, and R.S. Kagiwada, "Efficient Bulk Wave Excitation on ST-cut Quartz", Electronics letters, 13, (1), 1977, pp. 37-38.

[4]. T.I. Browning and M.F. Lewis, "New Family of Bulk Acoustic Wave Devices Employing Interdigital Transducers", Electronics Letters, 13, (5), March 1977, pp. 128-130.

[5]. D.L. Lee, "S-Band SSBW Delay Lines for Oscillator Applications", IEEE Proc. 1980 Ultrasonics Symposium, pp. 245-250.

[6]. D.F. Thompson and B.A. Auld, "Surface Transverse Wave Propagation Under Metal Strip Gratings", IEEE Proc. 1986 Ultrasonics Symposium, pp. 261-266.

[7]. B.A. Auld and D.F. Thompson, "Temperature Compensation of Surface Transverse Waves for Stable Oscillator Applications", IEEE Proc. 1987 Ultrasonics Symposium, pp. 305-312.

[8]. T.L. Bagwell and R.C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q", IEEE 1987 Proc. Ultrasonics Symposium, pp. 319-324.

[9]. I.D. Avramov, "1 GHz Low Loss Coupled Resonator Filter Using Surface Skimming Bulk waves and Bleustein-Gulyaev Waves", Electronics Letters, 5, (27), February 1991, pp. 414-415.

[10]. P.S. Cross, W.R. Shreve and T.S. Tan, "Synchronous IDT SAW Resonators with Q Above 10,000", IEEE 1979 Proc. Ultrasonics Symposium, pp. 824-829.

[11]. C.K. Campbell and C.B. Saw, "Analysis and Design of Low-Loss SAW Filters Using Single-Phase Unidirectional Transducers", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC-34, No. 3, May 1987, pp. 357-367.

[12]. J.M. Robinson, "SAW Coupled Resonator Oscillator Technology", IEEE 1989 Ultrasonics Symposium, pp. 47-51.

[13]. T.E. Parker, "Characteristics and Sources of Phase of Phase Noise in Stable Oscillators", Proc. of the 41st Annual Symposium on Frequency Control, 1987, pp.99-110.

[14]. T.E. Parker and G.K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC-35, No.3, pp. 342-364.

[15]. I.D. Awramow, T. Thienel, " Ein hochstabiler abstimmbarer 750-MHz-Oszillator auf der Basis von Oberflaechentransversalwellen (OTW)", Nachrichtentechnik Elektronik, Berlin 40 (1990), 2 pp. 60-63.

[16]. I.D. Avramov, "High Q Metal Strip Resonators Using a SAW Design", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC-37, No.6, pp. 530-534.

[17]. I.D. Avramov, "Surface Transverse Wave Based 1 GHz FM Transceiver for Local Area Radio Networks", to be presented at the 5th Conference with International Participation "Acoustoelectronics'91", to be held in Varna, Bulgaria from 10-13 September 1991.

[18]. M.A Taslakov and I.D. Avramov, "1 GHz Surface Transverse Wave Oscillator with Analogue and Digital Temperature Compensation", to be presented at the 5th Conference with International Participation "Acousto-electronics'91", to be held in Varna, Bulgaria from 10-13 September 1991.

[19]. William J. Tanski and Herman van de Vaart, "The Design of SAW Resonators on Quartz with Emphasis on Two Ports", IEEE 1976 Proc. Ultrasonics Symposium, pp. 260-265.

[20]. T.N. Oliver, D.E. Bower and J. Dowset, "Low Loss Highly Stable SAW Devices on Quartz", Proc. 40th Annual Frequency Control Symposium-1986, pp. 269-274.

A HIGH PERFORMANCE WAVEGUIDE-COUPLED RESONATOR SAW VCO

Jorgen W. Andersen

Sawtek Inc.
P.O. Box 609501
Orlando, Florida 32860-9501

ABSTRACT

A quadrature balanced amplifier is used with a four pole waveguide-coupled resonator filter to form a feedback oscillator.

The effects of emitter degeneration under large signal (non-linear) conditions are examined. These effects govern the up-mixing of transistor low frequency (1/F) noise and can be used to reduce close-in amplifier residual phase noise. It has been found that a well designed loop amplifier can have lower close-in residual phase noise than the SAW filter itself. Further, it will be shown that degeneration need not degrade the noise floor.

The quadrature amplifier configuration permits matching and gain compression without disturbing the loop interface impedance; this has proven to be of great practical advantage.

Bias noise can negate the best of efforts to improve spectral purity. A low noise active bias scheme is described.

The oscillator is power-efficient, achieving 80 dBc/Hz at 10 Hz offset, with a SAW drive level of +11 dBm. The amplifier is biased at 10 mA per transistor.

INTRODUCTION

The development [1] of the four-pole waveguide-coupled resonator filter (RFW) provided an opportunity to design a high performance VCO with enough tuning range to eliminate the use of ovens in military phase-lock applications. The filter has a nearly linear phase shift of 230 degrees over its 1 dB bandwidth, as illustrated in Figure 1, yet the amplitude roll-off precludes π point ambiguity in the frequency of oscillation. The band center group delay is 3.6 μS.

Power density limitations prohibit the high RF drive levels commonly used in SAW oscillators, however, the low insertion loss and high loaded Q of the RFW make high drive levels unnecessary.

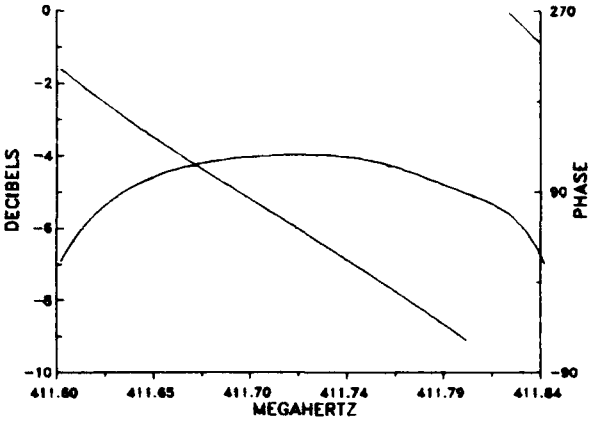


Figure 1 Measured response of the waveguide-coupled resonator filter. Note low insertion loss and phase linearity.

While much has been presented [2]-[3] about desirable amplifier characteristics, there remains a paucity of information about the design of amplifiers suitable for use in SAW oscillators. It is hoped that this work will provide some insight into sustaining device design.

AMPLIFIER DESIGN

The fundamental frequency loop gain of an oscillator in equilibrium is essentially unity. Practical start-up considerations demand that the initial (small-signal) gain be greater than unity. Because there is finite power available, power limiting must occur, mechanizing gain compression.

At this point, a choice must be made between a self-limiting amplifier and some form of external limiting. It can be shown that external limiting results in a larger degradation of the noise floor

CH2965-2/91/0000-239 $1.00 © 1991 IEEE

than self-limiting, therefore it is not considered further.

It is well known that collector (voltage), or output, limiting is highly undesirable since a stored charge is developed in the base circuit and causes serious noise degradation. Because of the foregoing, input (current) self-limiting is the logical choice for SAW oscillators.

The non-linear nature of a self-limiting oscillator can be expressed as a power series expansion of the voltage or current transfer characteristic. The coefficients of the second and higher order terms govern how much 1/F (low frequency) noise is mixed with the carrier. For the common emitter amplifier the expansion is that of e^x, wherein the ratio of the second and higher order coefficients to that of the first order term is substantial (.5, .166, .042, etc.). The relative magnitude of the higher order coefficients can be reduced by using emitter degeneration [4], a form of negative feedback.

Despite the reduction in up-mixed noise, degeneration is not a panacea, since collector current cut-off each carrier cycle becomes the non-linear process mechanizing power limiting and gain compression. The larger the ratio of small signal gain to unity loop gain, ie. gain compression, the larger the non-linearity. While conducting, however, the amplifier is more linear than without degeneration.

To further complicate matters, degeneration degrades the amplifier noise figure. It remains, then, to try to find a good compromise between excess gain, the amount of degeneration, and noise figure.

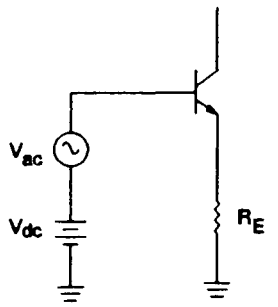


Figure 2 Intrinsic circuit extracted from surrounding circuit reactances.

Consider the simple circuit in Figure 2. Neglecting emitter circuit reactance, which can be canceled at the operating frequency, this can be shown to be the intrinsic circuit embedded within the terminal reactances of an actual device. Since Shockley's equation does not contain frequency variables, the circuit can be used to study the relative effects of gain compression and degeneration. The results of this exercise are similar to those of Frerking [5]. The differences are that numerical techniques are used, thereby avoiding some of his simplifying assumptions, and that the results are computed in terms of gain compression rather than terminal voltages and currents, which would be difficult to use at UHF frequencies.

The relationship of base voltage to collector current is given by,

$$V_b = \frac{KT}{q} \ln \frac{I_c}{I_s} + I_c R_E \qquad (1)$$

Unfortunately, for $R_E > 0$, a closed form solution for I_c as a function of V_b does not exist. However, inspection reveals that this relationship is monotonic and therefore is amenable to numerical solution by a binary search algorithm.

While a polynomial fit to I_c over some range of currents is possible, at some point of gain compression I_c goes to zero for $R_E > 0$, making it impossible to fit the complete voltage-current transfer function accurately. Though the Fourier coefficients of I_c are not equal to the power series coefficients, they are related, for a given amount of gain compression. Thus the harmonic content of I_c is a good indicator of trends and tendencies in the up-mixing of 1/F noise.

While space doesn't permit a complete description of the analysis, it roughly parallels Frerking's. Essentially, the peak value of V_{ac} is incremented from a very small fraction of the quiescent value of V_{dc}. As V_{ac} is increased, V_{dc} is adjusted to set the zero frequency Fourier coefficient of I_c equal to the zero signal value of I_c, ie., the quiescent collector current, which will be held constant by the bias circuits. This process is reiterated for each new peak V_{ac}. The gain compression is simply the ratio of the zero signal transconductance to the fundamental frequency transconductance for a given peak V_{ac}.

Figure 3 is a plot of the Fourier coefficients of I_c versus gain compression for two different values of emitter degeneration.

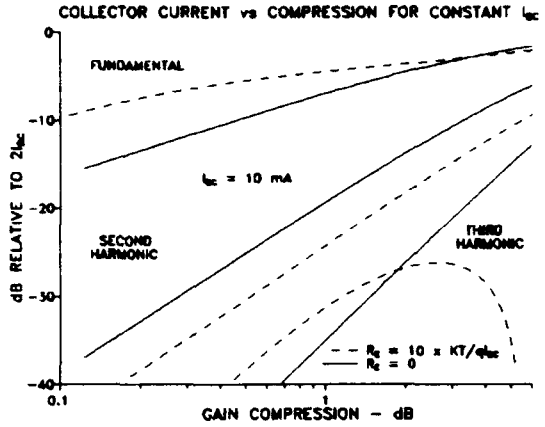


Figure 3 Spectral components of collector current vs. gain compression and emitter degeneration.

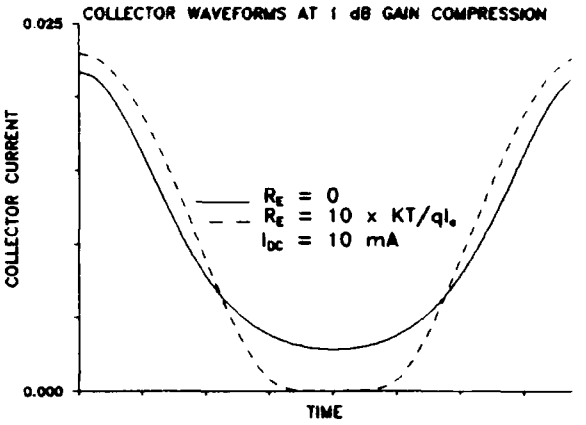


Figure 4 Collector current waveforms for a cosine wave input, used to compute harmonics at 1 dB gain compression.

Inspection of Figure 3 leads to several important conclusions:

1.) Emitter degeneration can result in substantially lower second harmonic content for a given degree of gain compression. Thus less 1/F noise will be upmixed to the carrier frequency.

2.) For the degenerated case illustrated, a higher output level is obtained than in the undegenerated case, for less than 3.5 dB gain compression, thus potentially offsetting any loss in noise figure.

3.) In any case, due to the larger harmonic content, excess loop gain is detrimental to low close-in residual phase noise and should be kept to a minimum.

4.) There is less fundamental frequency amplitude variation vs. gain compression with degeneration, suggesting better amplitude stability vs. temperature as the insertion loss of other circuit elements change.

Examination of the collector current waveforms is also revealing, as illustrated in Figures 4 and 5.

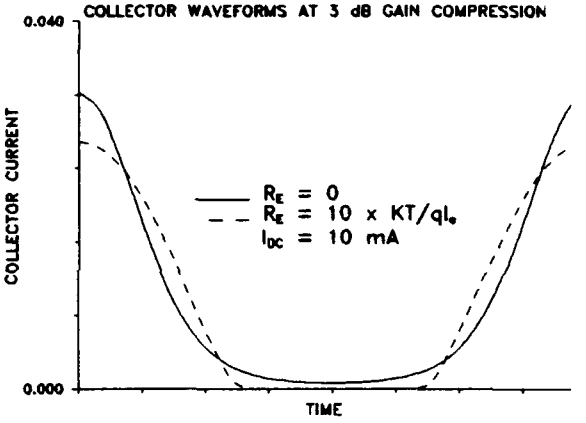


Figure 5 Collector current waveforms for a cosine wave input, used to compute harmonics at 3 dB gain compression.

Conventionally, in an amplifier, maximum output power is obtained with symmetrical limiting, and,

$$Z_{max} = \frac{V_{ce}}{I_{dc}} = 2\frac{V_{ce}}{I_{peak}} \qquad (2)$$

Note in Figures 4 and 5 that $I_{peak} > 2I_{dc}$, therefore $Z_{max} < V_{ce}/I_{dc}$, assuming Z_{max} is constant with frequency. This is a consequence of asymmetrical limiting (input limiting only) as compared to symmetrical limiting (both input and output limiting). One concludes, then, that there is less power avail-

able than suggested by the static load line represented by Eq. 2.

As mentioned earlier, degeneration degrades the device noise figure and must be accounted for in an evaluation of its use. Both the noise current and noise voltage must be identified since degeneration also changes the optimum source impedance of the device.

Figure 6, below, is a noise model of a bipolar junction transistor (BJT) [6].

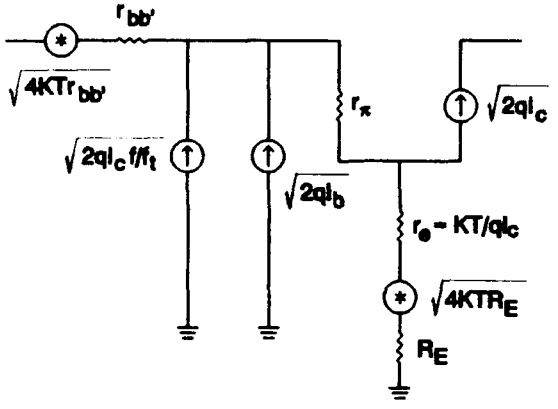


Figure 6 Noise voltage and current generators in a BJT amplifier with emitter degeneration.

The relationships given are normalized to 1 Hz bandwidth and are reasonably accurate up to $\approx F_\tau/3$. The term e_n is the input noise voltage, i_n the input noise current, and R_{opt} is the optimum source impedance. Resistances r_e and r_π are dynamic impedances and are noiseless.

$$e_n^2 = 4KT(r_{bb'} + R_E) + 2qI_c(r_e + R_E)^2 + 2qI_b r_{bb'}^2 + 2qI_c r_{bb'}^2 (f/f_\tau)^2 \quad (3)$$

$r_{bb'}$ = *base spreading resistance*

$$r_e = KT/qI_c$$

$$i_n^2 = 2qI_b + 2qI_c/H_{FE}^2 + 2qI_c(f/f_\tau)^2 \quad (4)$$

$$R_{opt} = \sqrt{e_n^2/i_n^2} \quad (5)$$

$$NF = 10 \log \frac{4KTR_{opt} + i_n^2 R_{opt}^2 + e_n^2}{4KTR_{opt}} \quad (6)$$

Note that the above relationships are expressed in terms of real resistance, meaning that is assumed that the circuit node represented by the input to the *intrinsic* circuit has been tuned real. The reactance in series with the emitter is not considered (though it can easily be added). As mentioned earlier, it is a simple matter to cancel (tune out) this reactance, so the model is quite useful. Furthermore, in practice, it has been found quite accurate as long as the base is tuned real, that is with no effort to tune out the emitter circuit inductance, other than keeping the leads short.

Examination of Eq. 3 reveals the primary effect of degeneration on the noise figure. The emitter circuit is in series with both the collector and base. While the open circuit noise voltage of R_E is a noise contributor, the second term in Eq. 3 dominates. Note that the full collector shot noise is in series with R_E and the resultant increase in input noise voltage is the primary result. At the same time the optimum source impedance rises significantly, since the input noise current remains unchanged.

Recall that a higher noise figure is much more sensitive to noise impedance mismatch, and will depart from the optimum noise figure much more rapidly, than a lower noise figure. Since degeneration raises noise figure to begin with, failure to rematch to the new noise impedance, R_{opt}, will compromise the design.

There is little relationship between R_{opt} and Z_{in}. This is one of the reasons a balanced quadrature amplifier is used. A conjugate impedance match will result in a poor noise figure. The only way to make the input impedance equal the noise impedance is through the use of negative feedback, a considerable complication.

Table I, below, is for a high quality transistor with I_c = 10 mA, r_{bb}' = 10Ω, H_{FE} = 270, F_τ = 4 GHz, and F = 411 MHz.

Table I

R_E	R_{opt}	NF - dB
0	63.4	1.4
r_e	78.6	1.6
$r_e x 10^{1/2}$	118	2.2
$r_e x 10$	254	3.9

As commonly represented, noise figure is a small signal parameter. To assume it remains the same under large signal conditions is erroneous. The shot noise currents are governed by the *mean* bias currents and remain constant in gain compression. While the fundamental frequency transconductance reduces in gain compression, the collector shot noise does not, therefore the large signal noise figure is higher than in the small signal case. Conservative practice is to consider the noise figure increased by the amount of gain compression.

With a knowledge of the peak collector current and the collector to emitter voltage, the power available without output (voltage) limiting can be determined from the maximum load resistance and the fundamental frequency collector current.

One is now in a position to develop a figure of merit consisting of the compressed gain power available, compressed gain noise figure, and the compressed gain harmonic content. Table II is such a formulation, for V_{CE} = 10 volts and the conditions given for Table I. The ratio of the compressed power available (without output limiting) and the compressed noise figure can be thought of as the maximum carrier to noise ratio.

As shown in Table II, there is little or no degradation of the noise floor with degeneration, yet the second harmonic levels can be reduced by up to 7 dB., meaning the frequency translated 1/F (baseband) noise can be reduced by a similar amount.

Figure 7 is a simplified diagram of the quadrature amplifier.

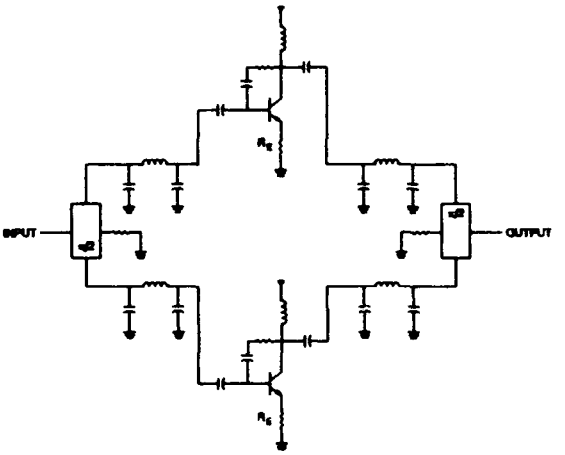


Figure 7 Quadrature amplifier.

The collector to base feedback is used to trim excess gain, since the use of feedback will not

Table II

Degeneration	$R_E = 0$	$R_E = r_e$	$R_E = r_e x 10^{1/2}$	$R_E = r_e x 10$
Gain Compression	Max CNR - dB 2F - dBc			
1 dB	181.3 12.3	182 15.8	182.2 18.1	181.1 19.3
2 dB	181.8 9.2	182.3 12.2	182.1 13.3	180.6 14.2
3 dB	181.3 7.4	181.8 9.9	181.4 11.1	179.8 11.6

degrade the noise floor nearly as much as the use of attenuators in the oscillator loop.

The impedance mismatch of the amplifiers in Figure 7, is isolated from the other loop elements by the directional properties of a quadrature hybrid, as illustrated in Figure 8.

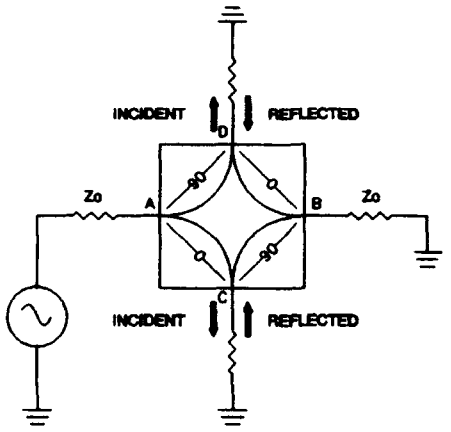


Figure 8 Directional characteristics of a quadrature hybrid.

As long as the reflection coefficients of the two loads are equal, the voltages reflected to port A will cancel since one reflection experiences a net 180 degree phase shift, while the other reflection remains in phase. The reflections to port B add and are absorbed by the termination.

Finally, with respect to the amplifier, it should be noted that the quadrature configuration has 3 dB higher CNR than illustrated in Table II. This is due to the fact that the collector noise adds as a mean square while the collector signals add algebraically in the output hybrid illustrated in Figure 7.

PHASE SHIFTER

The oscillator phase shifter also employs a quadrature hybrid, as illustrated in Figure 9. The tuned circuits exhibit unity reflection coefficient for all frequencies and the reflection angle changes 360 degrees as the frequency varies from zero to infinity. Likewise, at a fixed frequency, the reflection angle changes with changes in the resonant frequency. While finite Q and parasitic reactance preclude a full 360 degree phase shift with tuning, a control range of 240 degrees is easily achieved. The practical advantages include low insertion loss and low insertion loss variation and a constant interface impedance due to the previously discussed properties of the quadrature hybrid.

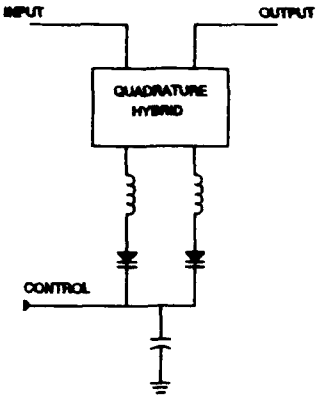


Figure 9 Balanced reflective phase shifter.

BIAS CIRCUIT

Basically, the bias circuit consists of a low noise reference diode, itself biased operationally, a DC amplifier with unity AC gain, and a low noise operational loop around the RF transistor, as illustrated in Figure 10.

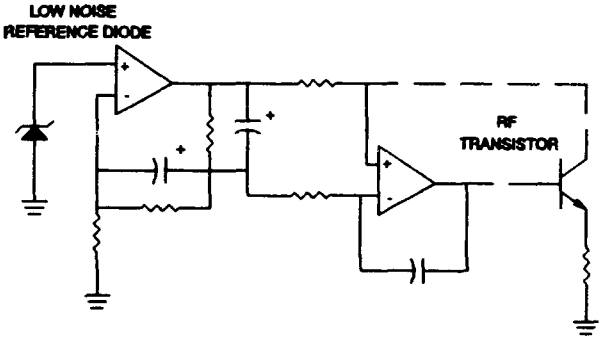


Figure 10 Simplified diagram of the bias circuit.

The common mode rejection ratio of the operational circuits exceeds 110 dB. Note also that the already low reference noise appears as a common mode signal to the second operational amplifier, therefore, the bias noise is essentially that of the second amplifier and the thermal noise of the resistors. The circuit has very high DC loop gain and the resultant bias stability greatly enhances the oscillator

amplitude stability, essential when working with the low excess gain advocated.

OSCILLATOR PERFORMANCE

Experiments have confirmed that degeneration and gain compression have a marked effect on close-in phase noise. Excellent performance at a modest drive level has been achieved, as shown in Figure 11, below.

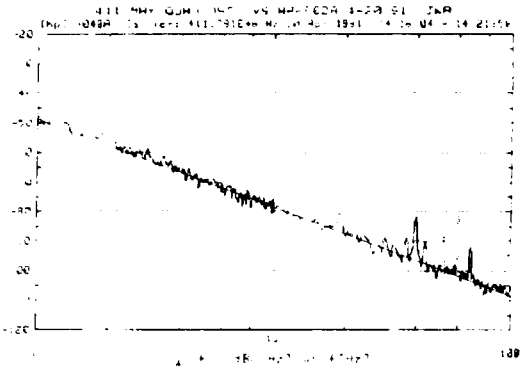


Figure 11 Close-in phase noise of a quadrature oscillator operating at 411 MHz and +11 dBm.

It has been found that the limiting factor in close-in phase noise is the SAW itself. A socketed board was constructed to permit a "plug and test" comparison of SAW performance using the same oscillator circuitry. This was possible because the variations in band center insertion loss and insertion phase were insignificant. The transfer characteristics, including group delay, for each filter were verified with a network analyzer. The group delay variations were also insignificant (approximately 100 nS out of the nominal 3.6 uS). The oscillator phase noise at 1 Hz offset varied from 50 dBc/Hz to 35 dBc/Hz. Subsequently, the residual phase of the filters was measured and found to vary by a similar amount, from 140 dBc/Hz to 125 dBc/Hz at 1 Hz offset. Finally, the residual phase of the amplifier was measured. This proved to be a challenge since the amplifier operates at a lower level than that required by the phase noise test set.

The amplifier residual phase noise was measured by constructing an LC coupled resonator filter and attenuator so that it matched the insertion loss of the SAW filter. A short phase adjust cable was used to set the insertion phase to be equal to that of the SAW filter, then the total insertion loss was carefully measured to make sure the gain compression would be accurately replicated. Finally, the *total* group delay of the oscillator loop was measured. The loop was closed and the phase noise measured on a conventional spectrum analyzer. With a knowledge of the loop group delay, the residual phase noise was computed from the closed loop phase noise. At 100 Hz offset, the computed residual phase was 169 dBc/Hz, which, extrapolated to 1 Hz offset at 10 dB/Decade, is 149 dBc/Hz.

When the amplifier residual phase is converted to oscillator phase noise with the 3.6 uS group delay of the SAW filter, the calculated oscillator phase noise is 56 dBc/Hz at 1 Hz offset, supporting the conclusion that the SAW filter is limiting performance.

It was not possible to demonstrate the full potential of this design since, at the time these test were performed, a second SAW device was not available that had the performance of that shown in Figure 11. Also, not having 411 MHz isolators, buffer amplifiers were used to prevent injection locking, thus degrading the measurement floor. The noise floor computed from measured noise figure, gain compression, and insertion loss is 177 dBc. Nevertheless, the performance illustrated in Figure 12 is quite good for a low power, non-ovenized oscillator operating from unconditioned power supplies.

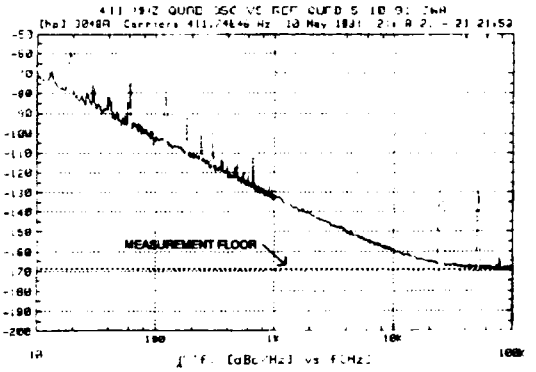


Figure 12 Two oscillator phase noise. Note indicated limit on noise floor.

Figure 13 is a representation of the complete oscillator.

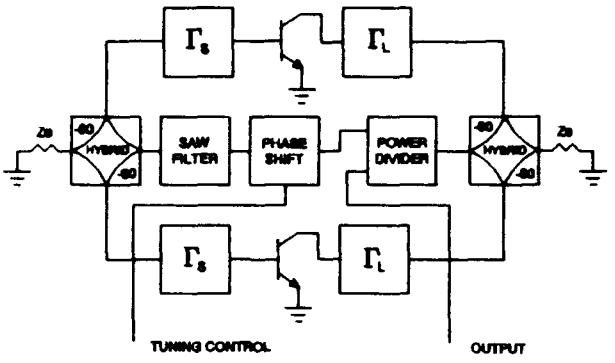


Figure 13 Simplified diagram of the quadrature oscillator.

CONCLUSION

The four pole-pair waveguide coupled resonator filter has been introduced as a frequency control element. Its low loss and high loaded Q make it particularly attractive for use in power limited applications. In furtherance of power conservation, the tuning bandwidth of the RFW eliminates the need for ovens in phase locked sources.

The quadrature balanced amplifier complements the RFW by providing constant impedance in both small and large signal conditions. Amplifier impedance matching does not disturb the loop interface impedance. The quadrature amplifier also provides a 3 dB extension of the individual amplifier's dynamic range.

Contributors to amplifier residual phase noise have been identified and analyzed. Means to minimize noise up-mixing have been described. It has been shown that emitter degeneration need not degrade an oscillator's noise floor.

A phase shifter with a tuning range consistent for use with the four pole RFW has been presented.

A low noise bias circuit has been introduced.

ACKNOWLEDGEMENTS

Deep appreciation for Mr. Thomas O'Shea's support and encouragement of this work is expressed.

Thanks to Mr. Brent Horine for his patient explanations of the RFW, which he brought to commercial practicality.

The author is indebted to Mrs. Jacqueline Hines for her many helpful comments throughout the course of this program and to Mr. Thomas Nowatski for his able assistance in everything from circuit modeling to lab work.

REFERENCES

[1] B. Horine, "Narrowband SAW Filters for IF Applications" in Proceedings of the 44th Annual Symposium on Frequency Control pp. 316-322

[2] G.K. Montress, T.E. Parker, and M.J. Lobota, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components", in Proceedings of the 43rd Annual Symposium on Frequency Control 1989 pp.349-359

[3] G.K. Montress and T.E. Parker, "Design Techniques for Achieving State-of-the-Art Oscillator Performance", in Proceedings of the 44th Symposium on Frequency Control 1990 pp. 522-535

[4] D. Halford, et.al., "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure" in Proceedings of the 22nd Symposium on Frequency Control 1968 pp. 340-341

[5] Marvin E. Frerking, "Crystal Oscillator Design and Temperature Compensation" Van Nostrand Reinhold Company • 1978 Litton Educational Publishing, Inc. LOC Catalog Card # 77-17876 ISBN: 0-442-22459-1

[6] C.D. Motchenbacher and F.C. Fitchen, "Low Noise Electronic Design" • 1973 John Wiley & Sons, Inc. ISBN 0-471-61950-7

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

GRATING ASSISTED SAW DIRECTIONAL COUPLERS

G. GOLAN, I. CHARUVY, M. ZILBERSTEIN, A. SEIDMAN and N. CROITORU

TEL AVIV UNIVERSITY, FACULTY OF ENGINEERING
DEPARTMENT OF PHYSICAL ELECTRONICS, RAMAT AVIV 69978, ISRAEL

ABSTRACT

Physical properties of novel grating assisted Y directional couplers, are examined using the coupled mode theory. A general formalism for the analysis of the lateral perturbed Directional Coupler properties, is presented. Explicit expressions for waveguide key parameters such as, coupling length, grating period and other structural characterizations, are obtained. The influence of other physical properties such as, time and frequency response or cut-off conditions are also analyzed. A plane grating-assisted directional coupler is presented and examined as a basic component in the *integrated acoustic* technology.

I. INTRODUCTION

The analysis of a perturbed channel waveguide plays a significant role in the design of Surface Acoustic Wave (SAW) directional coupler, since channeled waveguides form a basic building block in the *integrated acoustic* technology. The *acoustic directional coupler*, analogous to the optical or the microwave element of the same designation, consists of two or more parallel acoustical waveguide channels, spaced sufficiently close, so that energy may transfer from one channel to another[1,2]. Conventional coupling takes place cumulatively over a substantial length, only when the two acoustical waves, in each of the adjacent channels propagate with the same phase velocity. The amount of energy coupled between these two channels is determined by the modal overlap of each of the modes in its separate channel.[3,4]

The integration of waveguides and distributed reflectors for *integrated acoustics* applications, have been studied by our group.[5] Using the *coupled mode theory*, it was shown that a periodically perturbed, conventional $\Delta V/V$ acoustical waveguide channel, can serve as an acoustical reflector in a cavity structure (ΔV is the relative acoustic velocity reduction due to a deposited metallic strip over a piezoelectric substrate). The general method, given in [5,6], is applicable to any perturbed waveguide structure in the sense that it can analyze and obtain the number of the transversal modes, their cut off frequencies and their reflecting response, without having to solve their corresponding eigenvalue equations.

Lately, novel types of optical grating assisted directional couplers were presented[7,8]. These devices consist of grating assisted, frequency selective coupling, between two forward propagating channel modes of a directional coupler. The waveguides in that case, are *non synchronous*; i.e., each one of them supports a different optical mode, with a distinct phase velocity. Therefore, the only possible way for these two different modes to couple energy, is with the assistance of a perturbed grating structure that provides the necessary phase matching.

There are two main objectives for the following investigation:

1. To develop a basic understanding of the acoustical principles, required in guided surface acoustic wave devices, and 2. To incorporate these principles in the design and fabrication of

CH2965-2/91/0000-247 $1.00 © 1991 IEEE

various kinds of advanced acoustic directional couplers.

In this paper we present first, a study of the spectral properties of a conventional acoustic directional coupler, and the grating assisted co-directional coupler. Then, we compare these properties to those of the contra-directional coupler. In particular we give expressions for the fractional bandwidth and the frequency tunability of these structures. Design formulae for the exact determination of the center frequency of operation and the full coupling interaction length, as well as other device parameters, are also developed.

II.SEPARATEDDOUBLEGUIDEDIRECTIONALCOUPLERS

In principle, there are four fundamental types of directional couplers[9]. The first device configuration is the parallel, double guide directional coupler, shown in Fig. 1(a). This directional coupler is a coupled mode device. An input acoustic signal, launched at port 1, excites a resultant acoustic wave that is changing its amplitude and phase, while propagating down the double guide region. Coupling length, in such a device, is defined as the distance required for a whole amplitude peak to travel from the main guide to its collateral. A second device configuration is the continuous double guide directional coupler, shown in Fig. 1(b). This is again a coupled mode device and it can be considered to be the particular case of the parallel double guide coupler, in which the separation between its two coupling channels tends to zero. This particular configuration is of major interest, since it possesses a considerably wider bandwidth, although its coupling potential between the two adjacent channels is much smaller. The other remaining configurations, shown in Fig. 1 (c,d), are the non coupled mode devices. They will be referred to as Y Junction power splitters. An acoustic excitation at the input port of such a coupler will launch a single mode wave which, then, propagate into the tapered width guide, delivering equal power into both output arms, or tapered horns, respectively. Most of the Y junction power splitters are broadband, since the equal power split between the output ports is a consequence of physical symmetry and hence independent of frequency variations.

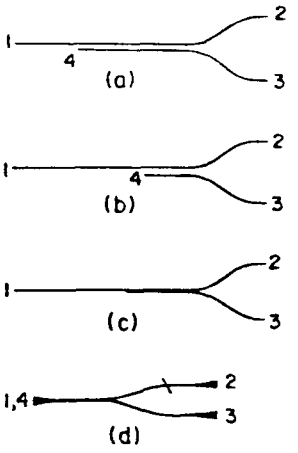


Figure 1. Four basic configurations of conventional SAW directional couplers.
a. The parallel double guide. **b.** The continuous double guide. **c,d.** non coupled mode, Y junction couplers.

Prior to the analysis of the double guide directional coupler, its theoretical model should be established. In this model, the necessary parameters for the design of a coupled mode directional coupler, are defined. As was mentioned before, the investigated structure consists of two parallel perturbed guides, deposited close enough to each other, for appreciable mode coupling.

III. SYSTEM ANALYSIS

The first work, relating to a double guide SAW structure, was published by Tiersten[10]. In addition to his single guide solution, he developed a perturbation technique for calculating coupling length in the weak coupling approximation. In the strong coupling case, however, two additional important features should be mentioned: 1. Coupling length which exhibits a low frequency cut off, under which coupling does not exist anymore. That means that a directional coupler is a frequency dependent device, that can utilize as a frequency filter. 2. Coupling length curves[9], which possesses a local minimum in the region of strong coupling. The coupling length is approximately linear with frequency around this

minimum. Thus, as a rule of thumb, strong coupling devices posses a wider bandwidth than weak coupling ones. A scheme of a basic directional coupler is shown in Fig. 2. The most important design parameters are $G=D/H$ and $n=C/D$, where H is the thickness of the metal guiding strip and C,D are the spacing between the couplers arms and the guides width, respectively. Due to physical symmetry, mode functions in these conventional devices must be symmetric or antisymmetric with respect to the center line. In order to illustrate how the double guide modes are generated, we assume an excited single mode wave, propagating in the main guide, toward the double guide portion of the coupler. At this section it generates a linear combination of the lowest order symmetric and antisymmetric guided modes. Since these two modes propagate with different velocities, the amplitude of the resultant combination is changed in such a way that the energy oscillates back and forth between the channels, as it propagates down the double guide region.[11]

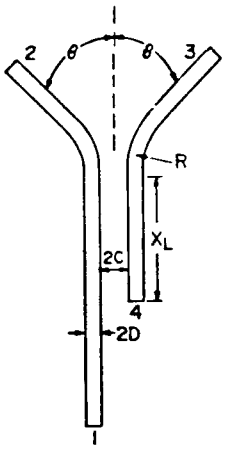


Figure 2. Fundamental structure of a plain directional coupler. β_a and β_b are the the propagation constants of each channel, respectively and separately.

The coupling length L_C, defined as the distance required for an amplitude peak to pass from one guide to its collateral, is:[12]

$$L_c = \pi/(k_s - k_a), \qquad (1)$$

where k_s and k_a are the symmetric and antisymmetric wave vectors, respectively.

In case of a slight difference between the channel parameters, namely dimensional variations or $1/V$ index difference, in such a manner that the individual propagation constants ($\beta_{a,b}$) of each channel are not equal anymore - energy transfer between channels is no longer possible. In order to enhance mode coupling in non identical channels, in this weak coupling approximation, periodic grid perturbations are planted along the wave propagation direction.[13] Energy transfer is then possible in the case of a grid periodicity Λ_r which is equal to the difference between the two separate propagation constants of each individual channel. This condition is expressed as follows:

$$\beta_a - \beta_b = 2\pi/\Lambda_r \ . \qquad (2)$$

Such a grating assisted structure, is presented in Fig. 3a. It consists of two dissimilar parallel acoustical waveguides, which can exchange energy due to the existence of a periodic grating, located at the guides boundaries. The grating assisted matching phenomenon is illustrated in Fig. 4. This figure shows the phased matched coupling between waveguide modes using periodic perturbations. $\beta_{a,b}$ are the mode indices and:

$$K_g = 2\pi/\Lambda_r \qquad (3)$$

is the grating constant of the perturbed section. It is worth mentioning that each guide, if isolated, would have possessed a different fundamental mode.

In this case, the waves are constructively reflected and coupled, due to the grating assisted perturbation structure. This structure serves, therefore, as a β matcher at its characteristic π/Λ_r, where the grooves along the waveguide boundary give an increased amount of slowness wave. The overall field can be approximated by a linear combination of a forward ($S_{(z)}$) and backward ($R_{(z)}$) propagating modes, which are coupled to each other by the applied z perturbation. $R_{(z)}$ and $S_{(z)}$ are the slowly varying amplitudes of the forward and backward acoustical

modes, propagating in each of the waveguide channels, respectively.

The coupling efficiency curve as a function of the exciting frequency, at $z=L$, is shown in Fig. 6. This figure demonstrates that the acoustic coupler behaves as an acoustic filter, owing its frequency selectivity to its strong dependency on the operating frequency. Calculated bandwidths in three cases examined, illustrated in Fig. 6, are equal. However, the required center frequency, for maximal coupling efficiency, is a function of the specific grating period-Λ_r.

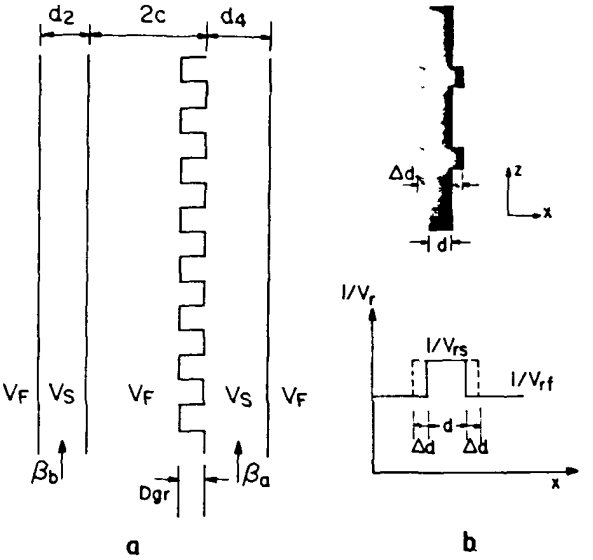


Figure 3. **a.** General structure of a grating assisted directional coupler. V_f and V_s are the SAW velocities of the fast surface and the slow channel, respectively. **b.** A perturbed waveguide section and its $1/V$ symmetric profile.

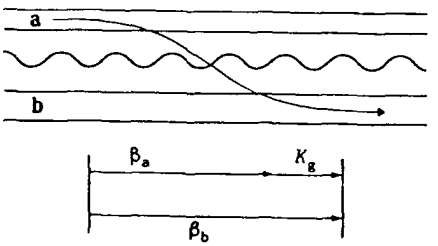


Figure 4. Phase match coupling between waveguide modes using periodic coupling. $\beta_{a,b}$ are the mode indices and K_g is the grating constant phase matcher.

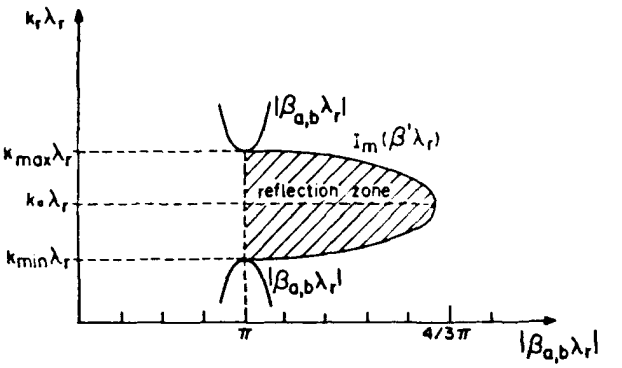


Figure 5. β behavior around the center frequency f_0 revealing the range of frequencies within where coupling occurs.

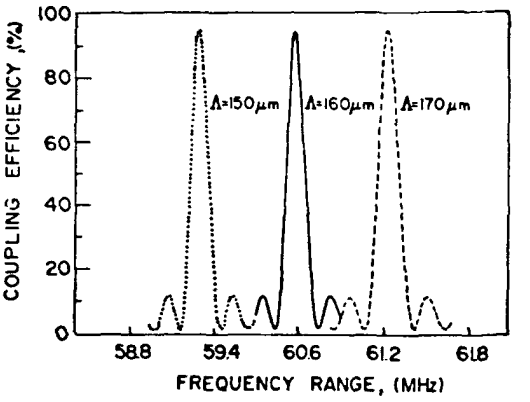


Figure 6. Calculated coupling efficiency of the proposed grating assisted co directional coupler for three different grating periods.

IV. DEVICE DESIGN CRITERIA

The above formalism was applied to the case of grating assisted directional coupler, made of two dissimilar metal strip channels, as illustrated in Fig. 3. The presented formalism is a one dimensional, first order scalar approximation, since its resulting errors, in this case, are negligible[15,16] (a technique for converting from the isotropic case to the anisotropic case is widely demonstrated in ref. 16). It should be emphasized that the whole calculation is based on a separate modal analysis of each of the participating waveguides in the structure, therefore, all the mathematical calculations are straightforward with minimum complexity.

The design of a practical coupler should provide for: 1. Efficient power transfer around the required center frequency (in order to minimize the coupler insertion loss). 2. Low back reflected power from a secondary channel to its primary, in order to assure low cross talk in the system. For instance, the spectral response of our proposed directional coupler, with a center frequency of 60 MHz has been examined for a particular set of design parameters (Fig. 3). Using a simple modal analysis, in order to match between the different propagation constants of each of the separated channels, for a center frequency of 60 MHz, resulted in, a grating period of 166 μm (Fig. 7). This grating periodicity is in inverse proportion with the frequency, within a given spectral range of up to 61.2 MHz (2% of the center frequency) (Fig. 7). The length of the proposed device was taken as 20 mm and its grating periodicity as 10 μm. Calculations of the required grating period and the device full coupling length as a function of the excitation frequency, within its bandwidth range are presented in Fig. 8. An almost linear dependency of the coupling length, within the bandwidth range, is clearly observed. It was found though, using the present formalism, that for a center frequency of 60 MHz, full coupling length is obtained after 9.6 mm. Any shorter coupler than that value, will result in a decrease in the coupling efficiency, followed by bandwidth expansion in the same proportion.

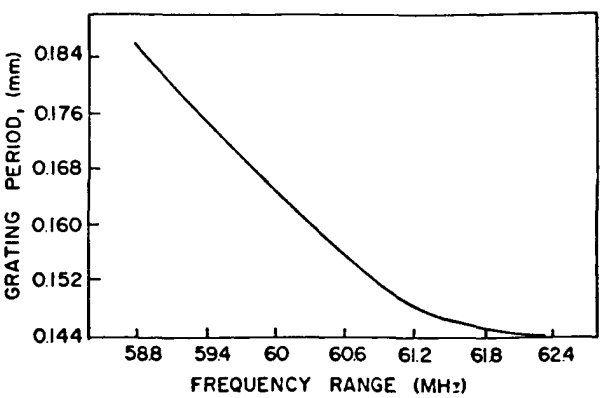


Figure 7. Grating period as a function of the required center frequency.

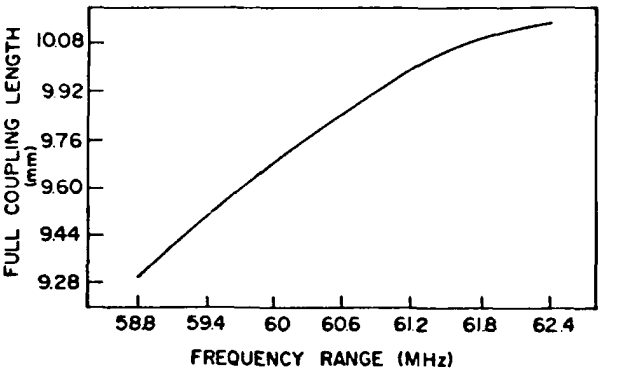


Figure 8. Full coupling length in a grating perturbed directional coupler as a function of the excitation frequency.

Coupling efficiency for the three proposed grating periods was calculated and the results are presented in Fig. 6. As expected, the bandwidth shown by these three curves is quite narrow (4%) and fits the calculated value, for a full coupling length. One way to increase the coupling efficiency of the proposed device is possible, by slightly reducing the full coupling length to a more practical value of 5 mm. This can be done by increasing the coupling coefficient through either an increase in the grating perturbation depth, or a decrease in the spacing between the coupled channels. Calculation of the full coupling length as a function of the desired grating depth, with the channels separation as a parameter, is illustrated in Fig. 9. Device parameters for optimum performances of various grating assisted directional couplers, can be derived From Fig. 8.

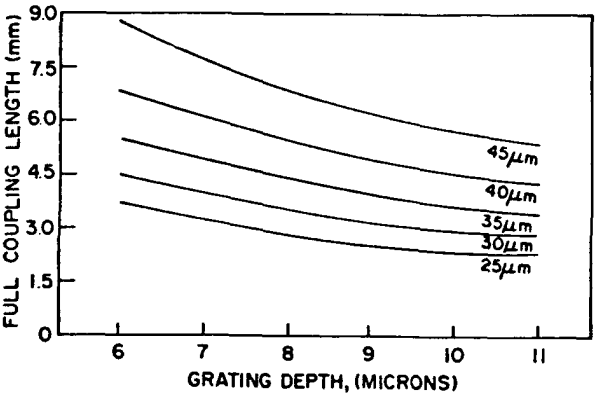


Figure 9. Full coupling length, calculated as a function of grating periodicity with the core to core separation between channels as a parameter.

Calculated values of the required grating constant, Λ_r, to assure coupling between two adjacent channels, spaced 2C apart, are presented in Figure 10. They clearly demonstrate that, as the two channels are being brought together, the difference in their wave vectors are increased so that, for full energy coupling, a shorter grating period is required (thus a decrease in the perturbations repetition rate). Above the spacing of 35 μm, the grating period parameter remains constant, coincidentally, with a large coupling coefficient decrease. Another interesting parameter is the coupling length for maximum power transfer from one channel to another. The expression for L_{opt} for full coupling, is reduced to a simpler form by nullifying the mismatch factor, generated by the wave vector differences between the guided channels. In the case of such a zero grating mismatch, the coupling length is given by:

$$L_{opt} = \frac{\pi}{2d\beta_{ab}\beta_{ba}} \quad . \qquad (4)$$

Calculations were carried out analytically for a grating period of 10 μm. Fig. 11 illustrates the calculated coupling length L, as a function of the spacing 2C between the adjacent channels. The value of the coupling length, L, was only approximated, by neglecting the mutual influence of each of the propagating constants on each other. Fig. 11 demonstrates a very interesting phenomenon, explained already by Marcuse [14] for optical directional couplers. When a zero crossing of one of the propagating modes occurs next to one of the core boundaries, coupling length increases drastically, to infinity.

Finally, the influence of a perturbed channel width on the coupling length was examined. The resultant curve is presented in Fig. 12. According to this curve, at a certain channel width of 50 μm, the coupling length gets its maximal value. It is also clearly seen that for thin channels, coupling length is small, since the grating amplitude is getting larger compared to the channel width.

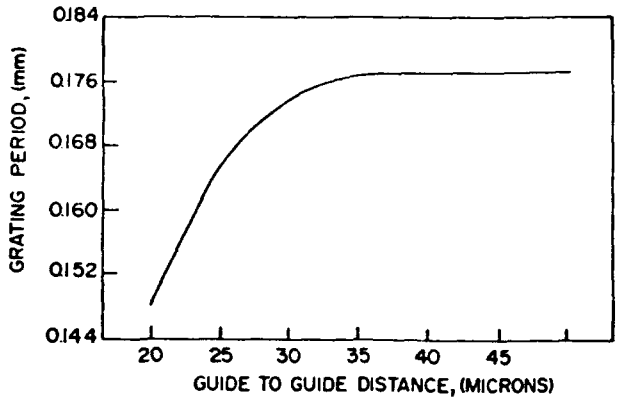


Figure 10. Grating period as a function of the distance between coupling channels, for a center frequency of 60 Mhz.

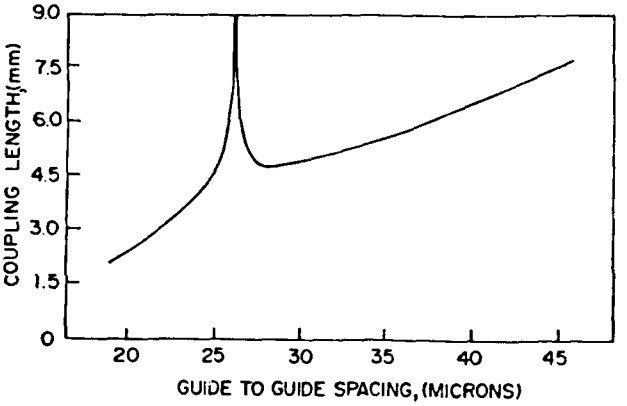


Figure 11. Coupling length as a function of guide to guide spacing.

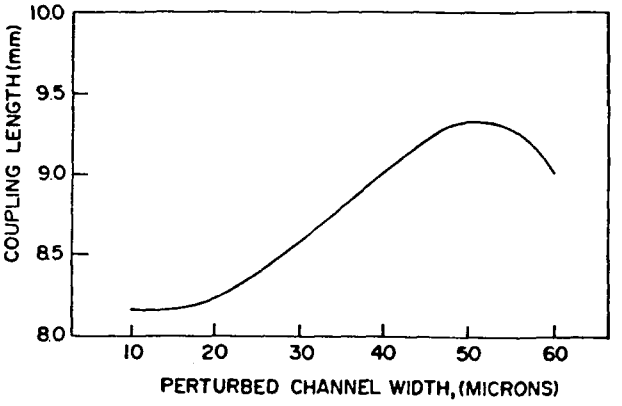


Figure 12. Coupling length as a function of the perturbed channel width.

CONCLUSIONS

A formalism for analyzing and laying out of acoustic, grating assisted directional coupler filters is presented. This formalism enables to obtain filter characteristics, such as center operating frequency, bandwidth, tunability, as well as dimensional parameters, concerning any device design. A practical example was analyzed and yielded parameters for a grating assisted directional coupler.

As a conclusion, this formalism may serve for a general design tool for almost any geometry, and give the criteria for grating parameters and location of grating assisted directional couplers.

REFERENCES

1. S.L. Chang, "A coupled mode theory for multi waveguide systems satisfying the reciprocity theorem and power conservation", **J. Lightwave Technology**, Vol. LT-5, No. 1, p. 174, 1987.

2. A. Yariv, "Coupled mode theory for guided wave optics", **I.E.E.E. J. Quantum Electronics**, Vol. QE-9, p. 919, 1973.

3. A. Hardy and W. Streifer, "Coupled mode theory for parallel waveguides", **J. Lightwave Technology**, Vol. LT-3, p. 1135, 1986.

4. H.A. Haus, W.P. Huang, S. Kawakami and N,A. Whitaker, "Coupled mode theory for optical waveguides", **J. Lightwave Technology**, Vol. LT-5, No.1, p. 16, 1987.

5. G. Golan, G. Griffel, A. Seidman and N. Croitoru, "Distributed feedback guided SAW micro resonators", **J. Appl. Phys**. 66(3), p. 1428, 1989.

6. G. Golan, G. Griffel, A. Seidman and N. Croitoru, "Modal analysis and cut off conditions of SAW waveguides", **I.E.E.E. Trans. on Ultrasonics Ferroelectrics and Frequency Control**, Vol. 35, No. 4, p. 503, 1988.

7. N.K. Dutta, A.B. Piccrilly, T. Cella and R.L. Brown, "Electronically tunable distributed feedback laser", **Appl. Phys. Letters**, Vol 48, No.22, p. 1501, 1986.

8. Y. Kotaky, M. Matsuda, H. Ishikawa and H. Imai, "Tunable distributed Bragg reflector laser with wide tuning range", **Electronics Letters**, Vol. 24, p. 503, 1988.

9. L.R. Adkins, and J. Hughes, "Investigations of SAW directional couplers", **I.E.E.E. Trans. on Sonics and Ultrasonics**, SU-25, p. 45, 1972.

10. H.F. Tiersten, "Elastic surface waves guided by thin films", **I.E.E.E. Trans on MTT**, Vol. MTT-17, p. 904, 1969.

11. H. Kogelnik, "Theory of dielectric waveguides", **Integrated Optics** , T. Tamir, 2nd. Ed. New York, Springer Verlag, Ch. 2, 1979.

12. H. Stall and A. Yariv, "Coupled mode analysis of periodic dielectric waveguides", **Optical Communication**, Vol. 8, p.5, 1973.

13. G. Golan, G. Griffel, A. Seidman and N. Croitoru, "Novel types of SAW micro reflectors", **J. Appl. Phys.**, 67(11), p. 6675 1990.

14. D. Marcuse, "Directional couplers made at nonidentical assymetric slabs" Part 2, **J. Lightwave Technology**, Vol. LT-5, No. 2, p. 268, 1987.

15. E. Marcantili, "Improved coupled mode equations for dielectric guides", **I.E.E.E. J. Quantum Electronics**, Vol. QW-22, No.6, p. 988, 1986.

16. R. V. Schmidt and L.A. Coldern "Thin films acoustic surface waveguides on anisotropic media". I.E.E.E. Trans on Sonics and Ultrasonics, Vol. SU-22, 1975.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

QUARTZ RESONATOR SENSORS IN EXTREME ENVIRONMENTS

ERROL P. EERNISSE AND ROGER W. WARD
Quartztronics, Inc.
1020 Atherton Drive, Bldg. C
Salt Lake City, UT 84123

ABSTRACT

This work presents new data on the tensile failure of quartz obtained from testing miniature force sensors of the double-ended tuning fork configuration. Also presented are Dauphiné twinning threshold data for temperatures ranging from room temperature to 175 °C, which were obtained from a number of AT-cut quartz crystal pressure transducers. One significant feature of the present twinning results is that the pressure transducers twin when subjected to a biaxial stress pattern. Previously published quantitative twinning results were obtained in experimental configurations designed for uniaxial stress. Calculations of the Gibbs Free Energy Difference between the two possible twin configurations have been performed for biaxial stress patterns for comparison to previously published uniaxial stress calculations. The calculated results show quantitatively that biaxial stress-induced twinning and uniaxial stress-induced twinning are consistent with one another in terms of threshold energy.

The results are of interest to the Frequency Control community for evaluation of brittle failure and Dauphiné twinning during resonator processing, mounting, sealing, and shock and vibration. In addition, calculations of the Gibbs Free Energy Difference for Dauphiné twinning in the SC-cut are presented for uniaxial as well as biaxial stress patterns for comparison to the AT-cut and X-cut. These results are useful for estimating the susceptibility of the SC-cut to twinning during manufacture and use.

INTRODUCTION

Considerable information has been obtained on the performance of quartz crystals used as sensors in extreme environments of temperature and stress. The results are useful for the Frequency Control community, particularly with regard to manufacturing yield losses due to brittle fractures and/or Dauphiné twinning during plating and mounting of resonator blanks and during sealing of crystal packages. Tensile failure of quartz, of concern not only during manufacture of mounted and sealed resonators, but also during shock and vibration, occurs at stress levels considerably lower than compressive failure [1-4].

Ferrobielastic twinning has been studied for years and is a constant concern in several phases of resonator fabrication and mounting [5-13]. The present experimental twinning results are supported by calculations of the Gibbs Free Energy Difference between the two twin states. Additional calculations of the Gibbs Free Energy Difference are presented for the SC-cut for comparison to the AT-cut and the X-cut used in published twinning studies. Both uniaxial stress conditions and biaxial stress patterns are included in the theoretical results to quantitatively compare the present experimental situation to earlier work. The fact that the present biaxial stress study is consistent with earlier uniaxial results confirms the use of the Free Energy Difference as a theoretical tool to predict twinning thresholds in complicated stress pattern situations.

TENSILE FAILURE

The emergence of flexurally vibrating quartz force sensors [14] has emphasized tensile failure under conditions where quantitative measurements of failure stress levels can be obtained. Figure 1 shows three different types of such force sensors. The middle device in Fig. 1 is a double-ended tuning fork. This device, as well as the three-beam device at the bottom of Fig. 1, is fabricated by photolithographic techniques and etching. The length of the tines is in the Y crystal direction. A study of the tensile failure of the two-beam device of Fig. 1 was performed. Figure 2 shows the result of testing 7 devices to failure. A bar chart is used in Fig. 2 to emphasize the apparent bimodal distribution of the results. Devices 1 and 5 failed at a level approximately one-half of the failure level of the other five devices. The bimodal distribution suggests that devices 1 and 5 failed with only one tine holding the ultimate tensile failure load. This would occur if one tine had a flaw such as a deep etch pit or etch channel that caused that tine to fail prematurely before the second tine. The remaining tine then carried the full load. The five devices with the larger failure load had both tines carrying the load until failure. The destruction of the devices by brittle failure was extensive enough that

CH2965-2/91/0000-254 $1.00 © 1991 IEEE

examination of the remaining structure provided no concrete conclusions as to whether one tine failed before the other.

Figure 3 shows the calculated failure stress in MPa if devices 1 and 5 are assumed to have only one tine holding the ultimate failure load, while the remaining five devices have two tines holding the ultimate failure load. The stress is the load/tine divided by the tine cross sectional area. The results shown in Fig. 3 appear to be consistent with this assumption. The failure level is 130 MPa minimum with an average of 140 MPa. Published results for failure in the Y direction are 83-91 MPa for small specimens used in geophysical studies [1] and 57-110 MPa for photolithographically etched single beams [3]. The presence of water has been shown to weaken quartz in tensile failure [2].

The results in Figs. 2 and 3 tend to be somewhat higher than the earlier published results. The devices tested were etched using photolithographic procedures from wafers with the major surfaces of the quartz wafer etched before metallization. Thus, the strength can be expected to be high due to the surfaces on all four sides of the tines being etched (the original etched wafer surfaces on the top and bottom and the two sides etched during the photolithographic fabrication).

The published compressive failure levels range from 2200 to 2685 MPa [2,4]. These numbers are much larger than the values discussed above for tensile failure. Such behavior is typical for brittle crystal materials. The fact that the tensile failure level is so much lower than the compressive failure level must be taken into consideration in designing sensors. Figure 4 shows an AT-cut pressure transducer structure designed such that external hydrostatic pressure causes no tensile stress anywhere in the structure. Finite Element Analysis was used to confirm the absence of tensile stress. Devices such as that depicted in Fig. 4 have been proof tested to 352 MPa external hydrostatic pressure at room temperature. The structure in Fig. 4 amplifies the external pressure such that the biaxial stress in the resonator is 1.4 times the external pressure. Thus, at proof pressure, the resonator in Fig. 4 is subjected to 493 MPa of biaxial stress.

TWINNING

The large compressive stress levels achievable in the structure of Fig. 4 can exceed the threshold for Dauphiné (electrical) twinning. Since quartz resonator sensors are used in oscillator circuits, twinning is catastrophic and must be avoided. Numerous studies have examined twinning thresholds for uniaxial stress. The published threshold level for room temperature twinning ranges from 400 to 550 MPa [5,11,12], so the proof pressure test of Fig. 4 mentioned above is at the threshold of twinning if there is a correlation between uniaxial stress-induced twinning and biaxial stress-induced twinning. The present study addresses this issue of biaxial vs. uniaxial stress-induced twinning.

Biaxial stress occurs in resonator processing when the quartz blank is subjected to thermal heat input to the major surfaces during sawing, lapping, polishing, and metal deposition. In addition, large temperature gradients across the surface of the blank, such as picking up a hot blank with cold tweezers, creates biaxial stress since the blanks are thin. Several published works use lasers to heat quartz blanks in order to study twinning [6,7,13]. It has been suggested that all of the published work using rams to generate uniaxial stress have actually generated biaxial stress because of the difference in Poisson's ratio for the metal platens and the quartz test pieces and the resulting friction and scuffing [9]. Thus, biaxial stress-induced twinning is of considerable interest.

Twinning induced by applied stress is termed ferrobielastic twinning [5-6]. The effect occurs because the two twin states of quartz have the same elastic compliance tensor components s_{ijkl} except for the s_{1123} component. When calculating the Gibb's Free Energy Difference ΔG between the existing state and the twin state (called here the Twinning Energy), other workers [5-7,11-13] have arrived at the expression

$$\Delta G = 4\, s_{1123}(T_{11}T_{23} - T_{22}T_{23} + 2\, T_{12}T_{13}). \qquad (1)$$

Here, T_{ij} is the applied stress expressed in the crystal axes system x_i. We use T''_{ij} as the applied stress in a right-handed coordinate system x''_i rotated by the angle ψ about the plate normal. The plate axis system is designated x'_i with x'_2 (or y') as the plate normal. When ΔG is negative, twinning is encouraged. The compliance term s_{1123} is positive for the 1949 IRE Standard [15] (AT-cut is +35° 15′) and negative for the 1978 IEEE Standard [16] (AT-cut is -35° 15′). We choose here the 1949 Standard for a positive s_{1123} so that the sign of ΔG is related to the transformation of the stress tensor T''_{ij} to T_{ij} in the crystal axes and so that our calculated results for ΔG will compare directly with the majority of the published work on twinning. Those published

results using the 1978 Standard can be compared by reversing the direction of the x′ axis.

Previous workers solved Eq. 1 for a uniaxial stress $T"_{33}$. Results for the Twinning Energy for uniaxial stress are shown in Fig. 5 for the X-cut, the AT-cut, and the SC-cut. These results, and all subsequent results, are plotted for a unit stress. All plots are shown to the same scale. Twinning occurs more easily for larger magnitudes of Twinning Energy. All plots use a solid line for negative Twinning Energy (twins occur) and dashed lines for positive Twinning Energy (no twins occur).

As shown by others for the X-cut [8,11,12], twinning occurs for uniaxial stress applied in a direction between the y′- and z′- axes, as seen in Fig. 5 (a). The AT-cut results in Fig. 5 (b) exhibit the four-leaf clover pattern commonly seen in various experiments [7,13]. The SC-cut also exhibits a four-leaf clover shape experimentally, as predicted in Fig. 5 (c), although experimentally one lobe was found by other workers to be larger under some conditions [13].

Note that, since the scales are the same for Figs. 5 (a)-(c), twinning can occur at the same general stress levels for the X-cut, AT-cut, and SC-cut along the most sensitive directions because the Twinning Energies are comparable.

Also note that uniaxial stress applied along the y′- axis of the X-cut (the y- crystal axis) can not induce twinning. This explains the fact that twinning has never been observed in any applications of the flexural devices shown in Fig. 1.

When the biaxial stress is uniform, $T"_{11} = T"_{33}$ (or $T"_{xx} = T"_{zz}$). This case is shown in Fig. 6 for the AT-cut and the SC-cut. The X-cut has a zero Twinning Energy for this case and is not shown in Fig. 6. First, consider the AT-cut used for pressure transducers. In comparing Fig. 6 (a) with Fig. 5 (b), we see that twinning can occur for approximately the same stress level for uniaxial stress applied in the most sensitive direction between the x′ and z′ axes and biaxial stress applied in any direction (the scale for the two plots is the same). Thus, twinning threshold results obtained from circular AT-cut pressure transducers is directly comparable to results obtained with uniaxial stress. This will be important later on when quantitative twinning threshold data will be presented.

Remembering that the scales are the same for Fig. 6(a) and Fig. 6(b), we note that the SC-cut is more difficult to twin with uniform biaxial stress because the Twinning Energy is much smaller for the same stress level.

Note in Fig. 4 that there are flats on the exterior of the structure. These flats are introduced intentionally to break up the circular symmetry. A non-uniform stress occurs in the resonator because of the presence of the flats. This non-uniform stress is used to improve the temperature dependence of the scale factor of the pressure transducer [17]. The typical non-uniformity obtained in the resonator is a 2:1 ratio of the in-plane resonator stress along the normal to the flats (chosen as the x- crystal axis in actual devices) compared to the in-plane resonator stress perpendicular to the flat normal (along the resonator z′- axis). The Twinning Energy is calculated for comparison to the uniform biaxial case of Fig. 6 and to the uniaxial case of Fig. 5 by selecting $T"_{xx} = 1.333$ and $T"_{zz} = .667$, so that the average stress is the same as the uniform biaxial case with $T"_{xx} = T"_{zz} = 1$.

The calculated results are shown in Fig. 7 for the X-cut, AT-cut, and SC-cut. Once again, the scale is the same for Figs. 5, 6, and 7. Note that the general shape and symmetry in Fig. 7 follows the corresponding plot for uniaxial stress in Fig. 5. In the case of the AT-cut pressure transducer with the flat normal along the x-crystal axis, the value of Twinning Energy in Fig. 7 (b) is larger by only 8.8% as compared to the uniform biaxial case in Fig. 6 (a). Thus, the use of flats to improve transducer temperature performance does not seriously degrade the full-scale pressure limits set by twinning considerations.

EXPERIMENTAL TWINNING RESULTS

A number of transducers have been tested to the twinning threshold by increasing applied pressure to the point where the resonator is found to have irreversibly twinned. By knowing the amplification of the particular structure used, the resonator stress at which twinning occurred can be calculated from the applied pressure. Experiments were carried out at 25°C and 175°C. The results are plotted in Fig. 8 as twinning threshold in MPa vs. temperature. One sample was twinned at 25°C and three samples at 175°C. Also included in Fig. 8 are the results of J. Laughner [12] measured with uniaxial stress applied to X-cuts at 45° between the y- and z- crystal axes (force direction shown with arrows in Fig. 5 (a) and Fig 7 (a)). As seen in Fig. 8, the results obtained here with nonuniform biaxial stress compare quantitatively in magnitude and temperature dependence with the earlier uniaxial results.

The fact that the present nonuniform biaxial stress results compare quantitatively with the earlier uniaxial stress results is explained by the equivalence of the Twinning Energy for the two cases. The ratio of the Twinning Energy for the nonuniform biaxial stress case in Fig. 7(b) to the Twinning Energy for the uniaxial case at 45° in Fig. 5 (a) is 1.34. The Twinning Energy increases as the square of the stress, so the expected ratio for stress threshold should be 1.16, with the nonuniform biaxial stress case twinning more easily. The difference in the two cases, 16%, is within typical experimental scatter seen for both the present nonuniform biaxial stress results and the published uniaxial stress results (as shown in Fig. 8 by the vertical bars).

As seen in Fig. 8, the nonuniform biaxial stress case may actually twin at a higher stress level than the uniaxial case, just the reverse of the prediction of the Twinning Energy calculations. This could be attributed to the pressure transducers having a continuous boundary at the resonator perimeter integral with the outer shell, so no surface damage is available to nucleate twins. In contrast, the uniaxial experiments use a sample with a free surface, and it has been shown that the presence of surface damage lowers the stress levels required for twinning [10].

CONCLUSIONS

Two failure mechanisms seen in quartz resonator sensors have been discussed. The tensile failure results presented here are comparable to other published results, but the present failure stress levels tend to be higher than published values. The difference may be due to the fact that the present samples are etched on all sides, so there is no surface damage that would lower the failure stress level.

The ferrobielastic twinning experimental results compare directly to published results for stress threshold vs. temperature. The fact that the present results, which are obtained with nonuniform biaxial stress, compare to published results obtained with uniaxial stress is explained by calculations of the Twinning Energy for the two cases.

The results of both the Twinning Energy calculations and the experiments contribute to the knowledge base of the Frequency Control community. Calculations for the SC-cut for biaxial stress cases are included for such purposes.

REFERENCES

1. R. B. Sosman, The Properties of Silica. New York: Am. Chem. Soc. Monograph, 1927, pp. 481-482.

2. C. H. Scholz, "Static Fatigue of Quartz," Jour. of Geophysical Research, vol. 77, No. 11, pp. 2104-2114, 1972.

3. T. Ueda, et. al., "Quartz Crystal Micromechanical Devices," in Proc. Intl. Conf. on Solid-State Sensors and Actuators-Transducers '85, 1985, pp. 113-116.

4. R. W. Ward, "The Constants of Alpha Quartz," in Proc. 38th Annual Symp. on Frequency Control, pp. 22-31, 1984.

5. K. Aizu, "Second-Order Ferroic State Shifts," Jour. Phys. Soc. Japan, Vol. 34, No. 1, pp. 121-128, 1973.

6. R. E. Newnham and L. E. Cross, "Secondary Ferroics and Domain-Divided Piezoelectrics," Ferroelectrics, Vol. 10, pp. 269-276, 1976.

7. T. L. Anderson, et. al., "Laser-Induced Twinning in Quartz," Phys. Stat. Sol., Vol. (a) 37, pp. 235-245, 1976.

8. J. W. Laughner, et. al., "Acoustic Emissions From Stress-Induced Dauphiné Twinning in Quartz," Phys. Chem. Minerals, Vol. 4, pp. 129-137, 1979.

9. J. Markgraaff, "Elastic Behavior of Quartz During Stress Induced Dauphiné Twinning," Phys. Chem. Minerals, Vol. 13, pp. 102-112, 1986.

10. D. A. Cocuzzi and J. W. Laughner, "Effects of Surface Abrasion and Impurity Levels on Stress-Induced Dauphiné Twinning in Alpha Quartz," in Proc. 43rd Annual Symp. on Frequency Control, pp. 617-621, 1989.

11. T. L. Anderson, et. al., "Coercive Stress for Ferrobielastic Twinning in Quartz," in Proc. 31st Annual Symp. on Frequency Control, pp. 171-177, 1977.

12. J. Laughner, "Induced Dauphiné Twinning in Quartz," Ph.D. Thesis, Penn. State Univ., 1982

13. K. L. Blisnuk, et. al., "A Comparison of Laser-Induced Twinning of Quartz Based on Characteristics and Supplier," in Proc. 44th Annual Symp. on Frequency Control, pp. 251-258, 1990.

14. E. P. EerNisse, et. al., "Survey of Bulk Resonator Sensor Technologies," IEEE Trans. Ultrasonics, Ferrolelectrics, and Frequency Control, pp. 323-330, 1988.

15. "Standards on Piezoelectric Crystals, 1949," Proc. I.R.E., Vol. 37, pp. 1378-1395, 1949.

16. "IEEE Standard on Piezoelectricity," IEEE Std 176-1978, The Inst. of Electrical and Electronics Engineers, Inc., 345 East 47th St., New York, NY 10017, 1978.

17. E. P. EerNisse, "Theoretical Modeling of Quartz Resonator Pressure Transducers," in Proc. 41st Annual Symp. on Frequency Control, pp. 339-343, 1987.

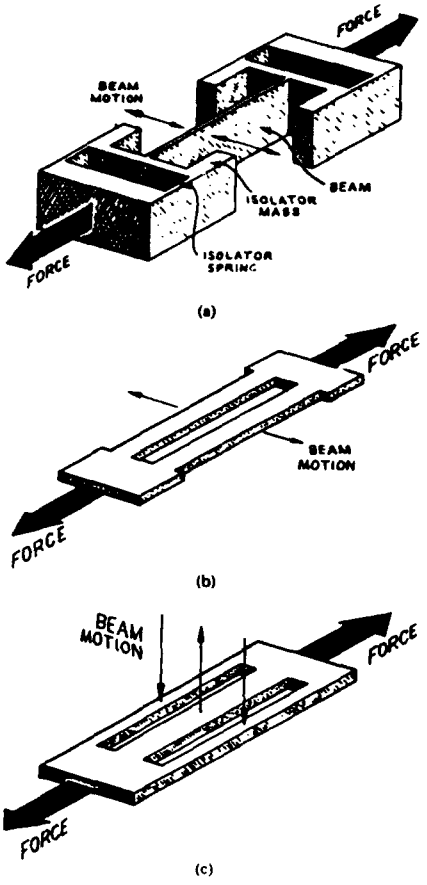


Figure 1. Examples of flexural mode force sensors. (a) Single-beam device. (b) Double-beam device. (c) Triple-beam device.

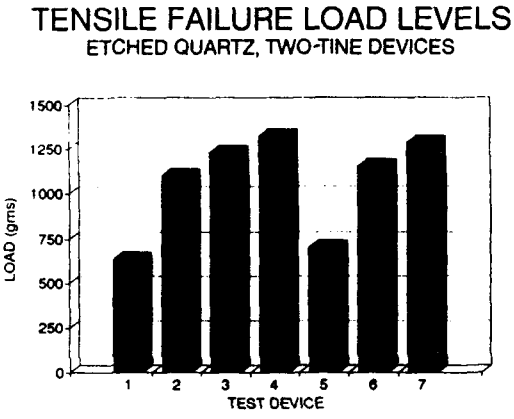


Figure 2. Failure load for seven devices of the two-beam configuration of Fig. 1 (b).

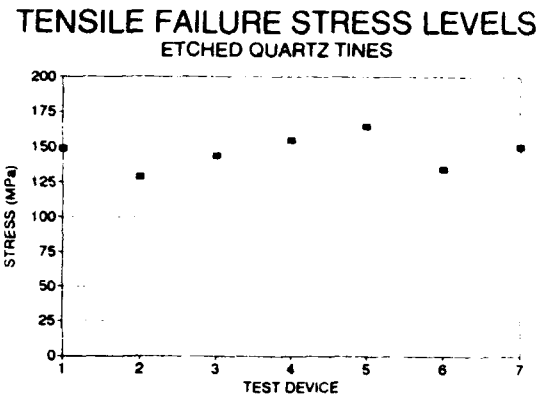


Figure 3. Failure stress level computed from data of Fig. 2.

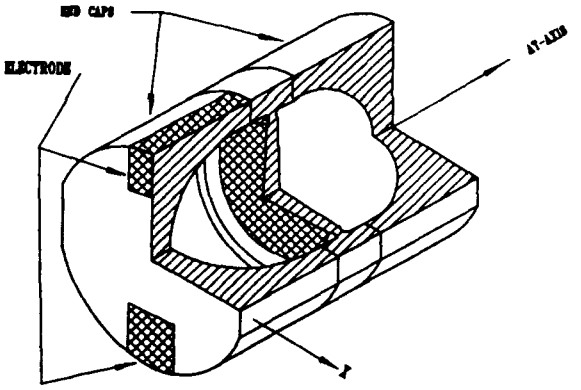


Figure 4. AT-cut pressure transducer configuration designed to avoid tensile stress in the quartz when subjected to high external pressures.

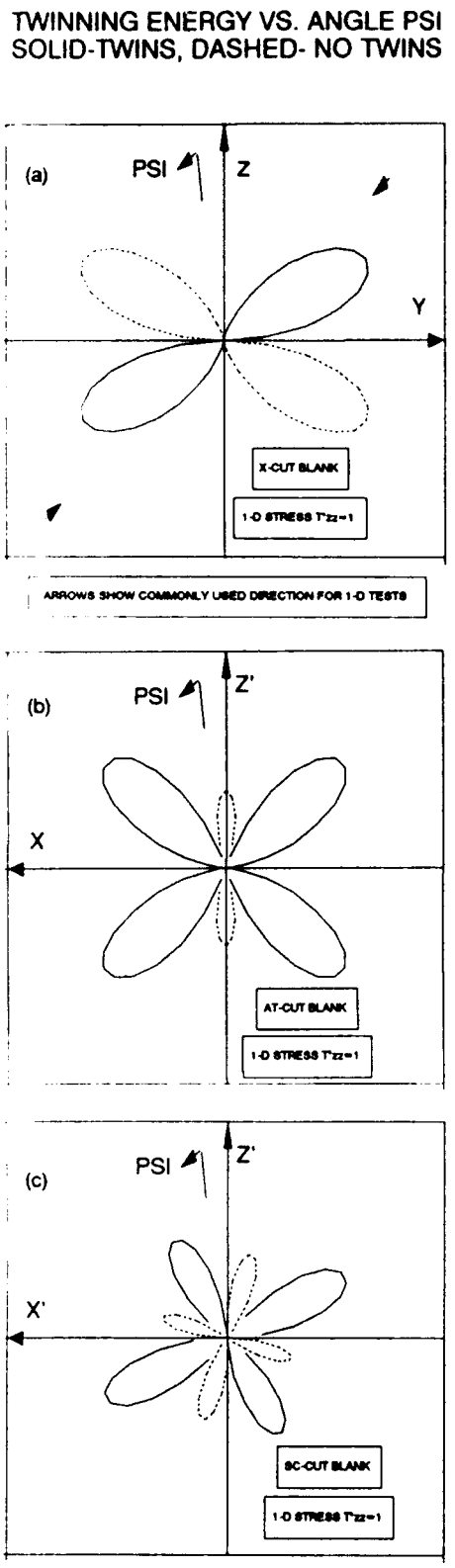


Figure 5. Calculated energy difference between the two twin states for uniaxial stress. (a) X-cut. (b) AT-cut. (c) SC-cut.

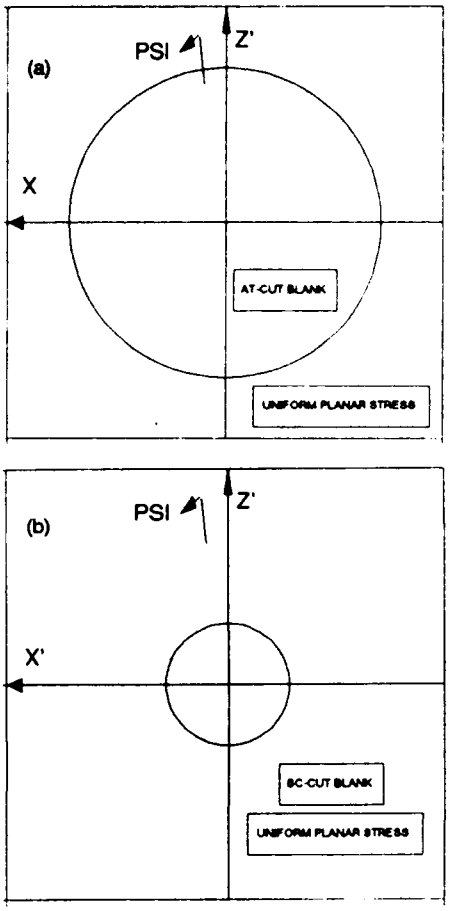


Figure 6. Calculated energy difference between the two twin states for uniform biaxial stress (note that the X-cut does not twin for this stress pattern). (a) AT-cut. (b) SC-cut.

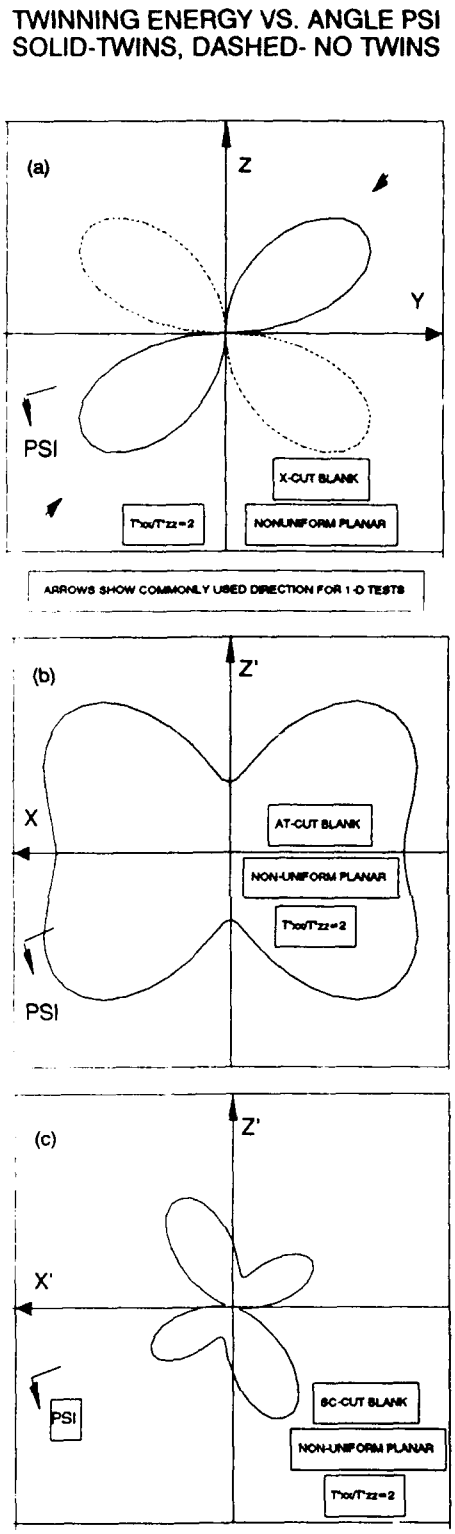


Figure 7. Calculated energy difference between the two twin states for nonuniform biaxial stress for the case of $T''_{xx}/T''_{zz} = 2$. (a) X-cut. (b) AT-cut. (c) SC-cut.

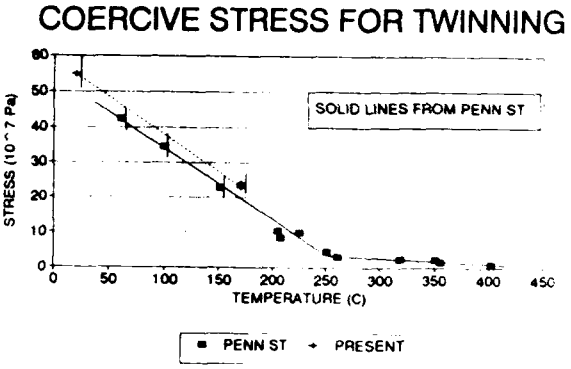


Figure 8. Measured threshold stress for twinning in AT-cut pressure transducers vs. temperature. Data from Penn. State [12] is included for comparison.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PIEZOELECTRIC VIBRATORY GYROSCOPE USING FLEXURAL VIBRATION OF A TRIANGULAR BAR

Satoru FUJISHIMA, Takeshi NAKAMURA and Katsumi FUJIMOTO

Murata Mfg. Co., Ltd.
Nagaokakyo, Kyoto, Japan

1. Introduction

Recently, car navigation systems have reached the stage of practical use and have become standard features of high-end cars. The directional sensors of these systems mainly use earth magnetism sensors, but they are prone to disturbances in the earth magnetism from the car body.

Piezoelectric vibratory gyroscopes were developed about 20 years ago, but they have not yet succeeded in practical use.

We have recently developed piezoelectric vibratory gyroscopes using flexural vibration of a triangular bar, which have higher sensitivity, lower drift and smaller configuration.

2. Basic principles of the piezoelectric vibratory gyroscope

Fig.1 shows the basic equivalent circuit of a piezoelectric vibratory gyroscope. In Fig.1 when an equivalent mass m vibrating in x axis direction rotated around z axis by rotation speed Ω, the motion equation regarding to the rotating axis is given by equation(1). [1]

$$\begin{cases} j\omega m \dot{x} = F_x + 2m\Omega\dot{y} + (m\Omega^2 x + m\dot{\Omega} y) \\ j\omega m \dot{y} = F_y - 2m\Omega\dot{x} + (m\Omega^2 y - m\dot{\Omega} x) \end{cases} \quad (1)$$

where the first term in right-hand side shows the outer force, the second term shows the Coriolis force (Fc), the third term shows centrifugal force (Fn) and the fourth term shows the force by the change of rotation speed . Generally, angular frequency(ω) of vibrating mass is so high that we may neglect the third and fourth term in equation (1).

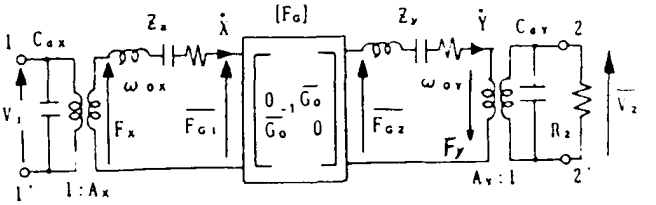


Fig.1 Basic equivalent circuit of a piezoelectric vibratory gyroscope.

In the case of vibratory gyroscope, $j\omega m$ in equation (1) is replaced by equivalent mechanical impedance z_x or z_y as follows:

$$\begin{cases} z_x \dot{x} = F_x + 2m_y\Omega\dot{Y} = F_x - G_o\dot{y} \\ z_y \dot{y} = F_y - 2m_y\Omega\dot{X} = F_y + G_o\dot{x} \end{cases} \quad (2)$$

Equation (2) is a motional equation in approximately, and the equivalent circuit is shown in Fig.1. [2]

When a vibratory gyroscope is drived and detected by piezoelectric transducers, Fig.2 shows the basic system diagram of Fig.1.

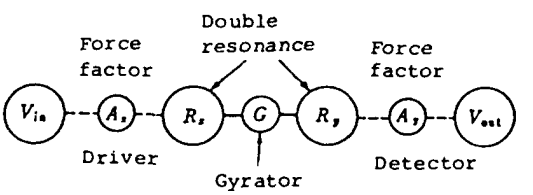


Fig.2 Basic system diagram of a piezoelectric vibratory gyroscope.

CH2965-2/91/0000-261 $1.00 © 1991 IEEE

Fig.3 shows the outline of a piezoelectric vibratory gyroscope using a square bar. In Fig.3, when a square bar vibrating in x axis direction by resonance frequency f_x is rotated around z axis by angular velocity Ω, the Coriolis force (Fc) is generated and the bar vibrates in y axis direction by the same frequency. So, we can detect the angular velocity by measurement of the square bar vibration in y axis. In this case, the Coriolis force is given as follows:

$$Fc = 2m\Omega\dot{y} \qquad (3)$$

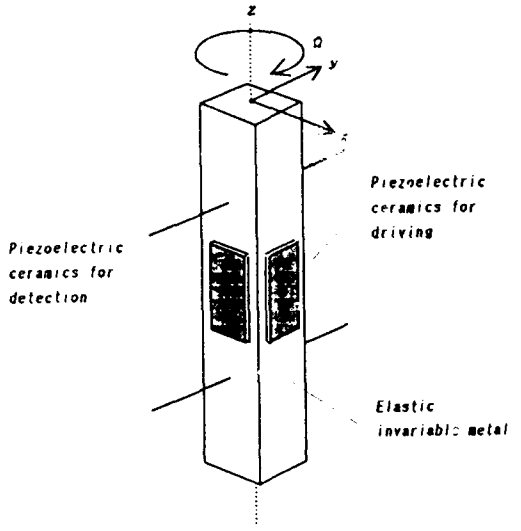


Fig.3 Outline of a piezoelectric vibratory gyroscope using a square bar.

3. Structure of a gyroscope using a triangular bar

In order to make higher sensitive gyroscope by simple structure, we have developed a new gyroscope using a triangular bar instead of a square bar. [3] Fig.4 shows the gyroscope using a triangular bar. The flexural resonance frequency of a triangular bar is given as follows:

$$f_\Delta = \frac{A\,a}{4\pi l^2}\sqrt{\frac{E}{6\rho}} \qquad (4)$$

where A, a, l, E and ρ are a contstant, width, length, Young's modulus and the density, respectively.

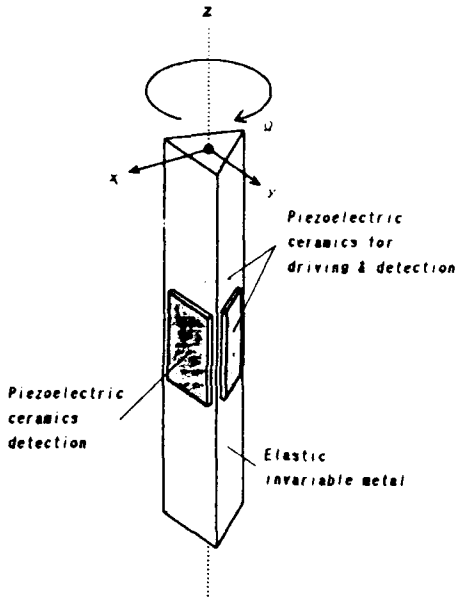


Fig.4 Outline of a piezoelectric vibratory gyroscope using a triangle bar.

According to the vibration analysis based on the finite-element-method (FEM) simulation, the node points are located at the 0.225ℓ as shown in Fig.5, which is same as a square bar. In this case, an euqivalent mass is given as follows:

$$m = M/4 = 0.11a^2l\rho \qquad (5)$$

where M is a total mass.

Using equations (3) (4) and (5), we can calculate the Coriolis force.

Elastic invariable metal (Elinvar) was used as the material for a triangular bar and three pieces of PZT-6 piezoelectric ceramics were glued in the center position. Two kinds of triangle bars were fabricated. One was mainly for automobile navigation at 7.85 KHz used 40 mm length and 3.5 mm width, and another was mainly for preventing camera shake at 25 KHz used 17 mm length and 2 mm width.

In order to make the zero temperature frequency characteristic, triangular bar metals were annealed at about 600°C before glueing. The temperature frequency characteristics were improved to less than ± 3 ppm/°C and mechanical Q values were more than 2,000.

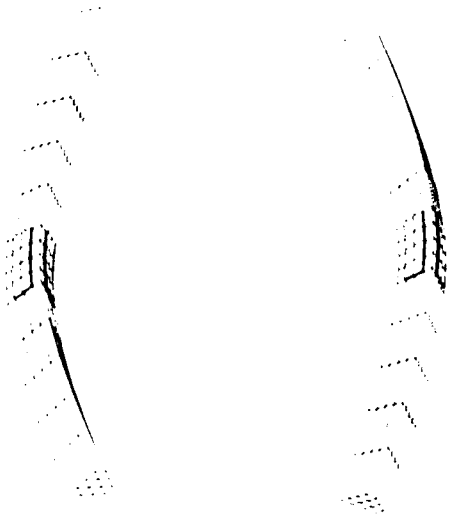


Fig.5 FEM figure of the flexural vibration of a triangular bar.

4. Adjustment of resonance frequency

Matching the driving frequency in x axis to the detector frequency in y axis enables high detection efficiency of output voltages to rotational angular velocity.

In the conventional piezoelectric vibratory gyroscope structure, the resonance frequencies of the driver and detector pieces have been tried to be same each other. However, the deviation of manufacturing accuracy causes difference of the resonance frequency between two axis. Therefore, it is extremely difficult to accurately adjust the oscillation frequency to detection frequency without disturbing the vibration mode.

In the frequency adjustment of a triangular bar, processing is performed on the ridgeline at the center of a bar. Trimming of the ridgeline rises up the resonance frequency only in one direction of the frexural mode. Thus, adjustment of one side does not affect the resonance frequency of the other two sides. Accordingly, the resonance frequency in y axis can be accurately adjusted to the resonance frequency in x axis.

5. Circuit configuration

The two pieces in three PZT ceramics were used as drivers, the remaining piece was used for feedback element to make an oscillator circuit. The two pieces were also used as detectors of the Coriolis force. The output was synchronously detected after differential amplification and a DC amplification circuit was used as shown in Fig.6. [4]

As the results of many experiments, our gyroscope has a high sensitivity to rotational angular velocity and a high S/N ratio. Furthermore, by using unique driving and detector circuit, the shock and vibration characteristics of the total configuration is very good more than 100 times comparing to the conventional vibratory gyroscope.

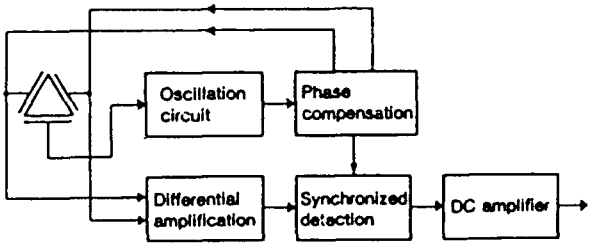


Fig.6 Circuit configuration of a piezoelectric vibratory gyroscope using a triangle bar.

6. Representative characteristics

The representative characteristics required for car navigation vibratory gyroscopes are sensitivity, linearity and durability.

Fig.7 shows an example of the sensitivity and linearity characteristics of our gyroscopes. Fig.8 shows an example of the drift characteristics comparing with conventional gyroscopes. Fig.9 shows an example of the vibration resistance characteristics comparing with conventional gyroscopes.

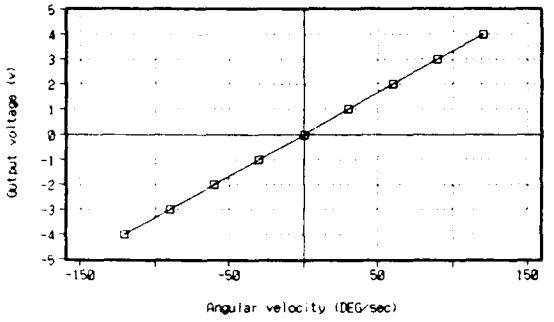


Fig.7 Sensitivity and linearity characteristics of a gyroscope using a triangle bar.

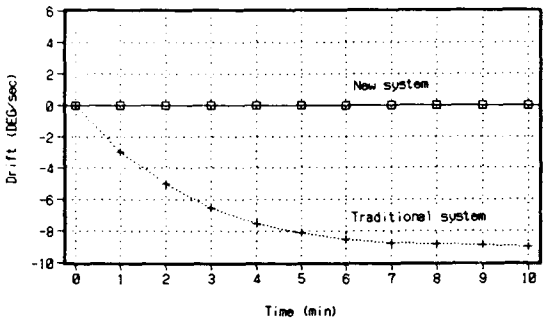


Fig.8 Drift characteristics of a gyroscope using a triangle bar comparing with an ordinary one.

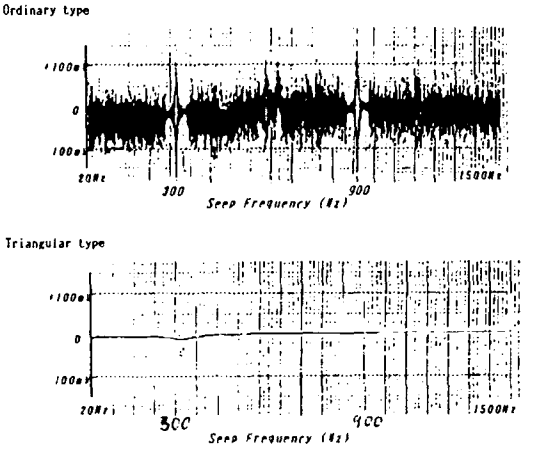


Fig.9 Vibration resistance characteristics comparing with an ordinary gyroscope.

7. Applications

Fig.10 shows a block diagram of the car navigation system used our vibratory gyroscope. In the navigation test by the block diagram, the angle was calculated by integrating the output of the gyroscope and the distance was calculated by integrating the speed of the car. Because of the nature of the system, even minute errors are integrated and the error accumulates as the car moves. Sign posts and correlations with maps are being considered as a means of calibration.

Fig.11 shows an example of the self-navigation test with our gyroscopes. The car was able to return to the starting point correctly after 4 mile self-navigation without any calibration.

We expect many applications of our gyroscopes in preventing camera shake, in position control of car and another robotts.

8. Conclusions

Considering basic principles of the piezoelectric vibratory gyroscope, we have succeeded in developing a unique vibratory gyroscope using flexural vibration of a triangular bar.

A simple configuration, high sensitivity and low drift are main merits of this gyroscope. This will be useful not only for navigation system but also for many other applications.

References

[1] M. Konno : Piezoelectric vibratory gyroscope, J.A.S.J. vol.45, No.5, pp.402-408, 1989 (in Japanese)

[2] S. Kudo, M. Konno, S. Sugawara, T. Yoshida : Consideration on equivalent mechanical circuits for vibratory gyroscope, Proc. 1990 Ultrasonics Symposium, pp.397-400

[3] T. Nakamura, K. Okano, T. Kasanami : Vibratory gyroscopic sensor using piezoelectrically drived metal bar, Proc. 150 Committy of Japan Society for The Promotion of Science, Jan. 1989 (in Japanese)

[4] T. Nakamura : Vibration gyroscope employs piezoelectric vibrator, JEE, pp.99-104, Sept. 1990.

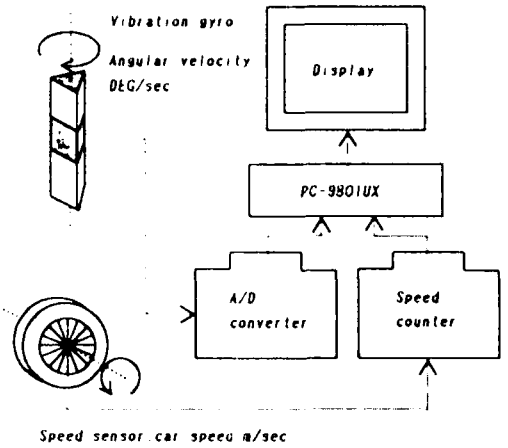


Fig.10 Block diagram of the navigation system using a piezoelectric vibratory gyroscope.

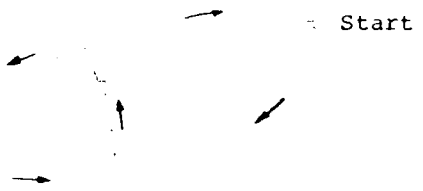
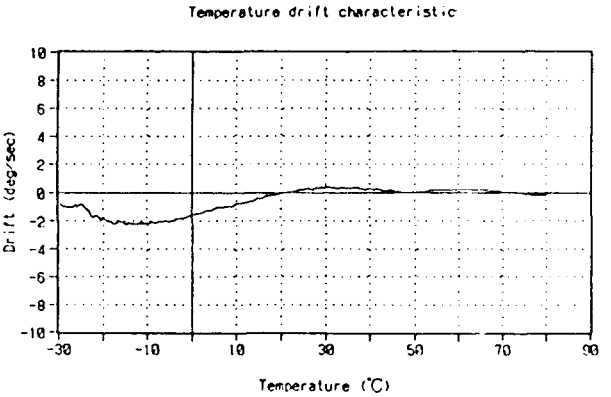


Fig.11 An example of the self-navigation test with a piezoelectric vibratory gyroscope.

Question:
What is the variation of null output over the temperature range for zero angular rate input to the gyroscope?

Answer:
Following figure shows an example of temperature drift characteristic of null output. We can guarantee less than ±3 deg/sec in temperature range from -10°C to 60°C for this gyroscope.



FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SOLDER BOND APPLICATIONS IN A PIEZOELECTRIC SENSOR ASSEMBLY

Paul T. Vianco
Jerome A. Rejent
Physical and Joining Metallurgy Division
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Abstract

A procedure was developed to use solder technology in the assembly of a single-crystal quartz accelerometer. 87.5Au-12.5Ge (wt.%) solder films 0.5×10^{-6}, 1.0×10^{-6}, and 2.0×10^{-6} m thick were formed by the electron beam deposition of individual layers of Au and Ge with thicknesses so that the bulk film composition equals the eutectic composition. Interdiffusion of the Au and Ge formed the solder; thermal-physical measurements showed the multilayer films to behave similarly to bulk 87.5Au-12.5Ge solder in process thermal cycles. The 2.0×10^{-6} m thick quartz/Au-Ge/quartz bonds had an adhesive tensile strength of 17 ± 2 MPa. The strength increased to 29 ± 3 MPa and 27 ± 12 MPa after thermal shock and thermal cycle exposures, respectively. The 1.0×10^{-5} m thick bonds exhibited strengths of 16 ± 3 MPa, 16 MPa, and 15 ± 8 MPa in the as-fabricated, post-thermal shock, and post-thermal cycled samples, respectively. The 0.5×10^{-6} m joints produced a large degree of scatter in the strength values. Accelerometers assembled with the 2.0×10^{-6} m thick joints demonstrated a significant improvement in temperature performance as opposed to units fabricated with a polyimide adhesive.

1. Introduction

A large family of sensors for detecting temperature, pressure, and acceleration is constructed of piezoelectric ceramics such as single crystal quartz, barium titanate, or lithium niobate. The fabrication of a growing number of these devices requires the physical joining of individual pieces of the ceramic. This paper describes the development of a thin solder bond technique that was used in the assembly of an accelerometer [1] made from single-crystal quartz.

A schematic diagram of the quartz digital accelerometer (QDA) is shown in Fig. 1. It consisted of two double-ended tuning forks (DETFs) arranged one on top of the other and separated by two spacers at either end. All parts were manufactured of single-crystal quartz with the z crystallographic direction out of the major plane. The x and y directions were across the width and along the length, respectively, of the tuning forks. The coefficient of thermal expansion and the elastic moduli were the same in the orthogonal x and y directions and are outlined in Table 1 [2]. These parameters are important in determining the residual stresses in the joint (which will be calculated later in this report). The dimensions of the bond areas measured 0.0010×0.0014 m. The substrate surfaces were coated with 45 nm of Cr on the quartz followed with 180 nm of Au. The Cr-Au thin films were retained from the photolithographic processing used to fabricate the quartz parts.

The performance of the accelerometer required that the joints be of a uniform thickness over the bond area. Non-uniformity was observed in initial assemblies, which used organic adhesives (Fig. 2), causing temperature-dependent frequency shifts that detracted from the sensor's performance. The effects of bond joint misalignment on the resonator properties were confirmed by finite element analysis [3]. It was also necessary that the joints be as thin as possible (preferably $<5\times10^{-6}$ m) and that the adhesive be precisely located on the bonding pads. Moreover, reproducibility of the accelerometer's performance required that the same bond thickness, uniformity, and lateral dimensions be met on each unit. Besides an adequate adhesive strength, other properties required of the bonding agent were low vapor pressure constituents to minimize changes to the tuning fork frequencies and thermal-physical

CH2965-2/91/0000-266 $1.00 © 1991 IEEE

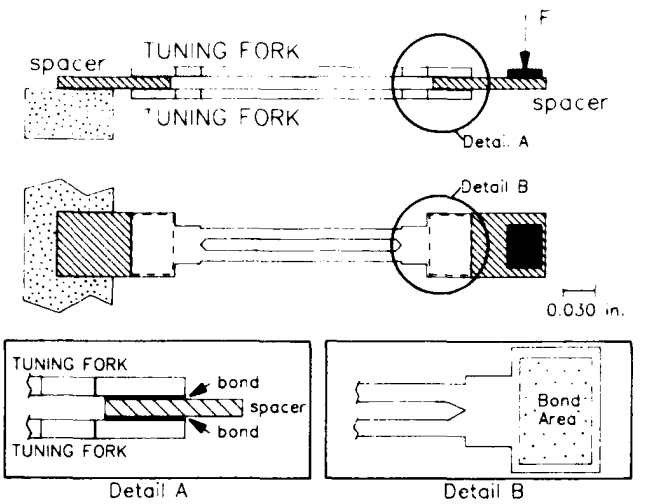


Figure 1. A schematic diagram of the quartz digital accelerometer assembly. The bonding surface locations have been highlighted.

Table 1. Thermal-Physical Properties of the x and y Directions of Single-Crystal Quartz.

Material	Quartz (x,y)
Coefficient of Thermal Expansion ($^{\circ}C^{-1}$)	$(-1.405 + 0.1147\,T - 3.526\times10^{-4}\,T^2 + 5.682\times10^{-7}\,T^3 - 3.185\times10^{-10}\,T^4)\times10^{-6}$ (T, °K)
Elastic Modulus (MPa)	$86814\,(1 - 4.359\times10^{-5}\,\Delta T - 1.110\times10^{-7}\,\Delta T^2)$ (T, °C; $\Delta T = T - 20^{\circ}C$)
Poisson's Ratio	0.25

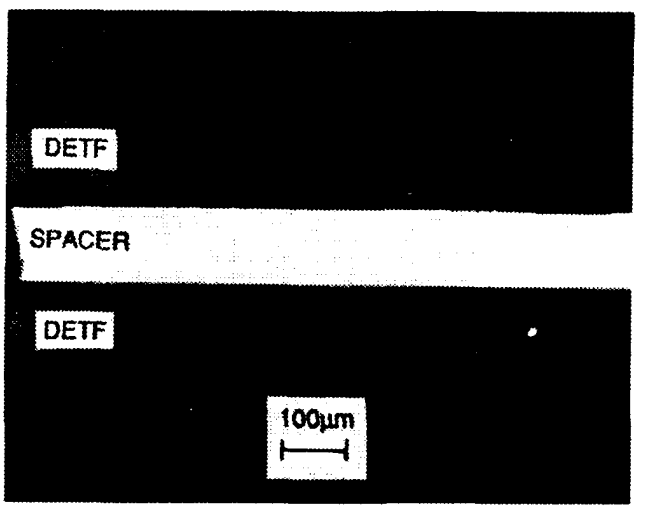


Figure 2. Optical micrograph of the bond joint of a QDA assembly in cross section showing non-uniform joint thickness, which plagued the use of the polyimide adhesive.

properties compatible with subsequent processing steps and the use environment.

The bonding material selected for this application was the eutectic gold-germanium solder, 87.5Au-12.5Ge (wt.%). The melting point is 361°C [4], well above the maximum secondary processing temperature of 310°C. The gold-rich and germanium-rich phases that constitute the eutectic solder readily wet the Cr-Au coated quartz surface without the requirement of a flux. The gold and germanium both have low vapor pressures. The properties of the 87.5Au-12.5Ge solder are listed in Table 2 [5,6]. The coefficient of thermal expansion is high relative to most metal systems, thus allowing the 87.5Au-12.5Ge alloy to more closely match the expansion coefficient of the quartz in the x and y directions. The thermal expansions of the quartz and solder deviated at elevated temperatures such that upon cooling from the reflow temperature, tensile residual stresses were introduced into the quartz and compressive stresses into the solder. The brittle nature of the quartz necessitated an evaluation of the magnitude of those stress values.

Table 2. Thermal-Physical Properties of 87.5Au-12.5Ge Solder.

Material	87.5Au-12.5Ge (bulk)
Coefficient of Thermal Expansion ($°C^{-1}$)	12.03×10^{-6}
Elastic Modulus (MPa)	$1000\,(138.8 - 0.2298\,T + 5.267 \times 10^{-4}\,T^2 - 4.629 \times 10^{-7}\,T^3)$ (T, °K)
Poisson's Ratio	$0.3843 + 3.084 \times 10^{-5}\,T$ (T, °C)

A first order approximation to the residual stresses in the solder and quartz substrates caused by thermal expansion mismatch between the two materials was derived from the one-dimensional, bi-metal beam problem developed by Timoshenko [7]. Creep of the solder is not figured in this model. However, the one-dimensional calculation was not adequate for the present situation because the bond area geometry required that both planar dimensions be considered. Therefore, a two-dimensional approximation will be presented. Shown in Fig. 3 is the geometry of the bond architecture used in the residual stress analysis.

The algebraic expressions required for the calculation of the forces in each of the layers were based upon (1) force balance, (2) moment balance, and (3) compatibility at the layer interfaces. The system of equations was developed in each of the x and y directions. The Poisson effect caused the two sets of equations to couple. The system of equations remained determinant with 12 equations for the following 12 unknowns: R_x, R_y, $P_{x,1}$, $P_{x,2}$, $P_{x,3}$, $P_{x,4}$, $P_{x,5}$, $P_{y,1}$, $P_{y,2}$, $P_{y,3}$, $P_{y,4}$, and $P_{y,5}$. The parameters R_x and R_y were the radii of curvature about the axis parallel to the x and y directions, respectively, and a_i is the layer thicknesses. $P_{x,i}$ and $P_{y,i}$ were the forces in the i^{th} layer ($i = 1,2,\ldots,5$) parallel to the x and y directions, respectively. The equations in the x direction only will be outlined below; the corresponding y expressions are the same except for an interchange of the x and y notations. The force balance is given by:

$$\sum_{i=1}^{5} P_{x,i} = 0 \qquad (1)$$

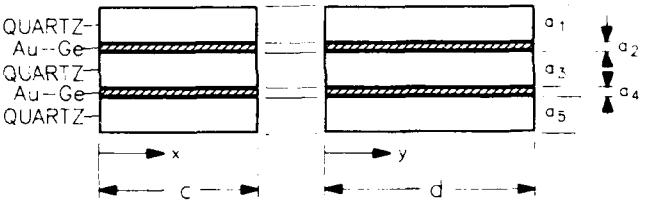


Figure 3. Schematic diagram of the geometry of the QDA bond joints used to calculate the residual stresses in the quartz substrates and solder films.

The moment balance is given by:

$$P_{x,1}\frac{a_1}{2} + P_{x,2}\left(a_1 + \frac{a_2}{2}\right) + P_{x,3}\left(a_1 + a_2 + \frac{a_3}{2}\right) + P_{x,4}\left(a_1 + a_2 + a_3 + \frac{a_4}{2}\right) + P_{x,5}\left(a_1 + a_2 + a_3 + a_4 + \frac{a_5}{2}\right) + \frac{c}{12\,R_y}\left\{E_{x,1}\,a_1^3 + E_{x,2}\,a_2^3 + E_{x,3}\,a_3^3 + E_{x,4}\,a_4^3 + E_{x,5}\,a_5^3\right\} = 0 \qquad (2)$$

$E_{x,i}$ is the elastic modulus in the ith layer. Finally, the compatibility requirement (which stipulates that there is no separation permitted at the layer interfaces) is given by the following expression:

$$\frac{1}{E_{x,i}\,a_i}\left\{\frac{P_{x,i}}{c}-\frac{\nu_i P_{y,i}}{d}\right\}-\frac{1}{E_{x,j}\,a_j}\left\{\frac{P_{x,j}}{c}-\frac{\nu_j P_{y,j}}{d}\right\} \quad (3)$$

$$+\frac{1}{R_y}\left(\frac{a_i+a_j}{2}\right)=(\alpha_{x,j}-\alpha_{x,i})\,\Delta T$$

where $j=i+1$. ν is the Poisson's ratio and α is the thermal expansion coefficient. The system of equations was solved by linear algebra techniques.

The loads in each layer were the sum of the values calculated in each of six temperature steps between 300°C and 20°C. The values of $1/R_x$ and $1/R_y$ were approximately zero as expected from the symmetry of the joints (Fig. 3). Therefore, the axial loads were the maximum values so that when divided by the cross sectional area normal to the load, they defined the stress level acting on the layer. The residual stresses for three layer thicknesses are summarized in Table 3. Symmetry also caused the stress levels to be same in similar-type layers. For the 2×10^{-6} m thick joint, the stresses in the quartz layers were tensile at 1.4 MPa and 2.9 MPa in the x and y directions, respectively. Correspondingly, compressive stresses in the solder layers were −160 MPa and −330 MPa for the x and y axes directions, respectively. The tensile stresses in the quartz were one to two orders of magnitude below the fracture strength of the quartz, thereby reducing the likelihood of failure to the quartz substrates. The stress in the quartz layers decreased in proportion to the solder layer thickness, thereby allowing the results in Table 3 to be extrapolated to other solder layer thicknesses. The compressive stresses in the solder film were relatively large because of the limited solder film thickness. The stress levels in the solder films were constant for the various thicknesses because the cross sectional area dropped proportionally with the load. The effect of the compressive stresses on film adhesion will need to be assessed experimentally, particularly because the constrained geometry of the solder joint may alter the failure properties. In summary, this calculation indicated that the thermal mismatch residual stress levels generated in the quartz were not sufficient to cause fracture at the proposed range of joint thicknesses. However, the effect of the corresponding compressive loads in the solder film awaited testing of the joints.

Thin film deposition techniques were developed to deliver a precise pattern of solder to the substrate. An initial attempt was made to sputter-deposit the films from an 87.5 Au-12.5 Ge source. Because sputtering is a mechanically based transfer of atoms from the target to the substrate, the procedure offered a greater likelihood of reproducing the solder composition on the substrate. The sputter-deposited films showed slight composition variations from the eutectic value; however, the discrepancies did not significantly affect the reflow parameters of the solder as determined by differential scanning calorimetry (DSC). However, bonding between the solder film and the metallized quartz substrates was poor because of severe oxidation of the Ge-rich phase. Moreover, oxidation of the films was enhanced by extensive porosity generated by the sputter-deposition process.

Table 3. Residual Stress Values in the Solder and Quartz Layers of the Bond Joint, Fig. 3.

Solder Layer Thickness	Residual Stress (MPa) Quartz		Solder	
10^{-6} m	x*	y#	x*	y#
0.5	0.37	0.75	−160	−340
1.0	0.71	1.5	−160	−340
2.0	1.4	2.9	−160	−330

* loads parallel to the 1.0 mm dimension
\# loads parallel to the 1.4 mm dimension

The sputter-deposited films clearly showed that exposure of the Ge-rich phase to air had to be limited for successful bonding. Therefore, an alternate procedure was developed whereby Au and Ge were deposited as separate layers in the sequence of Au-Ge-Au. The stacking order protected the Ge layer from excessive oxidation. The thicknesses of the layers were selected such that together the composition of the total film would equal that of the eutectic material. For a 2.0×10^{-6} m thick solder layer, the Au and Ge thicknesses were 658 and 684 nm, respectively. The anticipated process within the films was that upon heating the component layers would interdiffuse and form the eutectic solder. The three principal concerns of this technique were that (1) the required Au and Ge layer thicknesses for a proposed 2.0×10^{-6} m joint were large enough to generate *internal* stresses sufficient to cause decohesion during deposition or prior to reflow; (2) the film thicknesses had to allow the intermixing of the Au and Ge layers at a temperature near to or below the expected reflow temperature of 361°C characteristic of the cast alloy; and (3) it was necessary to determine whether there were physical artifacts in the films during intermixing that would disrupt the joint.

These three concerns were answered in a series of short experiments. The film integrity was verified by

depositing the multilayer 658 nm Au - 684 nm Ge - 658 nm Au coating on a polished fused silica plate that had already been coated with 45 nm Cr followed by 180 nm Au. All film thicknesses were verified to be ±1-2% of the nominal values. The films remained intact after deposition. The substrate was fractured to reveal the three layers (and the Cr-Au metallization) by scanning electron microscopy (SEM) as shown in Fig. 4. Even the fracture did not cause the film layers to delaminate. In fact, the bi-layer system of 684 nm Ge followed with 1316 nm Au also remained adherent to the substrate with no indication of separation between the Au and Ge components. These results confirmed that the multilayer films could be deposited at the required thicknesses without decohesion.

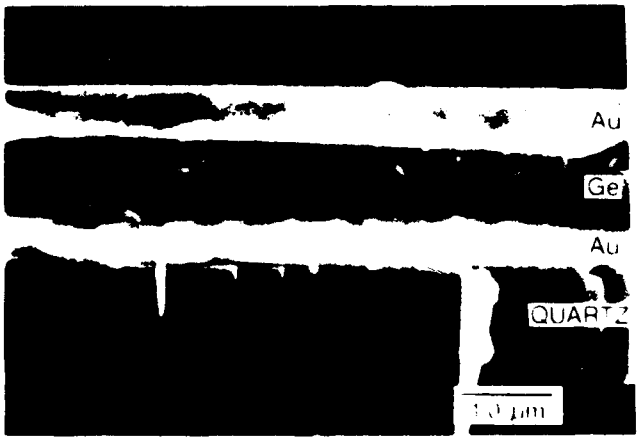


Figure 4. Scanning electron micrograph of the cross section of the Au-Ge-Au multilayer film (total thickness, 2.0 microns) on Cr-Au metallized fused silica.

The thermal-physical properties of the films were examined by DSC analysis. For this study, optically polished NaCl substrates were first coated with 45 nm Cr and 180 nm Au. Then, the multilayer 658 nm Au - 684 nm Ge - 658 nm Au film was deposited on the specimen. Two other film architectures were also examined: 684 nm Ge - 1316 nm Au and 342 nm Ge - 329 nm Au - 342 nm Ge - 329 nm Au. The coated salt substrates were then placed in a bath of 1:1 deionized water and ethyl alcohol, which slowly dissolved the NaCl substrate, releasing the films. The DSC measurements were made at a scanning rate of 10°C/min between 25°C and 420°C. Two temperature cycles were made with each sample. Shown in Fig. 5a is the initial DSC plot of the film system, 684 nm Ge - 1316 nm Au. The curve exhibited some small amplitude exotherms prior to the melting peak that were not present in the curve derived from a section of cast 87.5Au-12.5Ge solder ribbon, a portion of which is shown in Fig. 5b. The onset temperature (which designated the "melting" point) of the endothermic peak from the multilayer film, 362 ± 1°C, was not significantly different from that of the cast ribbon, 360 ± 1°C. A second thermal-physical parameter of the melting process was the heat of transformation (fusion), ΔH_f. The value of ΔH_f for the multilayer film was 75 ± 2 J/g as compared to 77 ± 2 J/g for the cast 87.5Au-12.5Ge. Similar results and DSC curves were obtained from the other two film structures. Therefore, these data clearly showed that the thermal-physical properties of the multilayer films were nearly identical to those of the cast alloy.

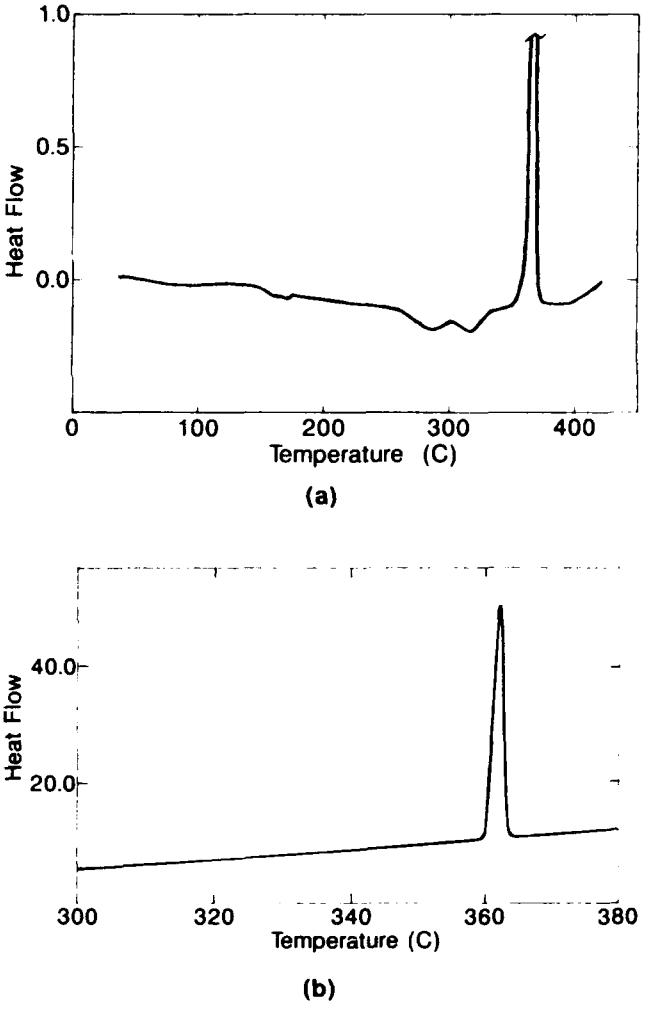
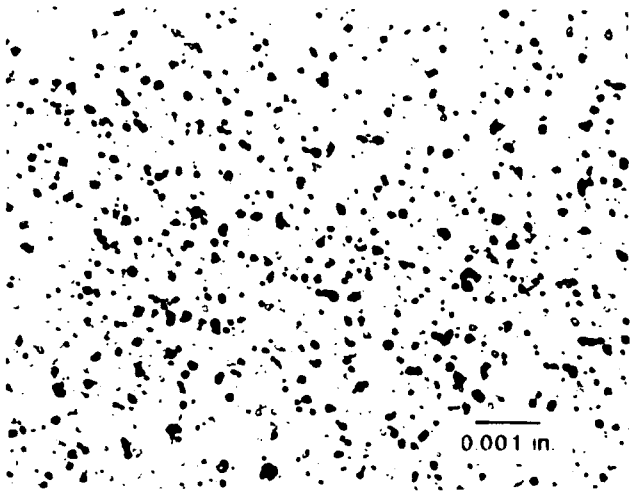
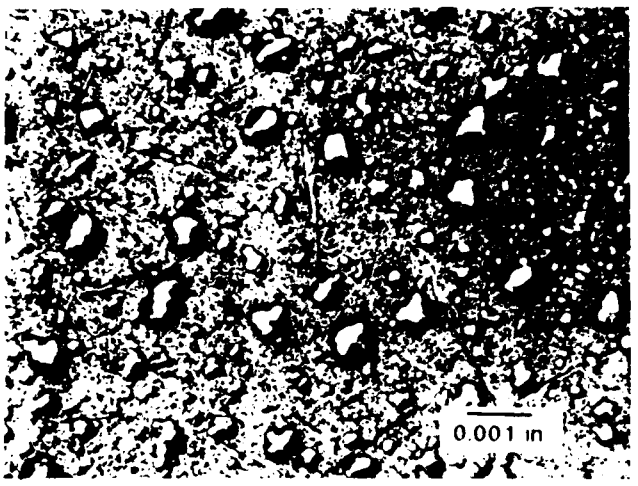


Figure 5. (a) DSC curve of the 684 A Ge - 1316 A Au multilayer film. (b) DSC curve of a sample of cast 87.5Au - 12.5Ge ribbon. In both plots, the scanning rate was 10°C/min.

Microscopic examination of the reflowed Au-Ge-Au film exhibited the two-phase structure of the eutectic system. The extent of interdiffusion/phase separation of the Au and Ge constituents is demonstrated in Fig. 6. Fig. 6a indicates the mixed morphology of the Au- and Ge-rich regions, which upon further heating in vacuum caused the solder to collect into small balls (Fig. 6b). Interestingly enough, when reflow was performed in air, the solder did not collect under its own surface tension but rather formed a complete coating over the whole substrate surface.



(a)



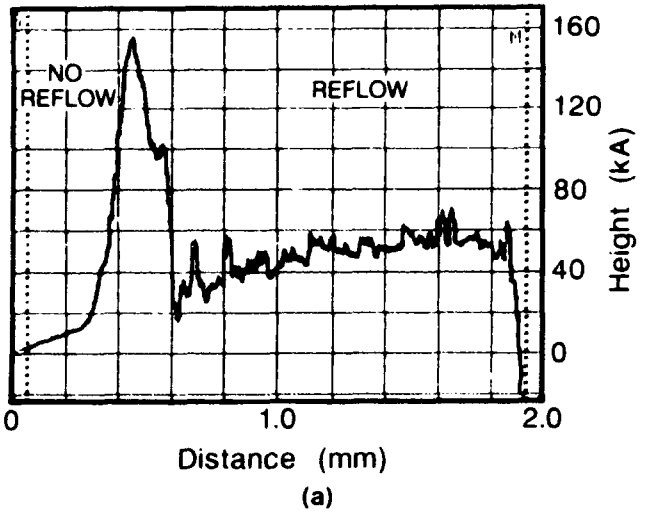
(b)

Figure 6. (a) Optical micrograph of the morphology of the multilayer film after heating in vacuum at 370°C for 5 min. (b) Heating the film in vacuum at 390°C for 15 min caused the solder to ball up under the effect of its surface tension.

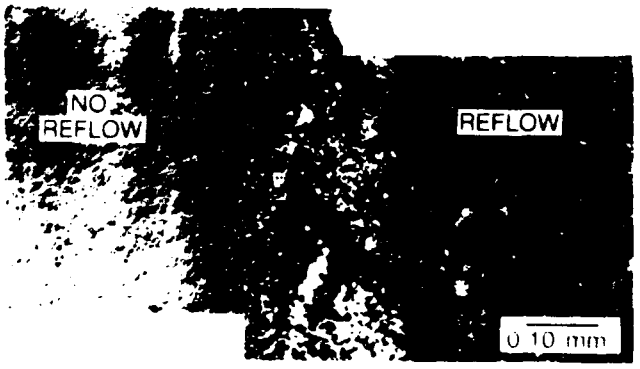
The third concern of the multilayer films was the appearance of physical artifacts that were part of the intermixing process and that would disrupt the bond formation. It was necessary to freeze the interdiffusion process so that observations and dimensional measurements could be made of any *intermediate* structures in the film. A procedure was developed whereby $0.0254 \times 0.0254 \times 0.00016$ m thick optically polished, fused silica substrates were deposited with the 45 nm Cr – 180 nm Au metallization followed by one of the three Au-Ge multilayer film sequences noted earlier to a total solder thickness of 2.0×10^{-6} m. Each wafer was placed edge-on onto a hot plate (in air) that was at 420°C, creating a thermal gradient within the substrate. Intermixing/reflow of the films was visually detected by a change to the surface finish, color, and luster. The appearance change progressed as a front up the substrate along the thermal gradient. The movement of the front was stopped by removal of the substrate from the hot plate. The film surface was examined by surface profilometry. A sample profile and a micrograph of the corresponding area taken from the Au-Ge-Au film are shown in Fig. 7. The boundary between reflowed and non-reflowed material had an elevation of 1.4×10^{-6} m above either area; the reflowed region had the same relative height as the non-reflowed area. Similar features were observed with the other Au-Ge film stacking patterns. This analysis confirmed that the starting thickness of the multilayer film was reproduced in the reflowed solder film.

In summary, these experiments verified that the multilayer Au-Ge films had the same thermal-physical characteristics as the cast 87.5Au-12.5Ge solder ribbon. Also, the films could be deposited to the observed thicknesses without a loss of adhesion to the metallized substrate. The films remained adherent after intermixing and reflow of the Au and Ge components and did not exhibit properties that would hinder joint formation.

Use of the thin film 87.5Au-12.5Ge solder as an adhesive required that a suitable bond strength be achieved between the (metallized) quartz substrates. Moreover the bond strength had to be maintained through thermal cycling conditions as well as after exposure to thermal environments required by subsequent processing of the sensor. An evaluation of the bond strength required the development of test samples (and procedures) that closely simulated solder joints in the actual device. That test program will be detailed in the following section.



(a)



(b)

Figure 7. (a) Surface profilometer trace normal to the front formed between the reflowed and non-reflowed regions of the Au-Ge-Au film (2.0 microns thick) on fused silica. (b) Optical micrograph of the boundary feature detected by the profilometer trace.

2. Experimental

The strength of the Au-Ge bond was evaluated by testing the joint in tension. It was critical that in order to model the actual joint and therefore predict its behavior in the assembled accelerometer, the bonded substrates had to have the same quartz orientation, similar dimensions, and identical surface finish including the Cr-Au metallization layer.

The substrates were fabricated from $0.0269 \times 0.0236 \times 0.000152 \pm 0.000010$ m single crystal quartz with the z direction perpendicular to the major faces. The surfaces were optically polished to a finish of 73 nm RMS. One substrate was coated on one side with the 45 nm Cr - 180 nm Au metallization followed by the three layer (Au-Ge-Au) solder film. The three layer architecture was used exclusively in this study. Total solder thicknesses of 0.5×10^{-6}, 1.0×10^{-6}, and 2.0×10^{-6} m were evaluated. After deposition, the wafer was cut up into sections approximately 0.00635×0.00635 m.

The second substrate was optically polished, z-oriented single-crystal quartz with the 45 nm Cr - 180 nm Au metallization deposited on one side of the wafer. Photolithography and quartz milling techniques were used to define a 0.00317×0.00317 m area of Cr-Au metallization centered on a 0.00635×0.00635 m section of quartz cut from the larger wafer. The bond area was defined by the 0.00317 m square area of the second substrate; the Au-Ge solder did not adhere to the bare quartz surface. A schematic diagram of the bonding sample substrates is shown in Fig. 8.

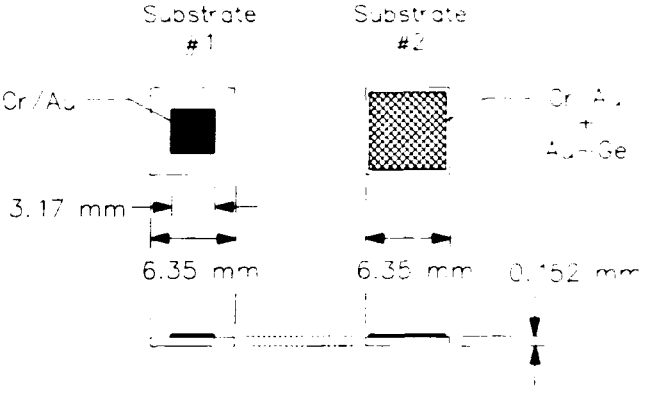
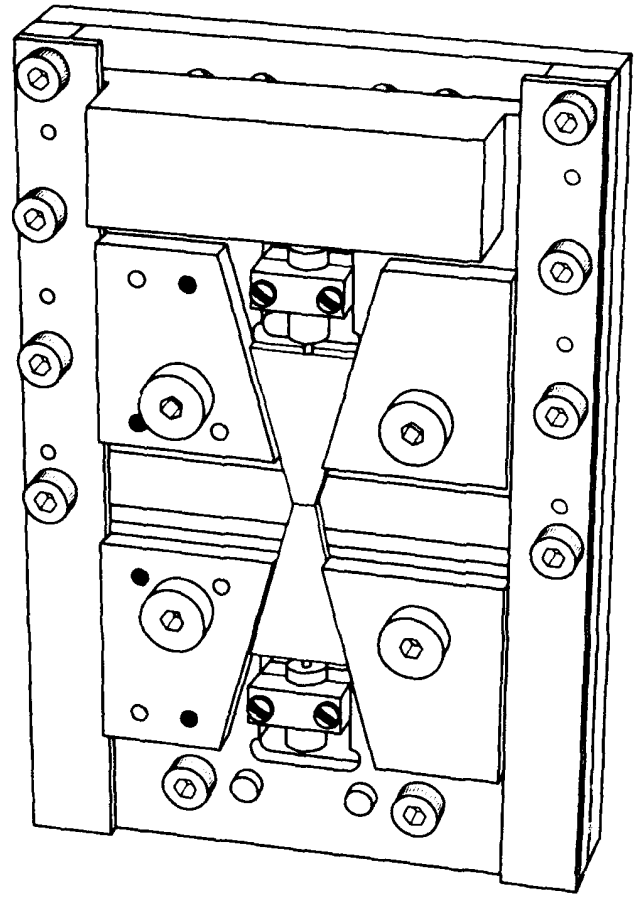
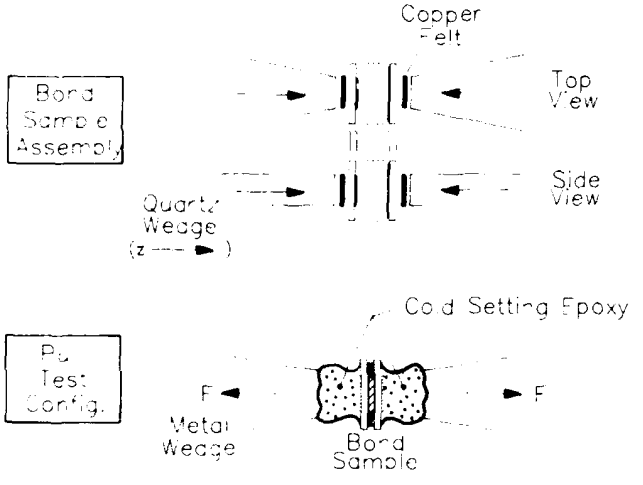


Figure 8. Schematic diagram of the bonding test substrates.

The bond sample was assembled in the fixture shown in Fig. 9a; the assembly process is diagrammed in Fig. 9b. The two substrate pieces were placed between two wedges that fit into the grips of the fixture. The quartz wedges were constructed of single-crystal quartz with the z direction parallel to that of the substrates in order to match their thermal expansion coefficients during reflow. A $0.00317 \times 0.00317 \times 0.000813$ m section of 20% filled copper felt was located between each quartz wedge and the substrate to assure a uniform distribution of pressure over the bond area.



(a)



(b)

Figure 9. (a) Schematic diagram of the fixture used to assemble the bonding strength test specimens. (b) Process sequence used to assemble the test samples.

Reflow was performed in vacuum (typically $<1.3\times10^{-3}$ Pa). Furnace control and temperature monitoring were performed through a Type K thermocouple embedded in a hole in one of the wedges. The peak temperature and time parameters for reflow were varied from 370°C to 390°C and 5 min to 30 min, respectively, with the goal of determining the lowest value of each that would produce satisfactory bonds.

Some samples were mechanically tested in the as-fabricated conditions. Other specimens were subjected to either (1) thermal shock exposure; (2) thermal cycling exposure; or (3) a static heat treatment developed to replicate subsequent processing steps (resulting from packaging procedures). The thermal shock exposure had the following parameters: (a) limits, −55°C to 125°C; (b) 200 cycles; (c) instantaneous temperature ramps; and (d) 10 min hold times at the temperature limits. The thermal cycle tests were performed as follows: (a) limits, −55°C to 125°C; (b) 100 cycles; (c) temperature ramps of 6°C/min; and (d) 120 min hold times at the temperature limits. Finally, static annealing treatments were conducted in air and vacuum with the following schedule: (a) 25°C to 300°C at 5°C/min; (b) hold at 300°C for 180 min; and (c) furnace cool.

After the quartz/Au-Ge/quartz pull test specimens had been fabricated, each was bonded between two metal wedges (similar to the quartz wedges used earlier) by means of a cold-setting epoxy in a fixture like that in Fig. 9a. The fixture and the sample were then transferred to an Instron™ 1130 load frame for the pull test. The crosshead speed was 1.7×10^{-4} m/s. The maximum load of the test divided by the bond area, 1.01×10^{-5} m^2, was used to calculate the bond strength (stress) values.

3. Results and Discussion

A. Adhesion tests – as-fabricated solder joints.

Observations of the bonding samples after assembly showed no indication of cracks or failures as would be expected from excessive residual stress generation. Shown in Table 4 are the pull strength data of the 1.0×10^{-6} and 2.0×10^{-6} m thick bonds formed by reflow at 390°C and 15 min. The load-displacement curves showed no evidence of ductility. The average strength (and ± one sigma) of the 2.0×10^{-6} m joints was 18±3 MPa. This value is a lower limit of the strength because test #4 exhibited premature failure in the quartz and epoxy; the actual Au-Ge bond remained intact. In all cases but test #4, the failure took place in the joint; specifically, fracture occurred between the metallization and the quartz substrate surface. Optical micrographs of the two complementary fracture surfaces are shown in Figs. 10a and 10b.

Higher magnification micrographs and a schematic interpretation of the failure path are shown in Figs. 10c, 10d, and 10e. It was apparent that the solder did not completely fill the joint but rather formed closely packed cells. The two zones in Fig. 10a and Fig. 10b were distinguished by the spacing of the cell-like morphology. It is likely that this phenomenon was a feature intrinsic to the reflow properties of the film under the confined geometry of the joint. Repeatability of the morphology between samples eliminated the source as variations to the flatness of the quartz surface that would cause incomplete filling of the joint.

Table 4. Adhesion Strength of the 1.0 and 2.0 Micron Thick Au-Ge Bonds.

Bond Thickness (m)	Test #	Strength* (MPa)	Failure Mode
2.0×10^{-6}	1	16	Bond
	2	19	Bond
	3	15	Bond
	4	21	Qtz. + Epoxy
1.0×10^{-6}	1	13	Bond
	2	18	Bond
	3	16	Bond

* Absolute (measurement) error = ±0.3 MPa.

Three samples with 2.0×10^{-6} m thick bonds were fabricated by dropping the reflow time period from 15 min to 5 min. The recorded adhesion strengths were 0.84, 2.6, and 13 MPa. The fracture surface of the 0.8 and 2.6 MPa tests samples clearly showed incomplete reflow of the solder. Increasing the time period from 15 min to 30 min had no effect on the bond performance.

Two samples were fabricated with the reflow cycle of 370°C, 5 min. The strength values were 0.7 and 1.2 MPa. The fracture surfaces clearly showed incomplete reflow; however, the Au and Ge were well intermixed. It was suspected that the solder film condition was similar to that of the boundary region in Fig. 7b. In conclusion, it was determined that the optimum reflow parameters for the test program were 390°C, 15 min.

In Table 4, data is presented from the 1.0×10^{-6} m thick joints in the as-soldered condition. The strength was 16 ± 3 MPa. In all cases, the failure was within the joint at the metallization-quartz interface of that substrate which was previously coated with the Au-Ge multilayer coating. Low-magnification optical micrographs (Fig. 11) indicated that the solder and metallization within the bond area had completely separated from the quartz surface. The different failure modes between the 1.0×10^{-6} m and 2.0×10^{-6} m thick joints clearly showed that the structure of the solder film was dependent on the joint thickness; yet, the strength values were not significantly different.

The adhesion strength data of the 0.5×10^{-6} m thick joints appears in Table 5. The first four tests (#1 to #4) showed a large degree of variability in the data (13 ± 10 MPa). The tests were repeated with a second group of four samples (#5 through #8) to confirm the strength variability. Similar scatter was indeed observed (13 ± 8 MPa). In all eight tests, failure occurred within the bond region. In five instances, the fracture surface indicated a lack of intimate contact between the two substrates, which was likely caused by obstructions such as dust particles present during assembly. In three cases, failure was observed at the metallization-quartz interface. Two of the three failures had morphologies similar to those of the 1.0×10^{-6} m thick joints, yet had strength values which were very low. The poor adhesion between the film and the quartz may reflect a general weakening of the joint (interfaces) caused by the thinner, more confined geometry. This thinner geometry limits the time-dependent deformation during cooling and the relaxation of a portion of the residual stresses (Table 3).

The scatter of the strength values caused the 0.5×10^{-6} m joint to be removed from further evaluation for sensor assembly.

In conclusion, it was determined that the 1.0×10^{-6} and 2.0×10^{-6} m thick multilayer Au-Ge solder films would produce bonds that were consistently strong enough (>4 MPa) to satisfy the sensor specifications.

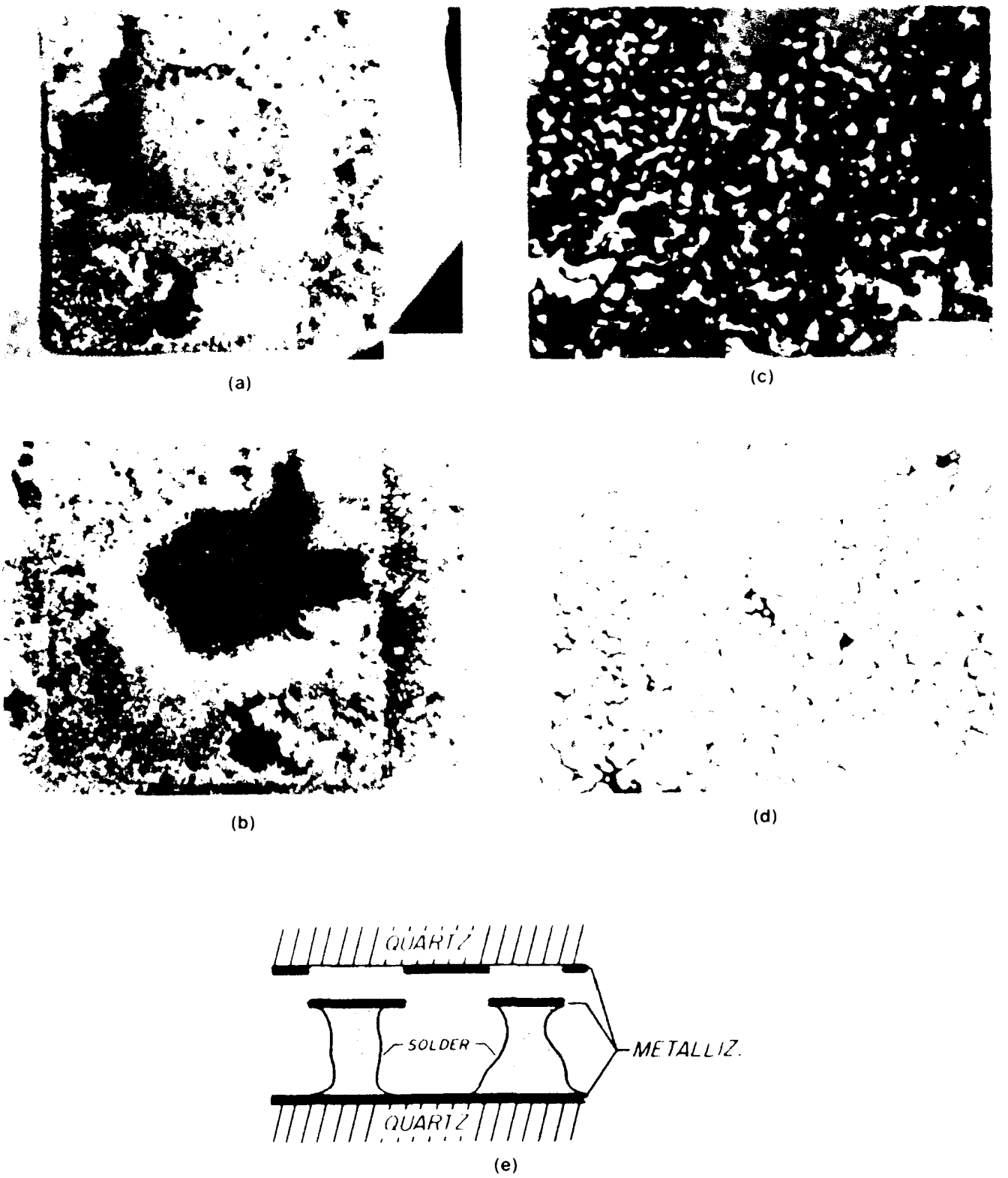
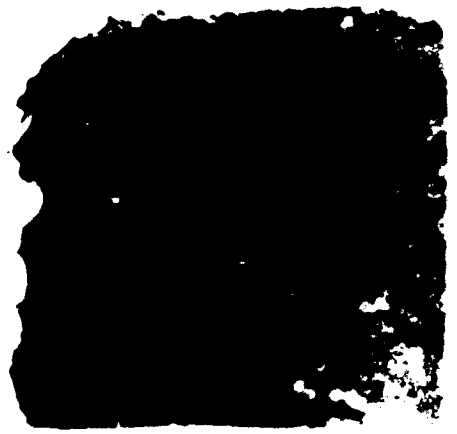
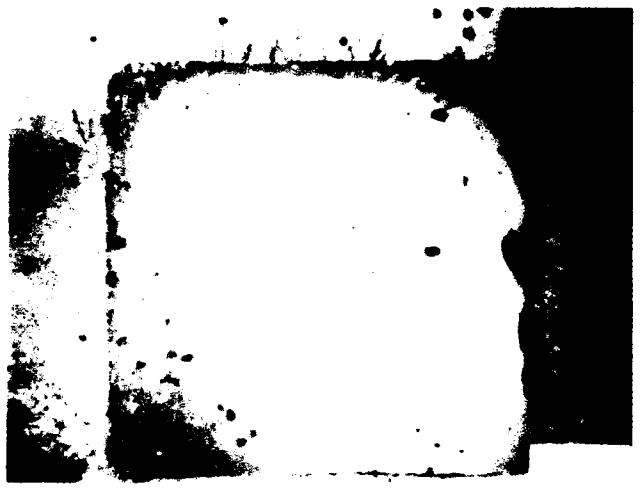


Figure 10. (a), (b) Low-magnification optical micrographs of the fracture surfaces of the patterned substrate and the solder plus metallized substrate, respectively, from a 2.0 micron thick Au-Ge solder joint after the pull test. (c), (d) High magnification optical micrographs of the same respective fracture surfaces. (e) A schematic diagram of the failure path.



(a)



(b)

Figure 11. Low-magnification optical micrographs of the fracture surfaces of (a) the patterned substrate and (b) the solder plus metallized substrate, respectively, from a 1.0 micron thick solder joint after the pull test.

Table 5. Adhesion Strength of the 0.5 Micron Thick Au-Ge Bonds.

Bond Thickness (m)	Test #	Strength (MPa)	Failure Mode
0.5×10^{-6}	1	23	Bond
	2	19	Bond
	3	6	Bond
	4	3	Bond
	5	17	Bond
	6	4	Bond
	7	8	Bond
	8	21	Bond

B. Adhesion tests – thermally shocked and thermally cycled solder joints.

Listed in Table 6 are the adhesion strength values of the 2.0×10^{-6} m thick solder joints after exposure to either 200 thermal shock cycles or 100 thermal cycles together with the as-fabricated results. The mean strength ($\pm$ one standard deviation) from the thermal shock data was 29 ± 3 MPa. Fracture took place entirely in the quartz or epoxy for two of the three tests while the entire bond remained intact. Therefore, the mean strength represented a lower limit value to the actual strength of the joints. The thermally cycled samples had a mean strength of 27 ± 12 MPa. The larger spread, ± 12 MPa, was caused by the extremely high value of 45 MPa in test #4. The mean strength level was likewise a lower limit as failures occurred in the quartz substrate or in the epoxy; the solder joints remained completely intact. It was apparent that the thermal exposure tests enhanced the adhesion strength of the joints to beyond the failure strength of the quartz substrate or the epoxy.

Table 6. Adhesion Strength of the 2.0 Micron Thick Au-Ge Bonds After Thermal Shock and Thermal Cycle Exposures.

Pre-Test Environment	Test #	Strength (MPa)	Failure Mode
As-Fabricated	1	16	Bond
	2	19	Bond
	3	15	Bond
	4	2[illegible]	Qtz. + Epoxy
200 Thermal Shock Tests	1	32	Qtz. + Epoxy
	2	29	Bond
	3	27	Qtz. + Epoxy
100 Thermal Cycle Tests	1	20	Qtz. + Epoxy
	2	19	Qtz. + Epoxy
	3	24	Qtz. + Epoxy
	4	45	Qtz. + Epoxy

The data from the 1.0×10^{-6} m thick bonds after thermal shock and thermal cycle exposures is presented in Table 7. No significant change was observed between the strength and failure modes of the thermal shocked and the as-fabricated bond strengths. In the thermal cycle tests, the premature failure of the epoxy in test #1 occurred well below the mean value of the as-fabricated tests. Moreover, the strength values of tests #1 and #4 were less than those of #2 and #3, in which the bond played a role in the failure. Therefore, tests #1 and #4 reflected premature failure of the epoxy and quartz, respectively, and were not included with the other data to establish a lower limit strength value. The mean strength from tests #2 and #3 was 20 MPa. This value was not significantly different from the as-fabricated mean strength. The #2 and #3 bonds did not remain entirely intact during fracture, though, as some failure did proceed through the quartz substrate in both cases. It was apparent that the strengthening trend noted in the 2.0×10^{-6} m thick joints was not reproduced in the 1.0×10^{-6} m thick bonds. One expects the similarity of exposure conditions to also improve the adhesion strength of the 1.0×10^{-6} m thick samples. However, such effects may have been offset by a higher residual stress level after initial sample fabrication and during the thermal fluctuations caused by the more constrained geometry of the thinner joint. The stress levels resulted in a degradation of the adhesion strength at the solder-quartz interface. In conclusion, the 1.0×10^{-6} m thick bonds retained sufficient adhesion strength to satisfy the accelerometer's performance requirements both in the as-fabricated condition and after thermal shock and cycling exposures.

The decision was then made to direct further study to the 2.0×10^{-6} m thick joints only. The exemplary performance of the thicker solder film and the possible degradation in the 1.0×10^{-6} m joints were key to this decision.

Table 7. Adhesion Strength of the 1.0 Micron Thick Au-Ge Bonds After Thermal Shock and Thermal Cycle Exposures.

Pre-Test Environment	Test #	Strength (MPa)	Failure Mode
As-Fabricated	1	15	Bond
	2	18	Bond
	3	16	Bond
200 Thermal Shock Tests	1	14	Bond
	2	18	Bond
100 Thermal Cycle Tests	1	13	Epoxy
	2	14	Qtz. + Bond
	3	26	Qtz. + Bond
	4	8	Qtz.

C. Adhesion tests - annealed solder joints.

Processing of the QDA after assembly of the tuning forks and the pacers required (1) the attachment of the device to the package frame using a polyimide adhesive and (2) the installation of the sapphire lids to the ceramic frame by means of an 80Au-20Sn (wt.%) solder seal. The polyimide attachment was performed in a flowing argon atmosphere. The Au-Sn solder seal of the lids and frame was performed in vacuum. To encompass the elevated temperatures reflected by both processes, the bond samples were exposed to the following annealing cycle: heating ramp of 7°C/min to 300°C; hold for 180 min; and then furnace cooling. The heat treatment was performed in both air and vacuum. The results appear in Table 8. The two samples exposed to the air anneal demonstrated exceptional increases of bond strengths to 40 MPa in each case. When the heat treatment was performed in vacuum, no significant increase was observed; the mean strength was 16 ± 9 MPa. The large spread in the data was caused by the low value of 3.0 MPa—there was no evidence in the fracture surface to suggest disqualification of that test. However, even without that sample, the mean strength of 20 MPa was well below the air-annealed values.

Table 8. Adhesion Strength of 2.0 Micron Thick Bonds Exposed to 300°C, 180 Min Air and Vacuum Annealing Cycles.

Pre Test Environment	Test #	Strength (MPa)	Failure Mode
As-Fabricated	1	16	Bond
	2	19	Bond
	3	15	Bond
	4	21	Qtz. + Epoxy
Annealed: 300°C - 180 min Air	1	40	Bond
	2	40	Bond
Annealed: 300°C - 180 min Vacuum	1	5	Bond
	2	17	Bond
	3	24	Bond
	4	20	Bond

These test results provided further evidence of the increased bond strength resulting from exposure of the joints to elevated temperature in the presence of an air atmosphere. A similar behavior was observed in a series of thin film adhesion tests of 45 nm Cr – 180 nm Au coatings on polished, single-crystal quartz substrates. In spite of the presence of interdiffusion between the Au and Cr [8], the adhesion strength of the film was increased by annealing in air and, to a lesser degree, in vacuum. It was hypothesized that the strength increase was caused by an oxide-oxide bonding at the metallization-quartz interface, particularly between the SiO_2 and the strong oxide former, Cr. A similar situation exists in the Au-Ge thin film solder bond except that both Cr and Ge play a role in the enhancement of the adhesion strength.

D. Multilayer Au-Ge film deposition logistics for the accelerometer.

The 2.0×10^{-6} m thick multilayer Au-Ge film was deposited onto both sides of the spacer parts (Fig. 1). The sequence of two layers of Au (each 658 nm thick) and a Ge layer (684 nm thick) between the Au films was used in the device assembly. The initial Au layer adhered very well to the Au film already on the spacers as part of the photolithographic process. An entire quartz wafer of 20 sets of spacers was coated at a time. The wafer was placed between two 1.27×10^{-4} m thick beryllium-copper plates, which had been chemically etched so as to leave apertures 9.1×10^{-4} $\times$ 1.2×10^{-3} m through which the Au and Ge passed onto the spacer. The alignment of the wafer, a spacer mask, and the aperture mask was maintained by two pins mounted in a steel base plate. A second plate was secured to the top of the assembly to prevent movement of the parts or masks, thereby allowing the aperture dimensions to be replicated by the film on the spacer.

Assembly of the accelerometers was performed either individually or as a batch as a wafer in which all twenty tuning forks from each of two wafers were bonded to 20 sets of spacers from a third wafer. The fixturing used for both processes was similar. The stacking of tuning fork and spacer wafers in the batch assembly process was performed on a base plate that maintained alignment of the parts to within $\pm5\times10^{-6}$ m in all dimensions. A cover plate was placed on top of the quartz wafers. The cover plate had a small spring-loaded pin located over each of the 40 bond joints. When the cover plate was fastened to the base plate, a small amount of pressure was applied to each joint to eliminate gaps between the parts caused by a slight warping of the quartz wafers.

The solder films were reflowed by heating in a vacuum of 1.3×10^{-3} to 13×10^{-3} Pa. The temperature climb to the 390°C reflow temperature included a 10°C/min ramp to 250°C followed by a 60 min soak to assure a uniform temperature of the fixturing and parts. The temperature was ramped from 250°C to 390°C at 10°C/min. The reflow time was extended from the 15 min value used in the study to 30 min in order to compensate for the thermal mass of the fixture. The parts were allowed to cool in the furnace prior to removal. Each accelerometer unit was broken from the wafers, mounted, and sealed in the ceramic package.

E. Performance of the solder-bonded accelerometers.

The realization of thin, uniform solder bonds was confirmed by cross-section micrographs of selected units. Also, alignment between the two tuning forks and the spacers was vastly improved over the devices assembled with the polyimide adhesive.

More important, the resonator performance of the tuning forks in the accelerometer assembly showed significant improvement.

A prime attribute of the double-ended tuning fork geometry was the temperature compensation offered by the fact that the output signal was a difference in frequency between the two individual fork frequencies. The difference frequency versus temperature plot of the activity imposed on the z-cut is a parabola centered about a "turnover" temperature of 20°C to 25°C; the frequency dropped off for higher or lower temperatures. Perfectly matched parabolas from the two tuning forks would cause the frequency difference, Δf, between the tuning forks to be relatively independent of temperature; that is, $d\Delta f/dT$ would be approximately zero.

A principal effect of misalignment within the sensor structure was a temperature shift of the two parabolas relative to one another, causing $d\Delta f/dT$ to NOT equal zero. Therefore, $d\Delta f/dT$ defined by fitting a straight line through the curve of Δf versus temperature was used as a figure-of-merit to describe the construction quality of each accelerometer. Shown in Fig. 12a is a histogram of the values of $d\Delta f/dT$ for a set of accelerometers made with the polyimide adhesive. The mean and one standard deviation of the distribution were 15 mHz/C and ± 24 mHz/C, respectively. In Fig. 12b is a histogram of units made with the Au-Ge solder. Note that the $d\Delta f/dT$ axis has been shortened with respect to that in Fig. 12a. The mean and one standard of the $d\Delta f/dT$ distribution was 5.0 mHz/C and ± 12 mHz/C, respectively. Clearly, these data demonstrated that use of the Au-Ge thin film solder bond improved device performance by moving $d\Delta f/dT$ closer to the ideal value of zero. Also, the variation in the performance as indicated by the sigma value was significantly reduced over the units assembled with the polyimide. In conclusion, the attributes of the solder bond technique were realized in the performance of the accelerometer assemblies.

4. Conclusions

1. Thin solder films of the eutectic alloy, 87.5 Au - 12.5 Ge, were used to join the single-crystal quartz tuning forks and spacers to construct an accelerometer.
2. The solder films were introduced onto the optically polished quartz substrate by the evaporation of multiple layers of Au and Ge, which together formed the eutectic composition. Total film thicknesses of 0.5×10^{-6}, 1.0×10^{-6}, and 2.0×10^{-6} m were studied.
3. Thermal-physical analysis of the solder films showed them to behave exactly as the cast solder foil. The interdiffusion process did cause a volume expansion relative to the multilayer film or the subsequent solder film.
4. The 1.0×10^{-6} and 2.0×10^{-6} m thick solder joints showed as-soldered strengths of 16 ± 3 and 18 ± 3 MPa, respectively. The variability of the as-deposited strength of the 0.5×10^{-6} m thick joints, e.g., 13 ± 10 MPa and 13 ± 8 MPa, from two separate test series caused it to be eliminated from further consideration.

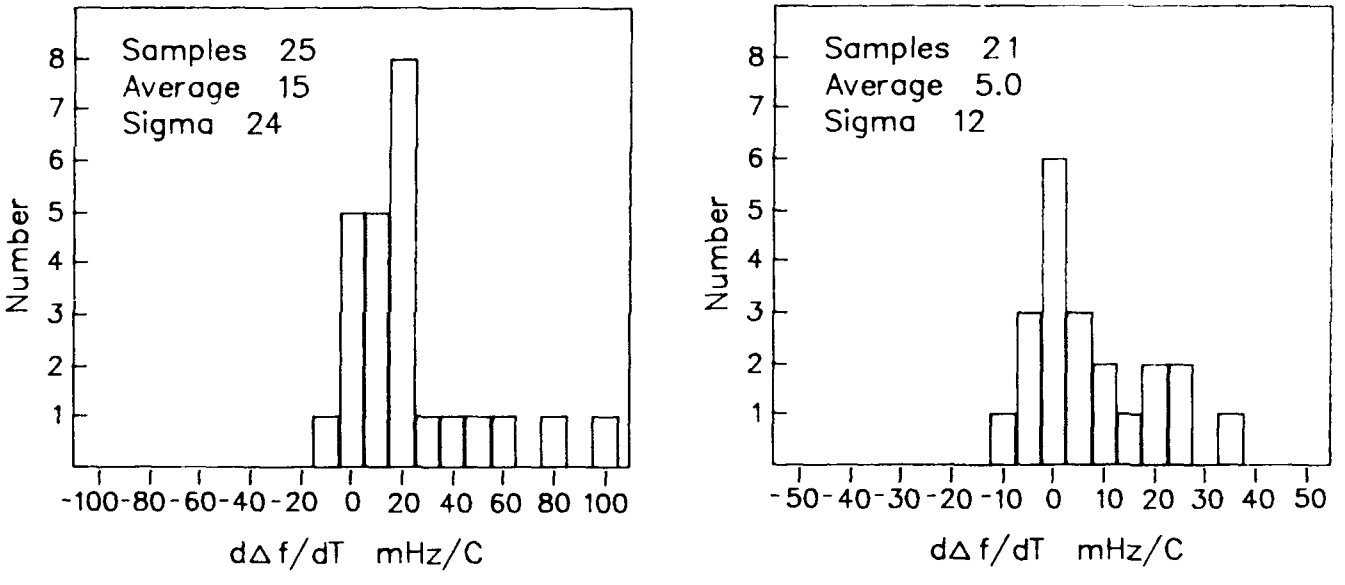


Figure 12. (a) Histogram of the values of d Δf/dT from accelerometers assembled with the polyimide adhesive. (b) Histogram of values of d Δf/dT from units assembled with the 2.0 micron thick Au-Ge solder bond.

5. The 1.0×10^{-6} and 2.0×10^{-6} m Au-Ge films were subjected to thermal shock and thermal cycle exposures, resulting in no significant change to the strength of the 1.0×10^{-6} m film and a substantial *increase* to the strength of the 2.0×10^{-6} m film. These data resulted in a preference for the 2.0×10^{-6} m thick films for device fabrication.
6. The strength improvement of the 2.0×10^{-6} m films was reproduced when samples were subjected to an air-annealing treatment which simulated next-step processing environments.
7. Use of the thin film Au-Ge solder on actual devices met the expectations of improved alignment between the two tuning forks, resulting in a significant improvement in the temperature performance of the sensor.

Acknowledgments

The authors wish to thank C. H. Sifford for performing the depositions, L. Roherty-Osmun for assembling the samples, and W. J. Kass for his review of the manuscript. This work was performed at Sandia National Laboratories for the U.S. Dept. of Energy under contract DE-AC04-76DP00789.

References

[1] D. Koehler, "Double Resonator Cantilever Accelerometer", U.S. Patent No. 4,479,385, Oct. 30, 1984.

[2] J. C. Brice, "Crystals for Quartz Resonators", *Rev. Mod. Phys.*, 57, (1985), 118.

[3] E. Reedy and W. Kass, "Finite-Element Analysis of a Quartz Digital Accelerometer," *IEEE Trans. on Ultra., Ferro., and Freq. Control*, 37, (1990).

[4] *Binary Alloy Phase Diagrams*, T. B. Massalski, ed. (ASM; Metals Park, OH, 1986), p. 264.

[5] J. H. Gieske, "Elastic Moduli of AuGe Eutectic Alloy from 23°C to 340°C", (Nondest. Test Report, Sandia National Laboratories, 1989).

[6] M. S. Lambert, "Thermal Expansion Coefficient Measurements on AuGe Eutectic Alloy Rod", (tech. memo., Sandia National Laboratories, 8-11-89).

[7] S. Timoshenko, "Analysis of Bi-Metal Thermostats," *J. Optical Soc. Amer.*, 11, (1925), p. 233.

[8] P. T. Vianco, W. R. Conley, and J. K. G. Panitz, "Resistivity, Adhesive Strength, and Residual Stress Measurements of Thin Film Metallizations on Single Crystal Quartz", *Proc. 44th Ann. Symp. on Freq. Control*, (1990), pp. 207-215.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ACCELERATION EFFECTS IN CRYSTAL FILTERS -- A TUTORIAL

R. C. Smythe

Piezo Technology Inc.
Box 547859, Orlando, FL 32854-7859

Abstract

The quasi-static method for the analysis of vibration-induced modulation in crystal filters is briefly reviewed, and a dynamic method, in which the filter is treated as a linear network with time-varying elements, is introduced. The method, which allows determination of both amplitude and phase modulation due to vibration is illustrated by examples. It is then applied to the analysis of the spectrum clean-up case, consisting of a frequency source with an output filter, both of which are undergoing the same acceleration.

Introduction

Under vibration, a crystal filter modulates signals passing through it, adding its own vibration-induced sidebands to those of the signals it passes. Both amplitude and phase modulation may occur. After briefly reviewing the quasi-static method, this paper presents a dynamic method for analyzing vibration-induced modulation in crystal filters, in which the filter is treated as a time-varying linear network. The dynamic analysis is then applied to the spectrum clean-up case in which a frequency source, such as a crystal oscillator, is followed by an output filter.

It is assumed throughout that the modulation is due entirely to the vibration-induced frequency variation of the crystal resonators, but the approach can readily be extended to include other components where appropriate. Also, while the discussion is restricted to bandpass filters, the method of analysis is also applicable to bandstop filters, and to filters other than quartz crystal filters.

Resonator G-Sensitivity

The acceleration sensitivity of quartz crystal resonators and oscillators has recently been reviewed by Filler [1, 2]. For accelerations which are not too large, the circular frequency, $\omega_s(\vec{a})$, of a resonator subjected to an acceleration, $\vec{a}$, is

$$\omega_s(\vec{a}) = \omega_s(1 + \vec{\Gamma} \cdot \vec{a}) , \quad (1)$$

where $\vec{\Gamma}$ is the acceleration-sensitivity vector of the resonator and ω_s is its zero-acceleration frequency.

For sinusoidal vibration,

$$\vec{a} = \vec{A}\cos\omega_v t , \quad (2)$$

$$\omega_s(t) = \omega_s + \Delta\omega_s \cos\omega_v t , \quad (3)$$

where

$$\Delta\omega_s / \omega_s = \vec{\Gamma} \cdot \vec{A} . \quad (4)$$

Quasi-Static Analysis

Although a filter under vibration is a time-varying linear system, at vibration frequencies which are sufficiently small, the vibration-induced modulation can be estimated from a quasi-static analysis [3]. Consider first the simplest situation, in which all the resonators have identical acceleration sensitivity vectors, identically oriented, so that their vibration-induced frequency changes are equal. Clearly, a constant acceleration produces a simple translation in frequency of the steady-state filter response. If, instead of a constant acceleration we have a very low-frequency sinusoidal vibration, the filter response can be thought of as oscillating back and forth along the frequency axis at the vibration frequency. That is, the filter center frequency, ω_o, varies according to equation (3), with ω_o replacing ω_s. This is, of course, an approximation, but it is a good one, provided that the vibration frequency is sufficiently low. A drawback is that the method does not provide an estimate of the error arising from the quasi-static assumption.

CH2965-2/91/0000-281 $1.00 © 1991 IEEE

The effect on a sinusoidal signal of frequency ω is then well-approximated by its first-order change in phase and amplitude,

$$\Delta\phi_v = \Delta\omega_o \cdot (d\theta/d\omega) \quad (5a)$$

and

$$\Delta V_v = \Delta\omega_o \cdot (dG/d\omega) . \quad (5b)$$

where ϕ and V are the phase and amplitude of the signal, and θ and G are the phase shift and gain of the filter.

For signals in the passband, $dG/d\omega$ is usually small enough to be neglected, so that the principal effect of vibration is phase modulation in accordance with equation (5a). Now, $d\theta/d\omega$ is just the negative of the group delay, $Tg(\omega)$, normalized values for which can be found in numerous references. Hence, for sinusoidal vibration the peak deviation is

$$\phi_v = (\vec{\Gamma} \cdot \vec{A})\omega_o Tg(\omega_o) . \quad (6)$$

A simple example will illustrate the use of the quasi-static method. Consider a two-pole Butterworth filter, with

$$|\vec{\Gamma}| = 1 \cdot 10^{-9}$$

$$f_o = 10 \text{ MHz}$$

$$BW_3 = 1 \text{ kHz} ,$$

where BW_3 is the 3 dB bandwidth of the filter. At center frequency the group delay is given by

$$Tg(\omega_o) = 1.414/\pi BW_3 = 0.45 \text{ ms.}$$

Then for a 5 G peak acceleration, the peak phase deviation is

$$\phi_v = 1 \cdot 10^{-9} \cdot 2\pi \cdot 10^7 \cdot 5 \cdot 0.45 \cdot 10^{-3}$$

$$= 1.414 \cdot 10^{-4} \text{ radians,}$$

which corresponds to $\mathcal{L} = -83$ dBc. At center frequency,

$$dG/d\omega = 0 ,$$

so that only phase modulation is present. Note that as a result of the quasi-static assumption, these results are independent of the vibration frequency (within the limits of that assumption.)

The simplicity of the foregoing analysis is partly a consequence of assuming that all the resonators have the same G-sensitivity. When this is not the case, a quasi-static analysis can still be carried out using network sensitivities -- the effect on the filter gain and phase shift of small changes in individual resonator frequencies -- but considerably more computation will be required.

Because of its simplicity, the quasi-static view is quite useful where valid. However, this is not always the case. A case of particular importance is the class of spectrum clean-up filters, or post-filters, discussed in a later section, in which the vibration frequencies may exceed the filter half-bandwidth. Here the quasi static method is clearly inadequate.

Dynamic Analysis

To overcome the limitations of the quasi-static assumption, the filter must be considered as a linear network with time-varying elements. As for many time-varying linear systems, the steady-state solution can be obtained, after some preliminaries, by essentially frequency-domain methods.

Consider a resonator whose equivalent circuit is given by figure 1a. For present purposes figure 1a represents the motional impedance of a quartz crystal resonator subjected to vibration; however, it could equally-well represent many other types of resonator. If its resonance frequency varies sinusoidally, as in equation (3), small frequency variations can be accurately represented by a sinusoidal variation of inductance L(t)

$$L(t) = L_0 + 2L_1 \cos\omega_v t , \quad (7)$$

where

$$L_1 = (-\Delta\omega_s/\omega_s)L_0 . \quad (8)$$

Substituting Eq. (4) in Eq. (8),

$$L_1 = (-\vec{\Gamma} \cdot \vec{A})L_0 . \quad (9)$$

it is shown in the Appendix that the effect of L_1 can be represented by a set of equivalent controlled voltage sources, one for each sideband. For the small perturbations usually of interest it is adequate to include only the first upper and first lower sidebands, as shown in figure 1b; however, the method is general.

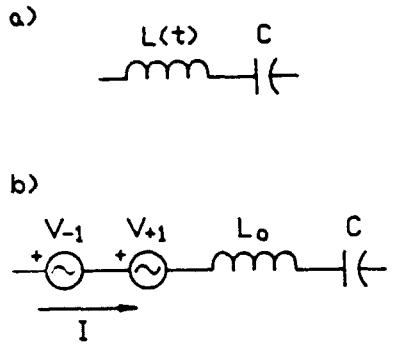


Fig. 1. Equivalent circuit of a resonator whose frequency is perturbed due to vibration. a) variation of resonance frequency is represented by L(t); b) small variations of resonance frequency can be represented by controlled voltage sources.

To illustrate, we again consider the two-pole Butterworth filter. The coupled-resonator equivalent network of figure 2, which can represent several physical realizations, both discrete-crystal and monolithic, will be used. In series with each motional arm are two controlled voltage sources, one for the lower sideband and one for the upper sideband. As outlined in the Appendix, the network is first analyzed at the carrier frequency, obtaining not only the output voltage, but also the carrier frequency currents in each resonator, which are needed to determine the amplitude and phase of the sideband generators. Once the controlled-source parameters have been determined, they can be used either simultaneously, or, since superposition applies, individually to calculate the output at each sideband frequency. Finally, carrier and sideband phasors are added together at the filter output and then demodulated to separate the amplitude- and phase-modulation components.

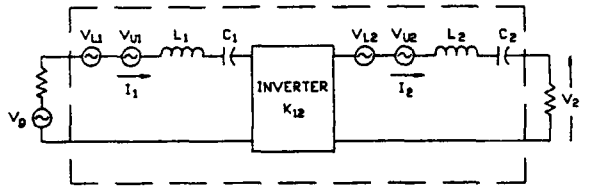


Fig. 2. Two-pole filter equivalent circuit.

The generality of this procedure makes it straightforward to consider cases in which the highest vibration frequency is not restricted by the quasi-static assumption. In addition, the resonators need not have identical G-sensitivities, and the relative contribution of each resonator is readily determined. All of these situations are easily handled once the analysis has been set up.

Figure 3 plots the single sideband power as a function of the vibration frequency when both resonators have the same acceleration sensitivity. The quantities in figures 3, 4, and 6-9 are normalized as follows: vibration frequency is normalized by one-half the 3 dB bandwidth of the filter, BW/2; single sideband power is normalized by $[P_c(\vec{\Gamma} \cdot \vec{A} \cdot \omega_o/BW)^2]$, where P_c is the carrier power.

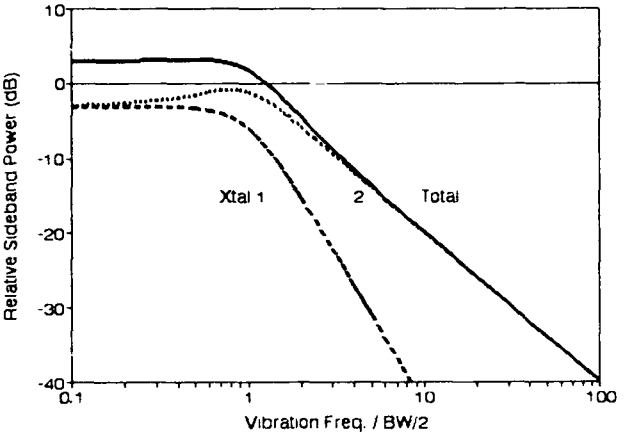


Fig. 3. Vibration-induced sideband power (normalized) for a two-pole Butterworth filter with $\vec{\Gamma}_1 = \vec{\Gamma}_2$, showing the relative contributions of the two resonators when the carrier is at band center.

In figure 3 the carrier is at band center; hence, upper and lower sideband levels are equal. At low frequencies the level corresponds to that given by quasi-static analysis. Also shown are the sideband levels due to each resonator. As would be expected from quasi-static theory, for low frequencies the two resonators contribute equally. However, for vibration frequencies in the filter stopband, the principal contributor is resonator 2, since it is connected directly to the load, while the sidebands generated by resonator 1 are attenuated by resonator 2. Since the Butterworth transmission characteristic is symmetric, and the carrier is at the filter center frequency, there is no amplitude modulation at ω_v, although there are small AM components at even multiples of ω_v; therefore, the sideband levels plotted in figure 3 are entirely due to phase-modulation.

It can be seen from figure 3 that if the two resonator G-sensitivity vectors are equal in magnitude and oriented anti-parallel, then at $\omega_v = 0$ their contributions exactly cancel, as shown in figure 4. For high vibration frequencies, no cancellation takes place, since the output

modulation is due almost entirely to resonator 2. While resonators having identical sensitivity vectors are not easily come by at the present state of the art, it is nevertheless useful to understand the basic effects.

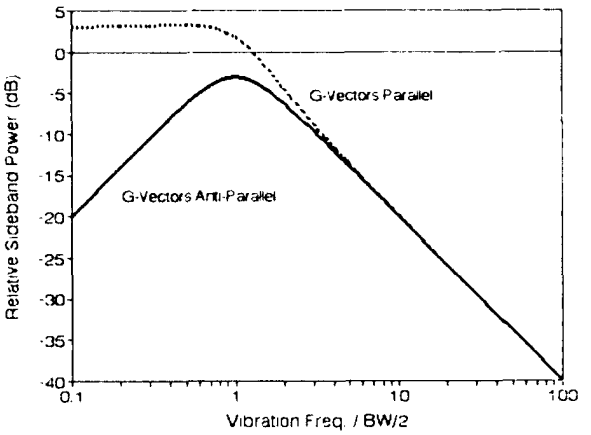


Fig. 4. Vibration-induced sideband power (normalized) for a two-pole Butterworth filter with $\vec{\Gamma}_1$ & $\vec{\Gamma}_2$ parallel and anti-parallel.

When the carrier does not lie at band center, the vibration induces both amplitude and phase modulation. Figure 5 shows the relative AM and PM levels as a function of carrier offset from band center, when the vibration frequency is $BW_3/4$. For small offsets, the modulation is principally phase modulation, while at large offsets it is principally amplitude modulation, the cross-over occurring near $(\omega_c - \omega_o) = \pm 1.4 \cdot BW/2$. Modulation is plotted in normalized form: phase modulation is $\phi_{peak}/(\vec{\Gamma} \cdot \vec{A} \cdot \omega_o/BW)$, where ϕ_{peak} is the peak phase deviation; the normalized amplitude modulation is $m_{am}/(\vec{\Gamma} \cdot \vec{A} \cdot \omega_o/BW)$, where m_{am} is the modulation index.

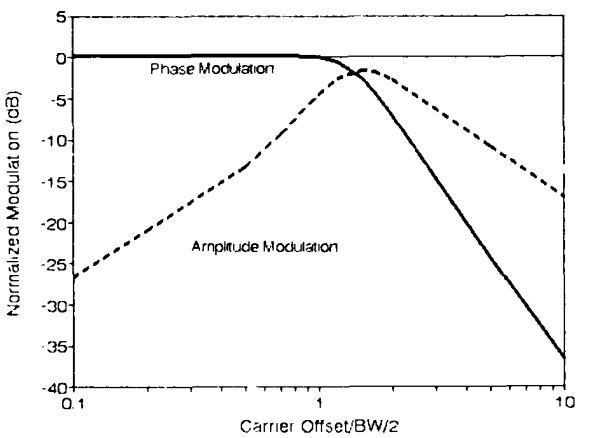


Fig. 5. Amplitude and phase modulation as a function of carrier offset for a two-pole Butterworth filter with $\vec{\Gamma}_1 = \vec{\Gamma}_2$.

Figure 6 compares the vibration-induced sideband levels for 2-, 4-, and 6-pole Butterworth filters in which all resonators have the same g-sensitivity and the carrier is at band center. The coupled-resonator equivalent circuit model is used in all cases. At low vibration frequencies, the filters have the sideband levels predicted by the quasi-static model, while at high vibration frequencies, all three filters have asymptotically identical sideband levels. In this region the sidebands are due almost entirely due to the resonator nearest the load.

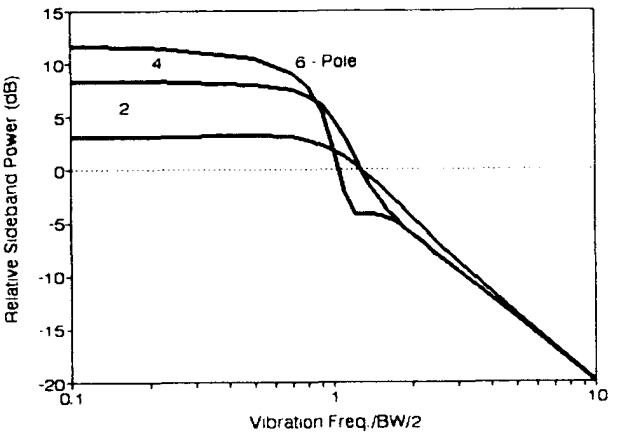


Fig. 6. Comparison of normalized vibration-induced sideband power for 2-, 4-, & 6-pole Butterworth filters. Carrier is at band center, and all resonators have the same $\vec{\Gamma}$.

Because all resonators are assumed to have the same g-sensitivity, the results in figure 6, although obtained for filters having the coupled-resonator topology, apply for other topologies as well. However, the relative contributions of individual resonators to the filter's vibration performance are of course topology-dependent.

Spectrum Clean-up Filters

Many frequency sources, including both oscillators and synthesizers, use spectrum clean-up filters to reduce various spurious output components, and to lower noise floors. Spectrum clean-up filters, or post-filters, are also sometimes considered as a means of reducing vibration-induced sideband levels in a frequency source. Therefore, we now consider a frequency source whose nominally sinusoidal output is frequency-modulated due to vibration and then passed through a filter which is also subject to vibration. (The two vibration environments will be assumed to be the same, although this is not necessarily the case; e.g., if either the source or the filter is protected by vibration isolators.) The filter in turn modulates both

the carrier and the vibration sidebands of the source; however, modulation of the latter, being a second-order effect, will be neglected. Hence, the vibration sidebands of the filter output are essentially the vector sum of the sidebands of the source, attenuated and phase-shifted by the filter, and the sidebands due to the modulation of the carrier by the filter.

Since clean-up filters are ordinarily as narrow as possible, quasi-static methods may not be applicable. This was shown for a single-resonator (one-pole) filter by Horton and Morley [4] who observed that for sinusoidal vibration at frequency f_v, equation (3), and a sinusoidal input at a carrier frequency, f_c, the output is a phase-modulated signal for which, for small phase deviation,

$$\mathcal{L}(f_v) = 20 \log(\vec{\Gamma} \cdot \vec{A} \cdot Q_L / [1 + (2Q_L f_v / f_c)^2]^{1/2}) , \quad (10)$$

where Q_L is the loaded Q of the resonator. The dynamic analysis method confirms this result. Since the 3 dB bandwidth of a one-pole filter is

$$BW = f_c / Q_L, \quad (11)$$

equation (10) can be written

$$\mathcal{L}(f_v) = 20 \log(\vec{\Gamma} \cdot \vec{A} \cdot (f_c / BW) / [1 + (f_v / BW/2)^2]^{1/2}). \quad (12)$$

From equations (10) and (12) it can be seen that the phase modulation exhibits a frequency dependence corresponding to a one-pole filter gain characteristic. For vibration frequencies within the filter passband, the denominator in equation (12) is approximately 1 and the vibration-induced phase noise is nearly constant at the level predicted by quasi-static analysis. On the other hand, for vibration frequencies far removed from the passband, equations (10) and (12) are asymptotic to the expression [1] for the single-sideband level for a frequency source having the same acceleration sensitivity vector,

$$\mathcal{L}(f_v) = 20 \log(\vec{\Gamma} \cdot \vec{A} \cdot f_c / 2f_v) . \quad (13)$$

Thus, at low vibration frequencies the filter's contribution to vibration-induced phase noise is less than that of a crystal oscillator using the same resonator (or, in general, any frequency source having the same acceleration sensitivity) while at high vibration frequencies the two contributions are the same. This is illustrated in figure 7. The curve labelled "filter" represents the phase modulation of a sinusoidal carrier by the filter; "oscillator," the phase modulation present on the unfiltered oscillator output; and "filter + oscillator" the phase modulation at the filter output, discussed below. The vibration frequency is normalized to one-half the 3 dB bandwidth, BW, of the filter; the sideband levels are shown relative to the unfiltered oscillator sideband level at BW/2.

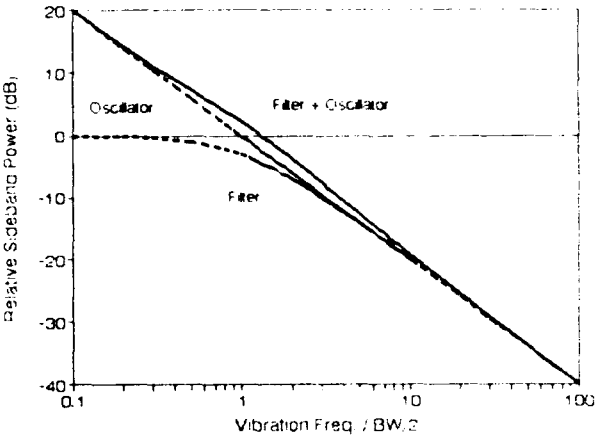


Fig. 7. Relative sideband levels for an oscillator with and without a one-pole post-filter when the filter resonator has the same acceleration sensitivity as the oscillator.

The curve labelled "Filter + Oscillator" represents the vector sum of the sidebands of the source, attenuated and phase-shifted by the filter, and the sidebands added by the filter. It can be seen that for identical gamma vectors identically oriented, the single-pole filter does not improve the output phase modulation, but rather degrades it by as much as 2.5 dB in the vicinity of the passband edges. (However, if the gamma vectors of the source and filter were oriented anti-parallel instead of parallel, some improvement would occur in the same region.) For vibration frequencies much greater than or much less than BW/2, the vibration-induced sideband level is essentially unchanged by the filter. This means that the filter may still play a useful role in cleaning up spurious responses, etc., but in order to reduce the vibration-related sidebands the filter resonator's acceleration sensitivity must be less than the frequency source's.

Figure 8 shows the sideband levels for a source with a two-pole Butterworth post-filter when the acceleration sensitivities of the two filter resonators are the same in magnitude and orientation as that of the oscillator and the center frequency of the filter equals the oscillator frequency. It is seen that the filter provides no net improvement in the vibration-induced sidebands and a degradation of 3 dB at the filter passband edge. If the

filter crystals are oriented anti-parallel to obtain cancellation, as for figure 4, the only benefit in this application is to remove this 3 dB "bump".

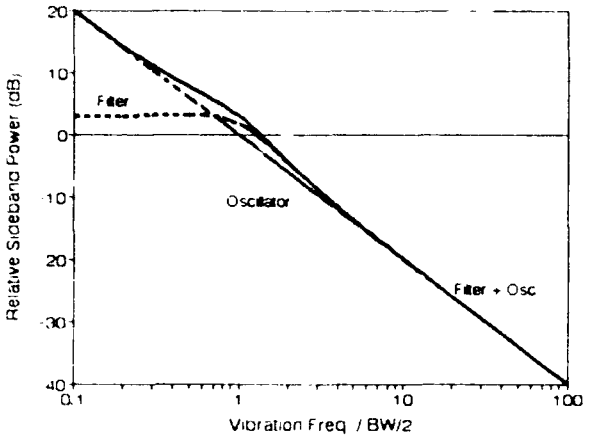


Fig. 8. Relative sideband levels for an oscillator with and without a two-pole Butterworth post-filter when the filter resonator has the same acceleration sensitivity as the oscillator.

The net effect of the filter is more easily seen in figure 9, which shows the vibration sideband reduction for the two-pole Butterworth as a function of normalized vibration frequency, with the gamma ratio, defined as $|\Gamma_{\text{filter crystal}}|/|\Gamma_{\text{source}}|$, as a parameter. Three curves are shown. A gamma ratio of 1 corresponds to the equal-gamma case, figure 6; a ratio of 0.5, to the case in which the acceleration sensitivity of the filter crystals is one-half that of the source; and a ratio of 0.2 to the case in which the acceleration sensitivity of the filter crystals is one-fifth that of the source.

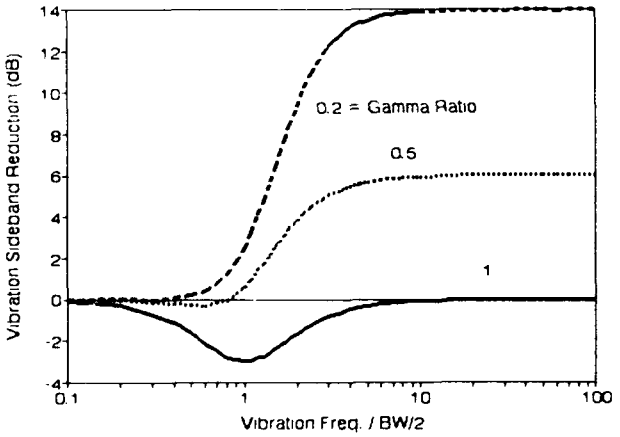


Fig. 9. Reduction in vibration sideband levels for a source with a two-pole Butterworth post-filter for three values of the gamma ratio (see text.)

Conclusion

Acceleration effects in bandpass crystal filters have been analyzed by treating the filter as a linear, time-varying network, thereby overcoming the limitations of the quasi-static model. By using controlled-voltage sources to represent the vibration-induced frequency variation of the crystal resonators, the problem is reduced to the analysis of a linear network with constant coefficients. With this approach, not only can sideband levels be accurately calculated, but also, by demodulating the calculated filter output, amplitude modulation and phase modulation effects can be separately determined.

Using this method of analysis, the use of spectrum clean-up filters for improving the vibration-induced sideband level of frequency sources has been examined and found to be of limited value.

While the entire discussion has been in terms of sinusoidal vibration, because of linearity the methods apply directly to multi-frequency and random vibration. Further, the method of analysis presented is not limited to crystal filters, but can be applied to any filter which can be represented by a lumped-element equivalent circuit.

Acknowledgments

A portion of this paper first appeared in a report [5] by the members of Working Group 4 (Acceleration Effects) of the IEEE Standards Coordinating Committee 27 on Time and Frequency. The author wishes to thank Dr. John Vig, Chairman of the Working Group, for suggesting that this work be expanded and presented separately. The author also wishes to thank M. M. Driscoll for asking a very important question.

Appendix: Circuit Representation and Analysis

For a linear network excited by a sinusoidal source whose frequency is ω_c and having one or more elements which are time-varying at a rate ω_v, it is readily apparent that the only frequencies present in the steady state are ω_c and sidebands at intervals of ω_v

$$\omega_i = \omega_c + i \cdot \omega_v \quad (i = ...,-1,0,+1,...). \qquad (14)$$

Moreover, since the network is linear, superposition applies; we can therefore consider the variation of each element separately and then add up their combined effects. In this Appendix we model the sinusoidal variation of resonator frequency by a

perturbation of the motional inductance. (This variation may be due to vibration, as in the present application, or to any other mechanism.) We then use Bolle's complex symbolism [6] to obtain the voltage-current relations for a sinusoidally varying inductor. In general, the relations include summations which are doubly infinite, but when the element variation is sinusoidal they conveniently simplify to a finite summation. This procedure yields an equivalent circuit in which the time-varying inductor is replaced by a fixed inductor and a voltage generator at each significant sideband frequency. While the method is quite general, we will consider only the case in which sidebands of second and higher orders can be neglected.

For a resonator whose frequency is varying sinusoidally at a rate ω_v, small frequency variations can be accurately represented by a sinusoidal inductance variation, Eq. (7). If we call $e_i = E_i\cos\omega_i t$ the voltage across L, and $i_i = I_i\cos\omega_i t$ the current through L, at frequency ω_i, then from equation (9) of Bolle, the voltages at, respectively, the first upper sideband, carrier, and first lower sideband frequency are

$$E_1 = j\omega_1[L_1I_2 + L_0I_1 + L_1I_0] \quad (15a)$$

$$E_0 = j\omega_0[L_1I_1 + L_0I_0 + L_1I_{-1}] \quad (15b)$$

$$E_{-1} = j\omega_{-1}[L_1I_0 + L_0I_{-1} + L_1I_{-2}]. \quad (15c)$$

Up to this point, we have made no simplifying approximations. In equations (15a) and (15c), for the upper and lower sideband voltages, we now neglect terms containing second-order sideband currents. We also neglect the terms containing first-order sideband currents in equation (15b). These simplifications are justified for most cases of interest. We then obtain

$$E_1 = j\omega_1[L_0I_1 + L_1I_0] \quad (16a)$$

$$E_0 = j\omega_0L_0I_0 \quad (16b)$$

$$E_{-1} = j\omega_{-1}[L_1I_0 + L_0I_{-1}]. \quad (16c)$$

In the first of these, E_1, the voltage across L at the upper sideband frequency, ω_1, is seen to be just the voltage drop across L_0 due to the upper sideband current, I_1, plus a voltage $V_1 = j\omega_1L_1I_0$. The latter can be represented by an equivalent controlled voltage source whose output is proportional to I_0. E_{-1} can be treated similarly, so that we obtain the equivalent circuit representation of figure 1b.

Using this equivalent circuit the complete network can then be analyzed as a linear network with constant coefficients using any modern linear network analysis program. The network is first analyzed at the carrier frequency with $L = L_0$, obtaining the current I_0, which allows us to calculate the equivalent generator voltage, V_1. The network, excited now by this voltage generator, is analyzed at the upper sideband frequency, ω_1. This step is then repeated at the lower sideband frequency.

In this manner, the contribution of each time-varying element can be calculated. For analysis programs which allow more than one generator, the total contribution of all the variable elements can be determined by a single analysis at each sideband frequency; otherwise, the effect of each element can be calculated separately, these results being combined to get the net effect. In the general case in which higher order sidebands must be considered, the method still applies, with obvious modifications; however, other methods may be preferable if there are a large number of sidebands.

From the network analyses, phasor representations of the carrier and each sideband of interest are obtained. These are then added together to construct the output signal, which is demodulated numerically to obtain the AM and PM components. This can be done most easily by shifting the carrier frequency to zero. If we call the frequency-shifted signal

$$v(t) = V(t)\cdot\exp(j\omega t + \phi(t)) , \quad (17)$$

then the peak phase deviation is given by

$$\phi_{pk} = \{\max[\phi(t)] - \min[\phi(t)]\}/2 , \quad (18)$$

and the amplitude-modulation index is given by

$$k = \{\max[V(t)] - \min[V(t)]\}/\{\max[V(t)] + \min[V(t)]\} , \quad (19)$$

where the maximum and minimum are taken over one period of the lowest vibration frequency. If it is known that the output is essentially only amplitude-modulated or only phase-modulated, then the modulation index and related quantities can be obtained directly from the sideband levels relative to the carrier.

References

[1] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: a Review," Proc. 41st Annual Symposium on Frequency Control, pp. 398-408, 1987.

[2] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: a Review," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, v. 23, no. 3, pp. 297-305, May, 1988.

[3] R. L. Clark and M. K. Yurtseven, "Spurious Signals Induced by Vibration of Crystal Filters," Proc. IEEE Ultrasonics Symp., pp. 365-368, IEEE Pub. 88CH2578-3, 1988.

[4] T. S. Payne, "Improved Tactical Miniature Crystal Oscillator," First Interim Report, SLCET-TR-86-0011-1, U.S. Army Laboratory Command, Fort Monmouth, New Jersey, Nov. 1987. The section of this report that deals with the vibration sensitivity of crystal filters was prepared by W. H. Horton and P. E. Morley.

[5] J. Vig, et al, "The Effects of Acceleration on Precision Frequency Sources," Research & Development Technical Rpt. SLCET-TR-91-3, U.S. Army Laboratory Command, Electronics Technology and Devices Laboratory, Ft. Monmouth, New Jersey, March 1991. AD A235470.

[6] A. P. Bolle, "Application of Complex Symbolism to Linear Variable Networks," IRE Transactions on Circuit Theory, v. CT-2, no. 1, pp. 32-35, March, 1955.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

THE INCREASE IN THE IN-PLANE ACCELERATION SENSITIVITY OF THE PLANO-CONVEX RESONATOR DUE TO ITS THICKNESS ASYMMETRY

H.F. Tiersten and Y.S. Zhou

Department of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

It has recently been shown that the in-plane acceleration sensitivity of contoured quartz resonators vanishes for a perfectly symmetric combined resonator plus support system. This indicates that a biconvex resonator will give lower resultant in-plane acceleration sensitivity than a plano-convex resonator because of its inherent additional symmetry. Since a plano-convex resonator is easier to fabricate, an analysis of the degradation in the in-plane acceleration sensitivity that arises from the contour being on one side only is performed. Under in-plane acceleration the single contour causes a state of flexure and asymmetric shear to exist in the resonator plate. The flexural biasing deformation is determined by means of our variational approximation procedure using the variational principle with all natural conditions for anisotropic static flexure. The very important shearing stresses varying through the thickness and accompanying strains are determined recursively. The resulting biasing states are employed in the existing perturbation equation along with the equivalent trapped energy mode shapes of the contoured resonators to calculate the degradation in the in-plane acceleration sensitivity caused by the loss of symmetry of the plano-convex resonator. The calculation shows that the in-plane sensitivity of the plano-convex resonator increases with decreasing radius of curvature and that the magnitude ranges over a few parts in 10^{12} per g.

1. Introduction

It has recently been shown[1] that the resultant in-plane acceleration sensitivity of contoured quartz resonators vanishes when the resonator and support system exhibit perfect symmetry with respect to each other. This was shown analytically and is a consequence of the symmetry of the biasing state and the mode shape. This implies that a biconvex resonator will tend to have lower resultant in-plane sensitivity than a plano-convex one on account of its inherent additional symmetry. In this work an analysis of the increase in the in-plane acceleration sensitivity resulting from the contour being on one side only is performed.

When the symmetrically supported plano-convex resonator is subject to in-plane acceleration, the fact that the contour is on one side only causes a state of flexure and shear to exist in the plate in addition to extension. The very important shearing stresses varying linearly through the thickness on account of the uniform extension through the thickness, which arises from the shear loading being applied on one side only, are determined recursively along with the accompanying deformation. The flexural biasing state arising from the point moments resulting from the single contour is determined by means of our variational approximation procedure[2,3]. The resulting biasing deformations are employed in the existing perturbation equation[4] along with the equivalent trapped energy mode shape[5] of the contoured resonator to calculate the increase in the in-plane acceleration sensitivity caused by the single contour. The calculated results show that the in-plane sensitivity increases with decreasing radius of curvature and lies in the range of a few parts in 10^{12} per g.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency obtained from the perturbation analysis[4] mentioned in the Introduction may be written in the form

$$\Delta_\mu = H_\mu / 2\omega_\mu, \quad \omega = \omega_\mu - \Delta_\mu , \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_\mu = - \int_V \hat{K}^n_{L\gamma} g^\mu_{\gamma,L} \, dV, \tag{2.2}$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g^μ_γ denotes the normalized mechanical displacement vector, and $\hat{K}^n_{L\gamma}$ denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g^μ_γ, and is given by

CH2965-2/91/0000-289 $1.00 © 1991 IEEE

$$\hat{K}^{n}_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M}, \tag{2.3}$$

where

$$\hat{c}_{L\gamma M\alpha} = T^{1}_{LM}\delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN}E^{1}_{KN} + c_{2L\gamma KM}w_{\alpha,K} + c_{2LK M\alpha}w_{\gamma,K}, \tag{2.4}$$

and

$$T^{1}_{LM} = c_{2LMKN}E^{1}_{KN}, \quad E^{1}_{KN} = \frac{1}{2}(w_{K,N} + w_{N,K}). \tag{2.5}$$

The quantities T^{1}_{LM}, E^{1}_{KN} and w_K denote the static biasing stress, strain and displacement field, respectively. The coefficients c_{2LMKN} and $c_{3L\gamma M\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g^{μ}_{γ} and $\hat{f}^{\mu}$ is defined by

$$g^{\mu}_{\gamma} \equiv \frac{u^{\mu}_{\gamma}}{N_{\mu}}, \quad \hat{f}^{\mu} = \frac{\hat{\varphi}^{\mu}}{N_{\mu}}, \quad N^{2}_{\mu} = \int \rho u^{\mu}_{\gamma} u^{\mu}_{\gamma}\, dV, \tag{2.6}$$

where u^{μ}_{γ} and $\hat{\varphi}^{\mu}$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\hat{K}^{\ell}_{L\gamma} = c_{2L\gamma M\alpha}u_{\alpha,M} + e_{ML\gamma}\hat{\varphi}_{,M},$$

$$\hat{\mathcal{D}}^{\ell}_{L} = e_{LM\gamma}u_{\gamma,M} - \epsilon_{LM}\hat{\varphi}_{,M}, \tag{2.7}$$

$$\hat{K}^{\ell}_{L\gamma,L} = \rho\ddot{u}_{\gamma}, \quad \hat{\mathcal{D}}^{\ell}_{L,L} = 0, \tag{2.8}$$

subject to the appropriate boundary conditions, and ρ is the mass density. Equations (2.7) are the linear piezoelectric constitutive relations and (2.8) are the stress equations of motion and charge equation of electrostatics, respectively. The upper cycle notation for many dynamic variables and the capital Latin and lower case Greek index notation is being employed for consistency with Ref.4 as is the remainder of the notation in this section.

The substitution of (2.3) in (2.2) yields

$$H_{\mu} = -\int_{V} \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} g^{\mu}_{\gamma,L}\, dV. \tag{2.9}$$

Since g^{μ}_{α} denotes the normalized mode shape in the contoured resonator and from (2.4) $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_{μ} can be evaluated when the resonator mode shape and biasing state are known.

3. Decomposition into Extensional and Flexural Biases

A plan view and cross-section of the plate with identical sidewalls on top and bottom are shown in Fig.1 along with the coordinate system. It should be noted that the sidewalls run all around the periphery. This defines the geometry of the support configuration considered in the static biasing problem. The support system is symmetric and the mode shape is located symmetrically with respect to it. Since, as noted in the Introduction, the acceleration sensitivity vanishes under these circumstances, we are interested in determining the biasing state resulting from the in-plane acceleration due only to the asymmetric portion of the plano-convex resonator shown at the bottom of Fig.1. This refers only to the matter between the spherical surface $X_2 = h$ and the plane surface $X_2 = h_1$. Accordingly, this results in the decomposition into flexural and extensional states, as shown in Fig.2, in which the actual loading applied to the surface $X_2 = h_1$ is decomposed into flexural and extensional loadings. The flexural loading consists of point moments per unit area $F^{(1)}_{B}$ applied along the center-plane $X_2 = 0$ of the flat plate of thickness $2h_1$, which in accordance with the immediately preceding discussion are given by[6]

$$F^{(1)}_{B} = -\int_{h_1}^{2h-h_1} X_2 \rho a_B\, dX_2 = -\rho a_B 2h(h - h_1), \tag{3.1}$$

where a_B denotes the acceleration in the X_B-direction. The corresponding extensional loading consists of shearing forces per unit area S_B applied along the surface $X_2 = h_1$, which in accordance with the same discussion are given by

$$S_B = -\int_{h_1}^{2h-h_1} \rho a_B\, dX_2 = -2\rho a_B(h - h_1), \tag{3.2}$$

where to lowest order in X_2 and X_3 the spherical surface is given by

$$h = h_o[1 - (X_1^2 + X_3^2)/4Rh_o]. \tag{3.3}$$

This applied shearing stress S_B is resisted by extension in the plane of the thin plate. As usual, this plane stress extension is taken to be uniform through the thickness. The directions of $F^{(1)}_{B}$ and S_B are as shown in Fig.2 because the accelerations a_B are positive.

Since the extensional portion of Fig.2 shows a shearing stress S_B on the top surface and zero on the bottom surface, there is a distribution of shearing stress throughout the thickness of the plate, for which we must solve. To this end we consider the free body diagram shown in Fig.3, where K_{B2} denotes the shearing stresses on any surface X_2 = constant in the range $+h_1$

to $-h_1$, and all stresses have been taken in the positive sense. The stress equations of equilibrium in the plane of the plate may be written

$$K_{AB,A} + K_{2B,2} = 0, \tag{3.4}$$

and we note that in the extension of thin plates the K_{AB} are independent of X_2. Equation (3.4) permits us to write

$$\int_{X_2}^{h_1} (K_{AB,A} + K_{2B,2})\, dX_2 = 0, \tag{3.5}$$

the integration of which yields

$$K_{2B}(h_1) - K_{2B}(X_2) = - K_{AB,A}(h_1 - X_2). \tag{3.6}$$

Since $K_{2B}(h_1) = S_B$ and $K_{2B}(-h_1) = 0$, on letting $X_2 = -h_1$ in (3.6) we obtain

$$K_{AB,A} = - S_B/2h_1. \tag{3.7}$$

The substitution of (3.7) in (3.6) yields

$$\hat{K}_{2B} = \frac{S_B}{2h_1}(h_1 + X_2), \tag{3.8}$$

where we have defined $\hat{K}_{2B}$ as $K_{2B}(X_2)$. Equation (3.8) gives the distribution of shearing stress through the thickness.

On account of the stubbiness of the cross-section of a sidewall, we model its resistance to the normal slope $\partial w_2^{(0)}/\partial n$ of the center-plane of the plate as a simple variable shear strip, as in earlier work[7], which yields

$$K^{(1)}_{(n)(n)} = \pm k \partial w_n / \partial n, \tag{3.9}$$

where in this work we have[8]

$$k = 2\mu h_1 t_f, \tag{3.10}$$

in which μ is the appropriate shear modulus depending on the direction in which the sidewall runs, and we have ignored the deformation in the sidewall to this approximation because $h_f \ll h$. Furthermore, it is clear that at the sidewall the vertical deflection $w_2^{(0)}$ vanishes and we have

$$w_2^{(0)} = 0, \tag{3.11}$$

at the sidewall.

4. Equations for Anisotropic Static Flexure

With the aid of Mindlin's plate equations[9-11] it has been shown that the equation of anisotropic static flexure with X_2 normal to the major surfaces can be written in the form[12]

$$K^{(1)}_{AB,AB} + F^{(1)}_{B,B} + F^{(0)}_2 = 0, \tag{4.1}$$

$$K^{(1)}_{AB} = \int_{-h}^{h} X_2 K_{AB}\, dX_2, \tag{4.2}$$

and in this instance $F^{(1)}_B$ is given in (3.1) and $F^{(0)}_2 = 0$. From Eqs.(4.25) of Ref.12 the constitutive equations for the stress-resultants take the form

$$K^{(1)}_{AB} = \frac{2}{3} h^3 \gamma_{ABCD} E^{(1)}_{CD}, \tag{4.3}$$

where Voigt's anisotropic plate elastic constants are given by

$$\gamma_{RS} = c_{RS} - c_{RW} c^{-1}_{WV} c_{VS}, \quad R,S = 1,3,5;$$
$$W,V = 2,4,6, \tag{4.4}$$

in the compressed notation, and where we have introduced the scheme shown. The plate strains $E^{(1)}_{CD}$ in (4.3) are given by

$$E^{(1)}_{CD} = \frac{1}{2}\left(w^{(1)}_{C,D} + w^{(1)}_{D,C}\right), \tag{4.5}$$

and from the relaxation of the stress resultants $K^{(1)}_{2L}$, the vanishing of the plate shear strains $E^{(0)}_{2A}$ and (4.5) we have the respective relations

$$E^{(1)}_W = - c^{-1}_{WV} c_{VS} E^{(1)}_S, \quad w^{(1)}_A = - w^{(0)}_{2,A},$$
$$E^{(1)}_{CD} = - w^{(0)}_{2,CD}, \tag{4.6}$$

which, respectively, are given in Eqs.(4.24), (3.34) and (3.37) of Ref.12.

It has been shown[3] that the biasing shearing stresses that are determined recursively in the classical theory of the flexure of thin plates may be written in the form

$$\hat{K}_{2B} = (3/4\, h^3)(h^2 - X_2^2)\, K^{(1)}_{AB,A}, \tag{4.7}$$

where the caret indicates that a quantity has been determined recursively and, hence, contains an order of smallness. The associated three-dimensional strain fields are obtained simply by inverting the linear constitutive equations

$$\hat{K}_{LM} = c_{LMKN} \hat{E}_{KN}, \tag{4.8}$$

with the result

$$\hat{E}_{KL} = c^{-1}_{KL2B} \hat{K}_{2B} = {}_o\hat{E}_{KL} + X_2^2\, {}_2\hat{E}_{KL}, \tag{4.9}$$

since all stresses that vary symmetrically with X_2 vanish except $\hat{K}_{2B}$, and ${}_o\hat{E}_{KL}$ and ${}_2\hat{E}_{KL}$ are clear

from (4.7) and (4.9). Similarly, the three-dimensional strain field associated with the shearing stresses $\hat{K}_{2B}$ that accompany any extension and are given in (3.8) are found by inverting the constitutive equations with the result

$$\hat{E}_{KL} = c^{-1}_{KL2B}\hat{K}_{2B} = {}_o\hat{E}_{KL} + X_2\,{}_1\hat{E}_{KL}. \quad (4.10)$$

Since from Eqs.(3.31) of Ref.5 we have $E^{(0)}_{KL} = 0$, when the plate deflection $w^{(0)}_2$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = {}_o\hat{E}_{KL} + {}_o\tilde{E}_{KL} + X_2(E^{(1)}_{KL} + {}_1\hat{E}_{KL}) + X_2^2\,{}_2\hat{E}_{KL}. \quad (4.11)$$

Although we now have the plate strains from (4.10), we cannot yet determine the $\hat{c}_{L\gamma M\alpha}$ from (2.4) because, as noted in Ref.5, we need the displacement gradients $w_{L,K}$ or equivalently the rotations

$$\Omega_{KL} = \frac{1}{2}(w_{L,K} - w_{K,L}). \quad (4.12)$$

As in Ref.5, we realize that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K^{(1)}_{2L}$ and the recursive inversion in (4.9) and (4.10) satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L}, \quad (4.13)$$

where

$$\Omega_{KL} = \sum_{n=0}^{2} X_2^n\,\Omega^{(n)}_{KL}, \quad (4.14)$$

and the $\Omega^{(n)}_{KL}$ are the plate rotations of nth order. As in the end of Sec.III of Ref.5 [after Eq.(3.32)] from the flexural solution without recursive strai· we obtain

$$\Omega^{(0)}_{13} = \text{arbitrary constant} = 0,$$
$$\Omega^{(1)}_{13} = 0,\ \Omega^{(0)}_{2A} = -w^{(0)}_{2,A}. \quad (4.15)$$

Including recursive strains for n = 1 and 2 and equating like powers of X_2 in the same way we obtain[1]

$$\hat{\Omega}^{(1)}_{13} = {}_o\hat{E}_{23,1} + {}_o\hat{E}_{23,1} - {}_o\hat{E}_{21,3} - {}_o\hat{E}_{21,3},$$
$$\Omega^{(2)}_{13} = \frac{1}{2}(E^{(1)}_{23,1} - E^{(1)}_{21,3} + {}_1\hat{E}_{23,1} - {}_1\hat{E}_{21,3}),$$
$$\Omega^{(1)}_{2A} = E^{(1)}_{2A} + {}_1\hat{E}_{2A} - {}_o\hat{E}_{22,A} - {}_o\hat{E}_{22,A},$$
$$\Omega^{(2)}_{2A} = {}_2\hat{E}_{2A} - \frac{1}{2}E^{(1)}_{22,A} - \frac{1}{2}\,{}_1\hat{E}_{22,A}. \quad (4.16)$$

The substitution of (4.15) and (4.16) in (4.14) yields Ω_{KL}, which with (4.11) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \quad (4.17)$$

gives us the desired three-dimensional displacement gradients when the plate deflection $w^{(0)}_2$ has been found.

5. Flexural Biasing Deformation

The flexural biasing problem due to an in-plane acceleration of a plano-convex resonator has been defined in Section 3 with the aid of Figs.1 and 2. The substitution of (3.1) with (3.3) into (4.1) yields

$$\frac{2}{3}h^3\gamma_{ABCD}w^{(0)}_{2,CDAB} - \frac{\rho}{R}a_Bx_B\left(2h_o - h_1 - \frac{x_Ax_A}{2R}\right) = 0, \quad |x_1| < c,\ |x_3| < c, \quad (5.1)$$

$$\frac{2}{3}h^3\gamma_{ABCD}w^{(0)}_{2,CDAB} = 0,\ |x_1| > c \text{ or } |x_3| > c, \quad (5.2)$$

where for computational convenience we have taken the liberty of replacing the circle of the intersection of the spherical contour with the flat plate by the circumscribed square. From Eqs.(3.11) and (3.9) with (3.10) we have the edge conditions

$$w^{(0)}_2 = 0 \text{ at } x_1 = \pm a \quad |x_3| \le b,$$
$$w^{(0)}_2 = 0 \text{ at } x_3 = \pm b, \quad |x_1| \le a,$$
$$K^{(1)}_{11} = \pm k_1 w^{(0)}_{2,1} \text{ at } x_1 = \pm a, \quad |x_3| \le b,$$
$$K^{(1)}_{33} = \pm k_3 w^{(0)}_{2,3} \text{ at } x_3 = \pm b, \quad |x_1| \le a, \quad (5.3)$$

where

$$k_1 = 2h_1\mu_1 t_f,\ k_3 = 2h_1\mu_3 t_f \quad (5.4)$$

and μ_1 and μ_3 are the appropriate shear moduli for the cut of quartz of interest.

We now transform the inhomogeneity from the differential equation (5.1) into the edge conditions (5.3). To this end we first let

$$f(x_1,x_3) = a_Bx_B\left(2h_o - h_1 - \frac{x_Ax_A}{2R}\right) \quad (5.5)$$

which is then expanded in a complex Fourier series thus

$$f(x_1,x_3) = \frac{1}{4}\sum_n\sum_m C_{nm}e^{i\alpha_n x_1}e^{i\beta_m x_3}, \quad n,m = 0, \pm1, \pm2\ldots, \quad (5.6)$$

where $\alpha_n = n\pi/a$ and $\beta_m = m\pi/b$. From the periodicity of the functions, we obtain

$$C_{nm} = \frac{1}{ab}\int_{-a}^{a}\int_{-b}^{b} f(X_1,X_3)\cdot \cdot e^{-i\alpha_n X_1} e^{-i\beta_m X_3}\, dX_1 dX_3 , \qquad (5.7)$$

where the expressions for the C_{mn} are too lengthy to present here[8]. We now write $w_2^{(0)}$ in the form

$$w_2^{(0)} = w_2^{*} + \hat{w}_2, \qquad (5.8)$$

where

$$w_2^{*} = \sum_n \sum_m A_{nm} e^{i\alpha_n X_1} e^{i\beta_m X_3} . \qquad (5.9)$$

The substitution of (5.7) with (5.8) into (5.1) [Eqs.(5.1) and (5.2) become identical when (5.6) is substituted into (5.1) and (5.2)] yields

$$A_{nm} = (3\rho/8h_1^3 R)C_{nm}/[\gamma_{11}\alpha_n^4 + \gamma_{33}\beta_m^4 + (2\gamma_{13} + 4\gamma_{55})\alpha_n^2\beta_m^2 + 4\gamma_{15}\alpha_n^3\beta_m + 4\gamma_{35}\alpha_n\beta_m^3], \qquad (5.10)$$

$$\gamma_{ABCD}\hat{w}_{2,CDAB} = 0 . \qquad (5.11)$$

The further substitution of (5.8) with (5.9) into (5.3) yields

$$\hat{w}_2 = -w_2^{*} = -\sum_m G_m e^{i\beta_m X_3} \text{ at } X_1 = \pm a ,$$

$$\hat{w}_2 = -B_o - \sum_n B_n e^{i\alpha_n X_1} \text{ at } X_3 = \pm b,$$

$$\hat{K}_{11} \mp k_1\hat{w}_{2,1} = -\sum_n G_m^{\mp} e^{i\beta_m X_3} \text{ at } X_1 = \pm a ,$$

$$\hat{K}_{33} \mp k_3\hat{w}_{2,3} = -B_o^{\mp} - \sum_n B_n^{\mp} e^{i\alpha_n X_1} \text{ at } X_3 = \pm b, \qquad (5.12)$$

where the constants B_o and $B_o^{\mp}$ arise from the w_2^{*} solution and are too lengthy to present here[8], as are the coefficients G_m, B_n, $G_m^{\mp}$ and $B_n^{\mp}$.

As a solution of (5.11) with (5.12), we write

$$\hat{w}_2 = w_2^{P} + w_2^{S} + \hat{w}_2 , \qquad (5.13)$$

in which all terms separately satisfy the homogeneous differential equation (5.11), w_2^P satisfies the constant terms and w_2^S the series terms in the edge conditions at $X_3 = \pm b$. The expression for w_2^P takes the form

$$w_2^{P} = -B_o + \frac{1}{b^2}(X_3^2 - b^2)(C_5 b^2 - C_2 X_3), \qquad (5.14)$$

where

$$C_2 = -\frac{1}{4}\frac{(B_o^{-} - B_o^{+})b}{2h_1^3\gamma_{33} + bk_3}, \quad C_5 = \frac{3}{4}\frac{(B_o^{+} + B_o^{-})}{2h_1^3\gamma_{33} + 3bk_3} . \qquad (5.15)$$

We take w_2^S in the form

$$w_2^{S} = \sum_n E_n e^{i(\alpha_n X_1 + \zeta_n X_3)} , \qquad (5.16)$$

the substitution of which in (5.11) along with the relation $\zeta_n = \kappa\alpha_n$ yields

$$\gamma_{33}\kappa^4 + 4\gamma_{35}\kappa^3 + (2\gamma_{13} + 4\gamma_{55})\kappa^2 + 4\gamma_{15}\kappa + \gamma_{11} = 0 , \qquad (5.17)$$

which is independent of n. Equation (5.17) yields four independent roots $\kappa^{(r)}$ (r = 1,2,3,4) independent of n. Hence, the sum in (5.16) can be written in the form

$$w_2^{S} = \sum_n \sum_{r=1}^{4} E_n^{(r)} e^{i\alpha_n\kappa^{(r)}X_3} e^{i\alpha_n X_1} . \qquad (5.18)$$

For each n the satisfaction of each exponential term in $(5.12)_{2,4}$ enables the determination of the $E_n^{(r)}$ in terms of the known inhomogeneities[8].

In order to satisfy the remaining inhomogeneous edge conditions variationally we substitute (5.13) into (5.8) and write

$$w_2^{(0)} = w_2^{*} + w_2^{P} + w_2^{S} + \hat{w}_2 , \qquad (5.19)$$

as the approximate solution function. We now take $\hat{w}_2$ in the form

$$\hat{w}_2 = \sum_m D_m e^{i(\nu_m X_1 + \eta_m X_3)} , \qquad (5.20)$$

where $\nu_m = m\pi/2\sigma a$ and σ is chosen to be an irrational number (here π) in order that neither $\cos \nu_m a$ or $\sin \nu_m a$ vanish. Let $\eta_m = \lambda\nu_m$ and substitute from (5.13) with (5.20) into (5.11) to obtain[13]

$$\gamma_{33}\lambda^4 + 4\gamma_{35}\lambda^3 + (2\gamma_{13} + 4\gamma_{55})\lambda^2 + 4\gamma_{15}\lambda + \gamma_{11} = 0 , \qquad (5.21)$$

which is independent of m. Equation (5.21) yields four independent roots $\lambda^{(p)}$ (p = 1,2,3,4) independent of m. Hence, the sum in (5.20) can be written in the form

$$\hat{w}_2 = \sum_m \sum_{p=1}^{4} D_m^{(p)} e^{i\nu_m X_1} e^{i\eta_m^{(p)} X_3} , \qquad (5.22)$$

where the $D_m^{(p)}$ are amplitude coefficients still to be determined. By satisfying the homogeneous edge conditions $\hat{w}_2 = 0$ at $X_3 = -b$ and $\hat{K}_{33} \mp k_3 w_{2,3} = 0$ at $X_3 = \pm b$ for each m, we obtain

$$D_m^{(p)} = g_m^{(p)} D_m^{(1)}, \tag{5.23}$$

where the $g_m^{(p)}$ are obtained from the resulting linear algebra[8]. This reduces the number of unknown coefficients from 4 to 1 for each m. The substitution of (5.23) in (5.22) enables us to write

$$\hat{w}_2 = \sum_m D_m^{(1)} e^{i\nu_m X_1} \sum_{p=1}^{4} g_m^{(p)} e^{i\eta_m^{(p)} X_3}, \tag{5.24}$$

as the approximate solution function.

Since the solution function (5.24) with (5.19), (5.9), (5.14) and (5.18) satisfies the differential equations (5.1) and (5.2) and the edge conditions are on $K_{(n)(n)}^{(1)}$ and $w_2^{(0)}$, all that remains of the variational principle given in Eq.(44) of Ref.2, in which all conditions appear as natural conditions, is

$$\oint_{C^N} (K_{(n)(n)}^{(1)} \mp k_{(n)} w_{2,n}^{(0)}) \delta \frac{\partial w_2^{(0)}}{\partial n} \, dS + \oint_{C^C} [-(\delta M_{NS,S} + N_B \delta K_{AB,A}^{(1)})(\bar{w}_2^{(0)} - w_2^{(0)})] \, dS = 0. \tag{5.25}$$

where

$$M_{NS} = N_D K_{DC}^{(1)} \mathscr{N}_C, \tag{5.26}$$

and N_D and $\mathscr{N}_C$ denote unit vectors, respectively, outwardly directed normal and tangential to C in the counterclockwise direction, as shown in Fig.2 of Ref.2. The bars over quantities in (5.25) denote that they are prescribed. The variation $\delta M_{NS,S}$ may readily be evaluated from (5.26) and $\hat{w}_2^{(0)} = 0$ on C^C. Substituting from (5.19) with (5.9), (5.14), (5.18), (5.24) and the homogeneous edge conditions mentioned above, which have been satisfied exactly by each term in the solution sum in (5.24), into (5.25) and employing (5.26), we obtain

$$\int_{-b}^{b} (K_1^{(1)} \mp k_1 w_{2,1}^{(0)}) \delta \hat{w}_{2,1} \, dX_3 \Big|_{X_1=-a}^{X_1=a} + \int_{-b}^{b} w_2^{(0)} \delta(\hat{K}_{1,1}^{(1)} + 2\hat{K}_{5,3}^{(1)}) dX_3 \Big|_{X_1=-a}^{X_1=a} + \int_{-a}^{a} \hat{w}_2^{(0)} \delta(\hat{K}_{3,3}^{(1)} + 2\hat{K}_{5,1}^{(1)}) dX_1 \Big|_{X_3=b} = 0, \tag{5.27}$$

where the notation $\Big|_{X_1=-a}^{X_1=a}$ means the quantity evaluated at a minus the quantity evaluated at -a and the $\delta\hat{K}_{AB}$ are obtained from (4.3) and $(4.6)_3$ with $\hat{w}_2^{(0)}$ instead of $w_2^{(0)}$ since the variation of prescribed quantities vanishes. Substituting from (5.19) with (5.9), (5.14), (5.18) and (5.24) into (5.27 and performing the integrations, we obtain

$$\sum_{n=-N}^{N} \sum_{m=-N}^{N} a_{mn} D_m^{(1)} \delta D_n^{(1)} + \sum_{n=-N}^{N} b_n \delta D_n^{(1)} = 0, \quad m,n = \pm 1, \pm 2, \ldots \pm N. \tag{5.28}$$

The expressions for a_{mn} and b_n are too lengthy to present here[8]. Since the variations $\delta D_n^{(1)}$ are arbitrary, we obtain

$$\sum_{m=-N}^{N} a_{mn} D_m^{(1)} = -b_n, \quad n = \pm 1, \pm 2, \ldots \pm N, \tag{5.29}$$

which constitute 2N inhomogeneous linear algebraic equations in the 2N unknowns $D_m^{(1)}$, the inversion of which gives the approximate solution.

6. The Equivalent Trapped Energy Mode Shape

A schematic diagram of the plano-convex resonator along with the associated coordinate system is shown at the bottom of Fig.1. It has been shown that the eigensolution for the dominant displacement of the harmonic modes in a contoured quartz resonator can be written in the form[14]

$$u_{1n} = \sin \frac{n\pi X_2}{2h} u_{noo} e^{i\omega_{noo} t}, \quad n = 1,3,5, \ldots, \tag{6.1}$$

where

$$u_{noo} = e^{-\alpha_n (\hat{X}_1^2/2) - \beta_n (\hat{X}_3^2/2)}, \tag{6.2}$$

in which

$$\alpha_n^2 = \frac{n^2 \pi^2 \hat{c}^{(1)}}{8Rh_o^3 M'_n}, \quad \beta_n^2 = \frac{n^2 \pi^2 \hat{c}^{(1)}}{8Rh_o^3 P'_n}. \tag{6.3}$$

and M'_n and P'_n are involved lengthy expressions that are defined in Sec.II of Ref.14 along with $\hat{X}_1$ and $\hat{X}_3$, which are orthogonal directions in the plane of the plate for which the scalar differential equations for the nth anharmonic family of modes does not contain mixed derivatives and R denotes the radius of the spherical contour. From Eq.(97) of Ref.14 in this work we have

$$\hat{X}_A = R_{AB} X_B, \tag{6.4}$$

where, from $Eq.(98)_1$ of Ref.14

$$R_{AB} = \begin{Vmatrix} \cos\hat{\beta} & -\sin\hat{\beta}_n \\ \sin\hat{\beta}_n & \cos\hat{\beta}_n \end{Vmatrix}, \qquad (6.5)$$

and $\hat{\beta}$ is defined in Eq.$(98)_2$ of Ref.14. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \times \left(\sqrt{\frac{M'_n}{\hat{c}^{(1)}}} + \sqrt{\frac{P'_n}{\hat{c}^{(1)}}}\right)\right], \qquad (6.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.14.

In addition to the thickness eigendisplacement u_{1n} given in (6.1) there are thickness eigendisplacements u_{2n} and u_{3n}, which are an order of magnitude smaller than u_{1n} but are required in this work because as shown in earlier work[3] the important biasing shearing deformation is an order of magnitude smaller than the biasing flexural deformation. However, since the u_2 and u_3 displacement components accompanying the larger u_1 component are known only for the flat plate, we fit the Gaussian mode shape given in (6.1) and (6.2) for the contoured resonator to a trapped energy mode in a flat plate in accordance with Sec.V of Ref.3, wherein the wavenumbers $\hat{\bar{\xi}}$, $\hat{\bar{\nu}}$, $\hat{\xi}^{SR}$ and $\hat{\nu}^{TM}$ are obtained from quantities contained in the Gaussian mode shape. As shown in Sec.V of Ref.3, when this has been done it is found from Eqs.(65) of Ref.14 that to the order of interest the $u_{\hat{n}}^{SR}$ displacement field for the SR region may be written

$$u_1^{SR} = A_+^{(1)SR}\left(1 + \frac{(c_{16}\xi^{SR} - c_{56}i\bar{\nu})}{\bar{c}^{(1)}}X_2\right) \cdot \cdot \sin\frac{n\pi X_2}{2h}\, e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\hat{\bar{\nu}}\hat{X}_3},$$

$$u_2^{SR} = \left(\frac{(r_2\xi^{SR} - r_4 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR}\cos\frac{n\pi}{2h}X_2 + iC_+^{(2)SR}\cos\kappa_2\frac{n\pi}{2h}X_2\right) e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\hat{\bar{\nu}}\hat{X}_3},$$

$$u_3^{SR} = \left(\frac{(r_5\xi^{SR} - r_3 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR}\cos\frac{n\pi}{2h}X_2 + iE_+^{(3)SR}\cos\kappa_3\frac{n\pi}{2h}X_2\right) e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\hat{\bar{\nu}}\hat{X}_3}, \qquad (6.7)$$

where

$$\xi^{SR} = R_{11}\hat{\xi}^{SR} - R_{31}i\hat{\bar{\nu}},\quad i\bar{\nu} = R_{33}i\hat{\bar{\nu}} - R_{13}\hat{\xi}^{SR}, \qquad (6.8)$$

and from Eqs.(72) of Ref.14, we know $C_+^{(2)SR}$ and $E_+^{(3)SR}$ in terms of $A_+^{(1)SR}$ and as in Ref.1 we do not write the relations for the corresponding displacement fields of the other regions. We note that $A_+^{(1)SR}$ here corresponds to B_S in Sec.V of Ref.5.

7. Calculated Results

From Sec.6 we now know g_{γ}^{μ} and from Secs.4 and 5 we know $\hat{c}_{L\gamma M\alpha}$ for in-plane acceleration of a plano-convex resonator resulting from the asymmetry caused by the contour being on one side only. Hence, we can now evaluate H_{μ} in Eq.(2.9). Such calculations have been performed using the known values of the second order[15] and third order[16] elastic constants of quartz. The calculated components of the in-plane acceleration sensitivity resulting from the thickness asymmetry of a plano-convex resonator for an SC-cut square plate are plotted in Fig.4 as a function of the radius of the contour for the first three harmonics. It can be seen from the figure that the in-plane acceleration sensitivity increases with decreasing radius of curvature. Since the practical range of R is between 15 and 20 cm, the increase in the in-plane acceleration sensitivity of a plano-convex resonator over that of an equivalent symmetric biconvex resonator is about a few parts in 10^{12} per g. Figure 5 shows similar calculations for a rectangular SC-cut plate with the practical aspect ratio a/b = 2 at the practical orientation of the rectangle[17-19], i.e., $\beta = -50^{\circ}$. The results are not substantially different from those shown in Fig.3.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No.DAAL03-88-K-0065 and the National Science Foundation under Grant No.MSS-8908188.

References

1. H.F. Tiersten and Y.S. Zhou, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," Proceedings of the 44th Annual Symposium on Frequency Control, U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No.90CH2818-3, 461 (1990)
2. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," J. Appl. Phys., 64, 4334 (1988).

3. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," J. Appl. Phys., 67, 60 (1990).

4. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).

5. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).

6. We have employed the convention that A,B,C,D take the values 1 and 3 and skip 2.

7. H.F. Tiersten and D.V. Shick, "On the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," J. Appl. Phys., 67, 2554 (1990).

8. For more detail see H.F. Tiersten and Y.S. Zhou, "The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Resulting from its Thickness Asymmetry," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.

9. R.D. Mindlin, "An Introduction to the Mathematical Theory of the Vibration of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, New Jersey (1955). Signal Corps Contract DA-36-03956-56772.

10. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).

11. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.13.

12. H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Doubly-Rotated Quartz Thickness-Mode Resonators," J. Appl. Phys., 50, 8038 (1979).

13. The four $\varkappa$ and four λ are, of course, identical. Nevertheless, the solution w_2^S and $\hat{w}_2$ are independent.

14. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).

15. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

16. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

17. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators with the Mode Shape Displaced with Respect to Rectangular Supports," J. Appl. Phys., 69, 2862 (1991).

18. H.F. Tiersten and Y.S. Zhou, "On the Influence of a Fabrication Imperfection on the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," to be published in the Proceedings of the 5th European Frequency and Time Forum, EFTF 91, Besancon, France (1991).

19. Although the aspect ratio for which the influence of offset for the case of normal acceleration exhibited a strong minimum for a particular orientation of support rectangle was denoted a/b = 0.5 in Refs.1 and 17, the same aspect ratio is denoted a/b = 2 in this work and that of Ref.18. This occurs because the direction in which X_1 points is always taken in the long direction and when the wavenumbers ν_m are taken in the long (X_1) direction it increases the rate of convergence of the static solution. With this convention the a/b ratio is never less than 1. For the flexural cases treated in Refs.1 and 17 the ν_m were taken in the X_3-direction rather than the X_1-direction. On account of this the orientation β of the support rectangle in Refs.1 and 17 differs from the orientation β in this work by $\pi/2$.

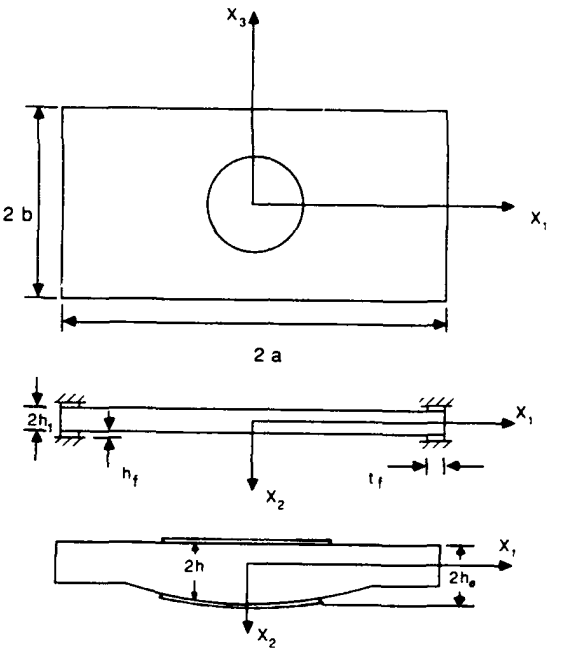


Figure 1 Plan View and Cross-Section of Plano-Convex Resonator

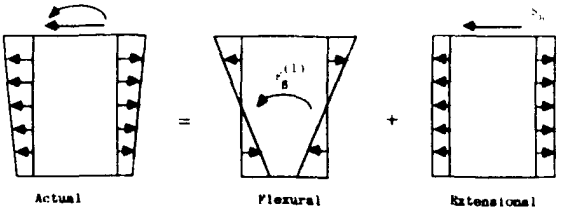


Figure 2 Decomposition into Flexural and Extensional Biasing States

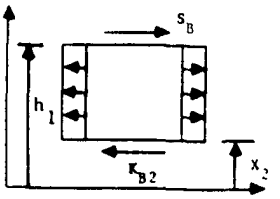


Figure 3 Free Body Diagram for Evaluation of Shearing Stresses Accompanying Extension

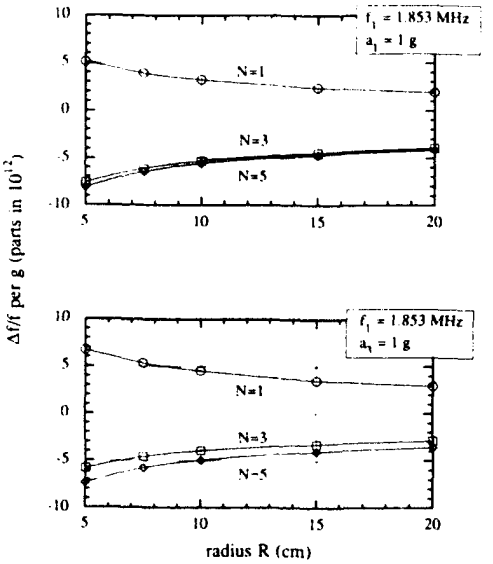


Figure 4 In-Plane Acceleration Sensitivity Resulting from Thickness Asymmetry of a Plano-Convex Resonator Versus Radius of Contour for SC-Cut Square Plate

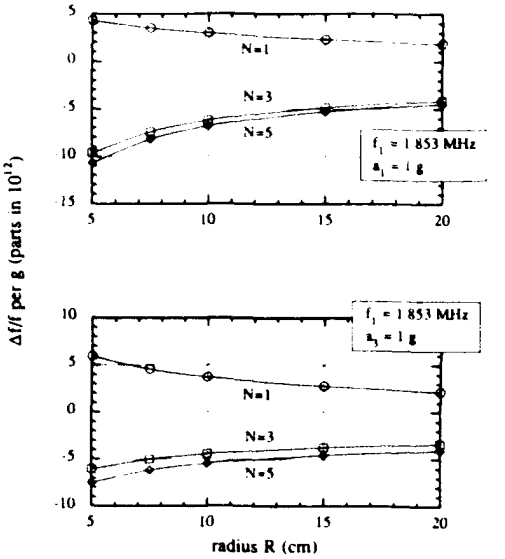


Figure 5 In-Plane Acceleration Sensitivity Resulting from Thickness Asymmetry of a Plano-Convex Resonator Versus Radius of Contour for Rectangular SC-Cut Plate

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN ANALYSIS OF THE NORMAL ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS STIFFENED BY IDENTICAL TOP AND BOTTOM QUARTZ COVER PLATES SUPPORTED BY CLIPS

Y.S. Zhou and H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Abstract

Recent work has shown that both the normal and resultant in-plane acceleration sensitivities of contoured quartz resonators vanish for a perfectly symmetric resonator and support system. It has also been shown that any loss of symmetry in the combined resonator plus support configuration results in a significant degradation in the acceleration sensitivity. Since it is essentially impossible to construct a perfectly symmetric resonator plus support configuration in practice, a stiffened structure is considered, which reduces the biasing deformation in the active region and provides isolation from the unavoidable variations in the details of the mounting devices. The structure consists of the active biconvex quartz resonator attached to identical top and bottom quartz cover plates by means of small sidewalls around the periphery. The mounting clips are attached to the top and bottom cover plates without touching the active plate. An analysis of the normal acceleration sensitivity of this configuration is performed. The variational principle for anisotropic static flexure in which all conditions arise as natural conditions is extended to include the influence of the cover plates, the active region and the clips. The biasing state is determined by means of our variational approximation procedure using the extended form of the variational principle. The calculated flexural biasing states are employed in the existing perturbation equation along with the mode shapes of the contoured resonator to calculate the normal acceleration sensitivities. Results are presented for a mispositioning of either a clip or the mode shape for a range of thickness of cover plates relative to the active plate.

1. Introduction

It has recently been shown[1,2] that the resultant acceleration sensitivity of contoured quartz resonators vanishes when the resonator and support system are perfectly symmetric with respect to each other. This was shown analytically and is a consequence of the symmetry of the biasing state and the mode shape. It was further shown[1,3] that for a small loss of symmetry in the combined resonator plus support system the acceleration sensitivity increases linearly with the loss. Since it is impossible to construct a perfectly symmetric configuration in practice, it is desirable to have a support configuration in which the linear increase in sensitivity with loss of symmetry is as low as possible. Consequently, in this work a stiffened structure is considered in which the active biconvex quartz resonator is sandwiched between identical top and bottom quartz cover plates to which it is attached by means of small sidewalls around the periphery. The mounting clips are attached to the cover plates without touching the active plate.

In order to perform the analysis for the biasing state for this stiffened structure, the variational principle for anisotropic static flexure in which all conditions arise as natural conditions[4] is extended to include the influence of the cover plates and the clips. The biasing deformation fields are calculated by means of a variational approximation procedure[4,5] using the extended form of the variational principle. The calculated biasing states are employed in an existing perturbation equation[6] along with the equivalent trapped energy mode shapes[7] for the contoured resonator to calculate the normal acceleration sensitivities arising from a loss of symmetry. The calculations reveal that increasing the thickness of the cover plates relative to the active plate reduces the normal acceleration sensitivity arising from either a misplacement of a clip or a mispositioning of the center of the mode shape. The influence of symmetric holes in the cover plates is included in the calculations. Results are presented for a range of thickness of cover plates relative to the active plate and hole size.

2. Preliminary Considerations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency[6] may be written in the form given in Eq.(2.1) of the previous paper (Ref.8), which we reproduce here for completeness

$$\Delta_\mu = H_\mu/2\omega_\mu, \quad \omega = \omega_\mu - \Delta_\mu, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively. It has further been shown that H_μ may be written in the form

CH2965-2/91/0000-298 $1.00 © 1991 IEEE

given in Eq. (2.9) of Ref.8, which we also reproduce here for completeness

$$H_{\mu} = - \int_V \hat{c}_{L\nu M\alpha} g^{\mu}_{\nu,L} g^{\mu}_{\alpha,M} \, dV, \qquad (2.2)$$

where $\hat{c}_{L\nu M\alpha}$ is defined in Eq.(2.4) of Ref.8 and all other quantities are defined in Sec.2 of Ref.8. Since g^{μ}_{α} denotes the normalized mode shape in the contoured resonator and from Eq.(2.4) of Ref.8 $\hat{c}_{L\nu M\alpha}$ depends on the biasing state, H_{μ} can be calculated when the mode shape in the resonator and biasing state are known.

The equation of anisotropic static flexure with X_2 normal to the major surfaces can be written in the form given in Eq.(4.1) of Ref.8, which we reproduce here for completeness

$$K^{(1)}_{AB,AB} + F^{(1)}_{B,B} + F^{(0)}_2 = 0, \qquad (2.3)$$

where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2 and

$$K^{(1)}_{AB} = \int_{-h}^{h} X_2 K_{AB} \, dX_2,$$

$$F^{(1)}_B = h[K_{2B}(h) + K_{2B}(-h)], \qquad (2.4)$$

and in this work

$$F^{(1)}_B = 0, \quad F^{(0)}_2 = K_{22}(h) - K_{22}(-h) - 2h\rho a_2 = -2h\rho a_2, \qquad (2.5)$$

since K_{2K} vanishes on the major surfaces and where a_2 is the acceleration in the X_2-direction. Equations (4.3) - (4.9) of Ref.8 apply here without change and for brevity we do not bother to reproduce them here. However, Eq.(4.10) of Ref.8 does not exist in this work because we are considering normal accelerations only here. As a consequence, Eqs.(4.11) and (4.16) of Ref.8 change their form in this work. Hence, although Eqs.(4.12) - (4.15) of Ref.8 remain the same, we repeat some here for completeness and because some of the in-between wording is changed. To this end we first note that since from Eqs.(3.31) of Ref.7 we have $E^{(0)}_{KL} = 0$, when the plate deflection $w^{(0)}_2$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = {}_o\hat{E}_{KL} + X_2 E^{(1)}_{KL} + X_2^2 \, {}_2\hat{E}_{KL}. \qquad (2.6)$$

Since, as in Ref.8, we also need the three-dimensional rotations Ω_{KL} given in (4.12) of Ref.8, we note that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K^{(1)}_{2L}$ and the recursive inversion in (4.9) of Ref.8 satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L}, \qquad (2.7)$$

where

$$\Omega_{KL} = \sum_{n=0}^{2} X_2^n \, \Omega^{(n)}_{KL}, \qquad (2.8)$$

and the $\Omega^{(n)}_{KL}$ are the plate rotations of nth order. As in the end of Sec.III of Ref.7 [after Eq.(3.32)] from the flexural solution without recursive strains we obtain

$$\Omega^{(0)}_{13} = \text{arbitrary constant} = 0,$$

$$\Omega^{(1)}_{13} = 0, \quad \Omega^{(0)}_{2A} = - w^{(0)}_{2,A}. \qquad (2.9)$$

Including recursive strains for n = 1 and 2 and equating like powers of X_2 in the same way we obtain[5]

$$\Omega^{(1)}_{13} = {}_o\hat{E}_{23,1} - {}_o\hat{E}_{21,3}, \quad \Omega^{(2)}_{13} = \frac{1}{2}(E^{(1)}_{23,1} - E^{(1)}_{21,3}),$$

$$\Omega^{(1)}_{2A} = E^{(1)}_{2A} - {}_o\hat{E}_{22,A}, \quad \Omega^{(2)}_{2A} = {}_2\hat{E}_{2A} - \frac{1}{2} E^{(1)}_{22,A}. \qquad (2.10)$$

The substitution of (2.9) and (2.10) in (2.8) yields Ω_{KL}, which with (2.6) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \qquad (2.11)$$

gives us the desired three-dimensional displacement gradients when the plate deflection $w^{(0)}_2$ has been found.

3. Unconstrained Variational Principle for Static-Flexure of Box Structure

In this section we present the unconstrained variational principle for static flexure of the box structure, a plan view and cross-section of which is shown in Fig.1. This principle is required for the determination of the flexural biasing state for the configuration shown in Fig.1 using the approximation procedure we employ. Before proceeding with the variational principle it is advisable to discuss the configuration in some detail. As already noted, the structure consists of the active biconvex quartz resonator, which is attached to identical top and bottom quartz cover plates by means of small sidewalls around the periphery. The cover plates have identical holes in them in order to provide access to the active resonator. The holes are identical as are the cover plates in order to achieve the desired symmetry. The mounting clips are attached to the cover plates without touching the active plate.

For the purpose of obtaining the variational principle we do not consider the rectangle shown in the plan view of Fig.1, but rather the asymmetric plan view shown in Fig.2, in which the outside edge of the cover plates consists of two smooth curves with outward unit normal N_A, which

intersect at ${}_oC_1$ and ${}_oC_2$. In addition, the cover plates have an inner edge of an opening consisting of two smooth curves, respectively, similar to the outer curves, which intersect at ${}_iC_1$ and ${}_iC_2$. Some portions of the outside edge of the cover plates are supported and others are subject to prescribed loadings, which may vanish in any part. Similarly for the inside edge of the openings in the cover plates in the derivation of the variational principle, even though in the problem to be treated the entire inside edge in each cover plate is free. The portions of any edge of a cover plate that are supported have $\bar{w}_2^{(0)}$ and $\partial\bar{w}_2^{(0)}/\partial n$ prescribed and are denoted c^C, where n denotes the normal curvilinear coordinate to the curve c^C and the bars denote that the quantities are prescribed. The portions of the edge of a cover plate which are subject to prescribed loadings, which consist of bending moments $\bar{m}$, twisting moments $\bar{t}$ and vertical shearing forces $\bar{v}$, are denoted c^N. Thus c^N and c^C denote the portions of an edge of a cover plate on which natural- and constraint-type conditions, respectively, are prescribed and we note that c^N and c^C can refer to different portions of the surface for different terms in the edge conditions on the cover plates depending on the actual condition at a point. Since we have outer and inner edges on the cover plates, we use c_o^N, c_o^C and c_i^N, c_i^C to refer to the outer and inner edges, respectively. The corners ${}_oC_1$ and ${}_iC_1$ are subject to prescribed vertical forces $\bar{T}_o = \bar{t}_o^+ - \bar{t}_o^-$ and $\bar{T}_i = \bar{t}_i^+ - \bar{t}_i^-$, respectively, where the ${}_oC_2$ and ${}_iC_2$ are presumed to be subject to prescribed corner displacements $\bar{w}_2^{(0)}{}_oC_2$ and $\bar{w}_2^{(0)}{}_iC_2$.

The active plate is attached to the identical top and bottom cover plates by means of short, stubby sidewalls that run around the periphery as shown in the cross-section in Fig.1. For flexure caused by normal loading the short, stubby sidewalls may be regarded as rigid without any loss in accuracy. This means that the sidewalls serve to keep the deflections and the normal slopes of the cover plates and the active plate the same at the sidewalls, i.e., around the outer periphery. In accordance with nomenclature we have used in previous work this makes each sidewall a surface of discontinuity, which in this work is denoted c^d for the curve of discontinuity that is taken to represent the sidewalls that run around the periphery.

As usual, in order that all variations may be regarded as independent when constraints exist, we introduce each constraint as a zero times a Lagrange multiplier[9,10] in the variational principle. Then, in accordance with the above discussion, the unconstrained variational principle for static flexure of the box structure may be written in the form

$$\delta\int_S (-U + F_2^{(0)}w - F_B^{(1)}w_{,B})dS + 2\delta\int_{\hat{S}} (-\hat{U} + \hat{F}_2^{(0)}\hat{w} - \hat{F}_B^{(1)}\hat{w}_{,B})dS + \sum_{r=o}^{i}\Big(2\delta\oint_{c_r^N}\Big[\Big(\bar{v} + \frac{\partial\bar{t}}{\partial s}\Big)\hat{w} - \bar{m}\frac{\partial\hat{w}}{\partial n}\Big]ds + 2\delta\oint_{c_r^C}[\lambda^r(\bar{w} - \hat{w}) + \mu^r(\bar{w}_{,n} - \hat{w}_{,n})]ds\Big) + 2\delta\oint_{c^d}[\lambda^d(\hat{w} - w) + \mu^d(\hat{w}_{,n} - w_{,n})]ds + 2\delta(\bar{T}\hat{w}\,{}_oC_1) + 2\delta(\bar{T}\hat{w}\,{}_iC_1) + 2\lambda\,{}_oC_2\,\delta\hat{w}\,{}_oC_2 + 2\lambda\,{}_iC_2\,\delta\hat{w}\,{}_iC_2 + 2\lambda^d\,{}_oC_1\,\delta(\hat{w} - w) + 2\lambda^d\,{}_oC_2\,\delta(\hat{w} - w) = 0, \qquad (3.1)$$

where S denotes the area of the active plate, $\hat{S}$ denotes the area of the cover plates

$$U = -\frac{1}{2}K_{AB}w_{,AB},\quad K_{AB} = -\frac{2}{3}h^3\gamma_{ABCD}w_{,CD},$$
$$\hat{U} = -\frac{1}{2}\hat{K}_{AB}\hat{w}_{,AB},\quad \hat{K}_{AB} = -\frac{2}{3}h^3\gamma_{ABCD}\hat{w}_{,CD}, \qquad (3.2)$$

λ^o, μ^o, λ^i, μ^i, λ^d, μ^d, $\lambda\,{}_oC_2$, $\lambda\,{}_iC_2$, $\lambda^d\,{}_oC_1$ and $\lambda^d\,{}_oC_2$ are Lagrange undetermined multipliers and for notational convenience we have taken the liberty of suppressing the superscripts (0), (1) and subscripts 2. The factor 2 appears in certain terms in Eq.(3.1) because there are two identical cover plates and a caret over a quantity indicates that it is associated with the cover plates. Taking the variations in (3.1), integrating by parts and using the surface divergence theorem twice[11], employing the chain rule of differentiation with respect to the orthogonal curvilinear coordinates n,s, integrating by parts along C_o and C_i to the respective intersections ${}_oC_1$, ${}_iC_1$, ${}_oC_2$ and ${}_iC_2$, and employing the relations

$$\partial n/\partial X_B = N_B,\quad \partial s/\partial X_C = \ell_C, \qquad (3.3)$$

where N_B and ℓ_C denote outwardly directed unit vectors normal and tangential to either C_o or C_i in the respective counterclockwise or clockwise direction, we obtain

$$\int_S [K_{AB,BA} + F_2^{(0)} + F_{B,B}^{(1)}]\delta w\,dS + 2\int_{\hat{S}}[\hat{K}_{AB,BA} + \hat{F}_2^{(0)} + \hat{F}_{B,B}^{(1)}]\delta\hat{w}\,dS + \oint_{c_o^N}\Big[2\Big(-\hat{M}_{Ns,s} - \hat{K}_{AB,B}N_A - \hat{F}_B^{(1)}N_B + \lambda^d + \bar{v} + \frac{\partial\bar{t}}{\partial s}\Big)\delta\hat{w} + 2(N_B\hat{K}_{AB}N_A +$$

$$+ \mu^d - \bar{m})\delta\hat{w}_{,n} + 2\delta\lambda^d(\hat{w} - w) + 2\delta\mu^d(\hat{w}_{,n} - w_{,n})\Big]ds$$

$$+ \oint_{C_o^C} 2[\delta\lambda^o(\bar{w} - \hat{w}) + \delta\mu^o(\bar{w}_{,n} - \hat{w}_{,n}) + \delta\lambda^d(\hat{w} - w) + \delta\mu^d(\hat{w}_{,n} - w_{,n})]ds + \oint_{c_i^N} \Big[2\Big(-\hat{M}_{Ns,s} - \hat{K}_{AB,B}N_A - \hat{F}_B^{(1)}N_B + \bar{v} + \frac{\partial\bar{t}}{\partial s}\Big)\delta\hat{w} + 2(N_B\hat{K}_{AB}N_A - \bar{m})\delta\hat{w}_{,n}\Big]ds$$

$$+ \oint_{c_i^C} 2[\delta\lambda^i(\bar{w} - \hat{w}) + \delta\mu^i(\bar{w}_{,n} - \hat{w}_{,n})]ds$$

$$+ 2(\hat{M}_{Ns}^+ - \hat{M}_{Ns}^- + \bar{T})_{i^C1}\delta\hat{w}^{i^C1} + 2(\hat{M}_{Ns}^+ - \hat{M}_{Ns}^- + \lambda^d{}^{o^C1} + \bar{T})_{o^C1}\delta\hat{w}^{o^C1} = 0, \quad (3.4)$$

where

$$M_{Ns} = N_D K_{DC}\epsilon_C \quad (3.5)$$

and for brevity[12] we have omitted the terms that serve to define the Lagrange multipliers

$$2\lambda^d = -M_{Ns,s} - K_{AB,B}N_A - F_B^{(1)}N_B, \quad 2\mu^d = N_B K_{AB}N_A$$

$$\lambda^o = -\hat{M}_{Ns,s} - \hat{K}_{AB,B}N_A - \hat{F}_B^{(1)}N_B + \lambda^d, \quad \mu^o = N_B\hat{K}_{AB}N_A + \mu^d$$

$$2\lambda^{d\,o^C1} = M_{Ns}^{+\,o^C1} - M_{Ns}^{-\,o^C1},$$

$$2\lambda^{d\,o^C2} = M_{Ns}^{+\,o^C2} - M_{Ns}^{-\,o^C2}$$

$$\lambda^{o^C2} = \hat{M}_{Ns}^{-\,o^C2} - \hat{M}_{Ns}^{+\,o^C2} - \lambda^{d\,o^C2},$$

$$\lambda^{i^C2} = \hat{M}_{Ns}^{-\,i^C2} - \hat{M}_{Ns}^{+\,i^C2}, \quad (3.6)$$

since all variations may be regarded as arbitrary. For the same reason, i.e., because all variations may be regarded as independent on account of the use of Lagrange multipliers with the constraint conditions, we obtain the flexural equation of equilibrium (2.3) for both the active plate and the cover plates from the surface terms in (3.4) and the material edge and corner conditions

$$\hat{M}_{Ns,s} + N_A\hat{K}_{AB,B} + N_B\hat{F}_B^{(1)} - \lambda^d = \bar{v} + \frac{\partial\bar{t}}{\partial s} \text{ on } c_o^N,$$

$$N_A\hat{K}_{AB}N_B + \mu^d = \bar{m} \text{ on } c_o^N, \quad N_A\hat{K}_{AB}N_B = \bar{m} \text{ on } c_i^N,$$

$$\hat{M}_{Ns,s} + N_A\hat{K}_{AB,B} + N_B\hat{F}_B^{(1)} = \bar{v} + \frac{\partial\bar{t}}{\partial s} \text{ on } c_i^N,$$

$$\hat{M}_{Ns}^- - \hat{M}_{Ns}^+ = \bar{T} \text{ on } {}_iC_1, \quad \hat{M}_{Ns}^- - \hat{M}_{Ns}^+ - \lambda^{d\,o^C1} = \bar{T} \text{ on } {}_oC_1, \quad (3.7)$$

along with the natural form of the constraint-type edge conditions

$$\hat{w} = w, \quad \partial\hat{w}/\partial n = \partial w/\partial n \text{ on } c_o^N + c_o^C,$$

$$\hat{w} = \bar{w}, \quad \partial\hat{w}/\partial n = \partial\bar{w}/\partial n \text{ on } c_o^C \text{ and } c_i^C, \quad (3.8)$$

from the remaining terms in (3.4).

Thus it is clear that the variational principle (3.1) with unconstrained variations yields the differential equation (2.3) for both the active plate and the cover plates, the natural edge and corner conditions (3.7) and the constraint type edge conditions (3.8). In addition, the Lagrange multipliers have been expressed in terms of derivatives of w and $\hat{w}$ in (3.6). Consequently, the variations may be obtained from (3.6) and substituted[12] in (3.4), which may then be used for obtaining an approximate solution for the flexural biasing state without any a priori conditions[13,14] on the approximating functions.

4. Flexure of Box Structure Supported by Clips

A plan view and cross-section of the box structure is shown in Fig.1 along with the coordinate system. The substitution of $(3.2)_{2,4}$ and (2.5) in (2.3) for both the active plate and the cover plates yields the equilibrium equation for flexure of the thin plates in the form

$$\frac{2}{3}h^3\gamma_{ABCD}w_{,CDAB} + 2h\rho a_2 = 0,$$

$$\frac{2}{3}\hat{h}^3\gamma_{ABCD}\hat{w}_{,CDAB} + 2\hat{h}\rho a_2 = 0. \quad (4.1)$$

Since the inner edges of the cover plates are free and the outer edges are free except at the clips, where they are simply supported and the active plate is attached to the cover plate by means of sidewalls around the periphery, the boundary conditions take the form

$$\hat{K}_{11}^{(1)} + \frac{1}{2}K_{11}^{(1)} = 0 \; |X_3| \le b, \; X_1 = \pm a,$$

$$\hat{K}_{11}^{(1)} = 0, \; |X_3| \le b_C, \; X_1 = \pm a_C,$$

$$\hat{K}_{33}^{(1)} + \frac{1}{2}K_{33}^{(1)} = 0 \; |X_1| \le a, X_3 = \pm b,$$

$$\hat{K}_{33}^{(1)} = 0, \; |X_1| \le a_C, X_3 = \pm b_C,$$

$$2\hat{K}_{13,3}^{(1)} + \hat{K}_{11,1}^{(1)} + \frac{1}{2}(2K_{13,3}^{(1)} + K_{11,1}^{(1)}) = 0 \quad \text{at } X_1 = \pm a, \; t \le |X_3| \le b,$$

$$2\hat{K}_{13,1}^{(1)} + \hat{K}_{33,3}^{(1)} + \frac{1}{2}(2K_{13,1}^{(1)} + K_{33,3}^{(1)}) = 0 \quad \text{at } X_3 = \pm b, \; t \le |X_1| \le a,$$

$$\hat{w} = 0 \text{ at } X_1 = \pm a, \; |X_3| \le t,$$

$$\hat{w} = 0 \text{ at } X_3 = \pm b, \; |X_1| \le t,$$

$$2\hat{K}_{13,3}^{(1)} + \hat{K}_{11,1}^{(1)} = 0 \text{ at } X_1 = \pm a_C, \; |X_3| \le b_C,$$

$$2\hat{K}^{(1)}_{13,1} + \hat{K}^{(1)}_{33,3} = 0 \text{ at } X_3 = \pm b_c, \quad |X_1| \le a_c,$$

$$w = \hat{w}, \; w_{,1} = \hat{w}_{,1} \text{ at } X_1 = \pm a, \; |X_3| \le b,$$

$$w = \hat{w}, \; w_{,3} = \hat{w}_{,3} \text{ at } X_3 = \pm b, \; |X_1| \le a, \qquad (4.2)$$

in which for simplicity we have written $(4.2)_{5\text{-}8}$ for the case of symmetric supports for which the acceleration sensitivity vanishes. However, in the actual problem treated we consider a clip to be mispositioned and modify $(4.2)_{5\text{-}8}$ to include the asymmetry.

Since as already indicated the problem defined in (4.1) and (4.2) cannot be solved exactly, an approximation procedure is employed. To this end we first transform the inhomogeneities from the differential equations (4.1) into the boundary conditions (4.2) by writing

$$w = w^* + w_S, \; \hat{w} = \hat{w}^* + \hat{w}_S, \qquad (4.3)$$

where w^* and $\hat{w}^*$ are polynomials up to fourth order in X_1 and X_3 with coefficients depending on the acceleration a_2. The polynomials w^* and $\hat{w}^*$ also satisfy certain of the edge conditions[12]. The substitution of (4.3) into (4.1) yields

$$\gamma_{ABCD} w_{S,CDAB} = 0, \; \gamma_{ABCD} \hat{w}_{S,CDAB} = 0. \qquad (4.4)$$

The further substitution of (4.3) with $(3.2)_{2,4}$ into (4.2) yields the transformed edge conditions, which for brevity we do not bother to write here[12].

As a solution of (4.4) we write

$$w_S = \sum_m D_m e^{i(\nu_m X_1 + \eta_m X_3)},$$

$$\hat{w}_S = \sum_m \hat{D}_m e^{i(\hat{\nu}_m X_1 + \hat{\eta}_m X_3)}, \qquad (4.5)$$

where $\nu_m = m\pi/2\sigma a$ and $\hat{\nu}_m = m\pi/2\sigma a$ and σ is chosen to be an irrational number (here π) in order that neither $\cos \nu_m b$ or $\sin \nu_m b$ vanish. The substitution of (4.5) into (4.4) yields

$$\gamma_{11}\nu_m^4 + \gamma_{33}\eta_m^4 + (2\gamma_{13} + 4\gamma_{55})\nu_m^2\eta_m^2 + 4\gamma_{15}\nu_m^3\eta_m + 4\gamma_{35}\nu_m\eta_m^3 = 0. \qquad (4.6)$$

Equation (4.6) yields four independent roots $\eta_m^{(p)} = \hat{\eta}_m^{(p)}$ $(p = 1,2,3,4)$, which are independent of m. Hence, the sums in (4.5) can be written in the form

$$w_S = \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} D_m^{(p)} e^{i\eta_m^{(p)} X_3},$$

$$\hat{w}_S = \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} \hat{D}_m^{(p)} e^{i\eta_m^{(p)} X_3}, \qquad (4.7)$$

where the $D_m^{(p)}$ and the $\hat{D}_m^{(p)}$ are amplitude coefficients still to be determined. By satisfying the homogeneous edge conditions $\hat{K}_{33} + \frac{1}{2} K_{33} = 0$ at $X_3 = \pm b$, $w_S - \hat{w}_S = 0$ at $X_3 = -b$ and $w_{S,3} - \hat{w}_{S,3} = 0$ at $X_3 = \pm b$ for each m, we obtain

$$D_m^{(p)} = g_{1m}^{(p)} D_m^{(1)} + g_{2m}^{(p)} D_m^{(2)} + g_{3m}^{(p)} \hat{D}_m^{(1)} = \sum_{\ell=1}^{3} g_{\ell m}^{(p)} \hat{D}_m^{(\ell)},$$

$$\hat{D}_m^{(p)} = \hat{g}_{1m}^{(p)} D_m^{(1)} + \hat{g}_{2m}^{(p)} D_m^{(2)} + \hat{g}_{3m}^{(p)} \hat{D}_m^{(1)} = \sum_{\ell=1}^{3} \hat{g}_{\ell m}^{(p)} \hat{D}_m^{(\ell)}, \qquad (4.8)$$

so that $\hat{D}_m^{(1)} = D_m^{(1)}$, $\hat{D}_m^{(2)} = D_m^{(2)}$, $\hat{D}_m^{(3)} = \hat{D}_m^{(1)}$. This reduces the number of unknown coefficients from 8 to 3 for each m. The substitution of (4.8) in (4.7) enables us to write

$$w_S = \sum_m e^{i\nu_m X_1} \sum_{\ell=1}^{3} \hat{D}_m^{(\ell)} \sum_{p=1}^{4} g_{\ell m}^{(p)} e^{i\eta_m^{(p)} X_3},$$

$$\hat{w}_S = \sum_m e^{i\nu_m X_1} \sum_{\ell=1}^{3} \hat{D}_m^{(\ell)} \sum_{p=1}^{4} \hat{g}_{\ell m}^{(p)} e^{i\eta_m^{(p)} X_3}, \qquad (4.9)$$

as the approximate solution functions.

Since the solution functions in (4.3) satisfy the differential equations (4.1), only the edge conditions remain in the variational principle (3.4) and those edge conditions that have been satisfied do not appear. Substituting from (4.3) and (4.9) and the homogeneous edge conditions mentioned above, which have been satisfied exactly by each term in the solution sums, into (3.4) and employing (3.5) and the fact that $\bar{t}$ and $\bar{v}$ vanish on the inside edges and on the outside edges of the cover plates, where there are no clips, we obtain

$$-\int_{-b}^{b} [(4\hat{K}_5 + 2K_5)_{,3} + (2\hat{K}_1 + K_1)_{,1}] \delta\hat{w}_S \Big|_{\substack{X_1=a \\ dX_3 \\ X_1=-a}}$$

$$-\int_{-a}^{a} [(4\hat{K}_5 + 2K_5)_{,1} + (2\hat{K}_3 + K_3)_{,3}]_S \, \delta\hat{w}_S \Big|_{\substack{X_3=b \\ dX_1 \\ X_3=-b}} +$$

$$+\int_{f-t}^{f+t}\hat{w}\delta[(4\hat{K}_5+2K_5)_{,3}+(2\hat{K}_1+K_1)_{,1}]_S\Big|_{\substack{X_1=a,\,f=f_1\\ dX_3\\ X_1=-a,\,f=f_3}}$$

$$+\int_{f-t}^{f+t}\hat{w}\delta[(4\hat{K}_5+2K_5)_{,1}+(2\hat{K}_3+K_3)_{,3}]_S\Big|_{\substack{X_3=b,\,f=f_2\\ dX_1\\ X_3=-b,\,f=f_4}}$$

$$+\int_{-b}^{b}(2\hat{K}_1+K_1)\delta\hat{w}_{S,1}\Big|_{\substack{X_1=a\\ dX_3\\ X_1=-a}}-\int_{-b}^{b}(\hat{w}-w)\delta(2K_{5,3}$$

$$+K_{1,1})_S\Big|_{\substack{X_1=a\\ dX_3\\ X_1=-a}}-\int_{-a}^{a}(\hat{w}-w)_S\delta(2K_{5,1}+K_{3,3})_S\Big|_{dX_1}^{X_3=b}$$

$$+\int_{-b}^{b}(\hat{w}_{,1}-w_{,1})_S\delta K_{1S}\Big|_{\substack{X_1=a\\ dX_3\\ X_1=-a}}+(4\hat{K}_5$$

$$+2K_5)\delta\hat{w}_S\Big|_{\substack{X=a,X_3=b\\ X_1=a,X_3=-b}}-(4\hat{K}_5+2K_5)\delta\hat{w}_S\Big|_{\substack{X_1=-a,X_3=b\\ X_1=-a,X_3=-b}}$$

$$+2\int_{-b_C}^{b_C}(2\hat{K}_{5,3}+\hat{K}_{1,1})\delta\hat{w}_S\Big|_{\substack{X_1=a_C\\ dX_3\\ X_1=-a_C}}+2\int_{-a_C}^{a_C}(2\hat{K}_{5,1}$$

$$+\hat{K}_3)_{,3})\delta\hat{w}_S\Big|_{\substack{X_3=b_C\\ dX_1\\ X_3=-b_C}}-2\int_{-b_C}^{b_C}\hat{K}_1\delta\hat{w}_{S,1}\Big|_{\substack{X_1=a_C\\ dX_3\\ X_1=-a_C}}$$

$$-2\int_{-a_C}^{a_C}\hat{K}_3\delta\hat{w}_{S,3}\Big|_{\substack{X_3=b_C\\ dX_1\\ X_3=-b_C}}-4\hat{K}_5\delta\hat{w}_S\Big|_{\substack{X_1=a_C,X_3=b_C\\ X_1=-a_C,X_3=-b_C}}$$

$$+4\hat{K}_5\delta\hat{w}_S\Big|_{\substack{X_1=-a,X_3=b_C\\ X_1=-a_C,X_3=-b_C}}=0,\qquad(4.10)$$

where the integrals with the slash are defined by

$$\int_{-b}^{b}=\int_{-b}^{f_1-t}+\int_{f_1+t}^{b}\quad\text{at } X_1=\pm a,$$

$$\int_{-a}^{a}=\int_{-a}^{f_2-t}+\int_{f_2+t}^{a}\quad\text{at } X_3=\pm b,\qquad(4.11)$$

in which f_i $(i=1,2,3,4)$ denotes the mispositioning of a clip as shown in Fig.1. The notation $\Big|_{X_1=-a}^{X_1=a}$ means the quantity evaluated at a minus the quantity evaluated at -a, the compressed notation for tensor indices has been introduced and the δK_{AB}, $\delta\hat{K}_{AB}$ are obtained from $(3.2)_{2,4}$ with w_S, $\hat{w}_S$ instead of w, $\hat{w}$ since the variation of prescribed quantities vanishes and the subscript S means evaluated with w_S, $\hat{w}_S$. Substituting from (4.3) with (4.9) into (4.10) and performing the integrations, we obtain

$$\sum_{n=-N}^{N}\sum_{k=1}^{3}\Big[\sum_{m=-N}^{N}\sum_{\ell=1}^{3}a_{mn}^{\ell k}\hat{D}_m^{(\ell)}+b_n^k\Big]\delta\hat{D}_n^{(k)}=0.\qquad(4.12)$$

The $a_{mn}^{\ell k}$ and b_n^k are very lengthy expressions, which provide no additional understanding and were left in the computer where they were generated. Since the $\delta\hat{D}_n^{(k)}$ are arbitrary, we obtain

$$\sum_{m=-N}^{N}\sum_{\ell=1}^{3}a_{mn}^{\ell k}\hat{D}_m^{(\ell)}=-b_n^k,$$

$$k=1\text{-}3,\ n=\pm1\ \pm2,\ \ldots,\ \pm N,\qquad(4.13)$$

which constitute 6N inhomogeneous linear algebraic equations in the 6N unknowns $\hat{D}_m^{(\ell)}$, the inversion of which gives the approximate solution.

5. The Mode Shape in the Biconvex Resonator

A schematic diagram of the biconvex resonator along with the associated coordinate system is shown in Fig.3. Although most of Sec.6 of Ref.8 appears here without change, there is one important change in an equation and Ref.8 is for a plano-convex resonator while this work is for a biconvex resonator. Consequently, we repeat some of the material of Sec.6 of Ref.8 here. It has been shown that the eigensolution for the dominant displacement of the harmonic modes in a contoured quartz resonator can be written in the form[15]

$$u_{1n}=\sin\frac{n\pi X_2}{2h}\,u_{noo}e^{i\omega_{noo}t},\quad n=1,3,5,\ldots,\qquad(5.1)$$

where

$$u_{noo}=e^{-\alpha_n(\hat{X}_1^2/2)-\beta_n(\hat{X}_3^2/2)},\qquad(5.2)$$

in which

$$\alpha_n^2=\frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3M'_n},\quad\beta_n^2=\frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3P'_n},\qquad(5.3)$$

and M'_n and P'_n are involved lengthy expressions that are defined in Sec.II of Ref.15 along with $\hat{X}_1$ and $\hat{X}_3$, which are orthogonal directions in the plane of the plate for which the scalar differential equation for the nth anharmonic family of modes does not contain mixed derivatives and 2R denotes the radii of the spherical contours. From Eq.(97) of Ref.15 in this work we have

$$\hat{X}_A=R_{AB}(X_B-d_B),\qquad(5.4)$$

where d_B represents the components of the displacement of the center of the mode shape from the symmetric position with respect to the support configuration and from Eq.$(98)_1$ of Ref.15

$$R_{AB} = \begin{Vmatrix} \cos \hat{\beta}_n & -\sin \hat{\beta}_n \\ \sin \hat{\beta}_n & \cos \hat{\beta}_n \end{Vmatrix} , \qquad (5.5)$$

and $\hat{\beta}$ is defined in Eq.$(98)_2$ of Ref.15. The eigenfrequencies corresponding to the eigensolutions for the harmonic modes are given by

$$\omega^2_{noo} = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \times \left(\sqrt{\frac{M_n}{\hat{c}^{(1)}}} + \sqrt{\frac{P_n}{\hat{c}^{(1)}}}\right)\right] , \qquad (5.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.15. The other displacement components of the equivalent trapped energy mode shape are discussed in Sec.6 of Ref.8.

6. Calculated Results

From Section 5 we now know g_γ^μ and from Sections 2 and 4 we know $\hat{c}_{L\gamma M\alpha}$ for normal acceleration, i.e., flexure of the box structure. Hence, we can now evaluate H_μ in Eq.(2.2). Since the perturbation integral H_μ has been shown to vanish for a perfectly symmetric mode-shape plus support configuration, i.e., when both d_A and f_i vanish in this work, the substitution of the biasing state in (2.2) and expansion to first order in d_A or f_i yields, with the aid of (2.1)

$$\Delta\omega/\omega)_d = C_A d_A, \quad \Delta\omega/\omega)_{f_i} = C_{f_i} f_i , \qquad (6.1)$$

where the C_A and the C_{f_i} are more readily calculated than obtained analytically. The f_i denote the mispositioning of a clip and are scalars, while the d_A denote the components of the planar displacement of the center of the mode shape from the center of the rectangle. Since the f_i are scalars, the C_{f_i} are scalars, but since the d_A are the components of a planar vector, the C_A are the components of a planar vector, the magnitude of which

$$C = \sqrt{C_A C_A} , \qquad (6.2)$$

is an important scalar, as are the C_{f_i}. In earlier work[1] C was referred to as an offset coefficient. We now change the name to a mispositioning coefficient. The C denotes the mispositioning coefficient for the mode center and the C_{f_i} denotes the mispositioning coefficient for the ith clip shown in Fig.1. Since for a perfectly symmetric system the normal acceleration sensitivity vanishes, the C and the C_{f_i} are the important quantities to study because when they are minimized, the influence of an error in fabrication is minimized. Since C and C_{f_i} will be plotted in this work and the significance of their value is not familiar, the relation between the values of C and the normal acceleration sensitivity for a range of mispositioning dimensions is shown by the straight lines in Fig.4.

Calculations of C and C_{f_i} have been performed using the known values of the second order[16] and the third order[17] elastic constants of quartz. In Fig.5 we have plotted the mispositioning coefficients as a function of the ratio of the thickness of a cover plate to the active plate for an SC cut with an orientation of the support rectangle $\beta = -50°$ and an aspect ratio a/b = 1.5. It can be seen from the figure that the mispositioning coefficients decrease as the thickness of the cover plates, i.e., the stiffness, increases. The figure also shows that the mispositioning of the mode center is much more deleterious than the mispositioning of the clips. The orientation of the support rectangle $\beta = -50°$ is used because it was found in earlier work[1] that for a simply-supported active plate a sharp minimum in C was found at this orientation for an aspect ratio a/b = 2.0. This orientation of the support rectangle of an SC cut is maintained in all calculations presented in this work, as is the ratio $a_c/a = 0.5$ for the openings in the cover plates. Similar calculations have been performed for the same plate, but with an aspect ratio a/b = 1.8 and are plotted in Fig.6, which shows that the mispositioning of the mode center has become less significant and the mispositioning of the clips a bit more. This large change in C for a relatively small change in aspect ratio is a result of the fact that the simply-supported plate has a sharp minimum at $\beta = -50°$ when a/b = 2.0.

In Fig.7 we have shown the mispositioning coefficients as a function of the ratio of thicknesses for the critical aspect ratio a/b = 2.0. It can be seen from the figure that the influence of the mispositioning of the mode center is very low throughout the range of thickness ratio and that the influence of the mispositioning of the clips decreases as the thickness ratio increases. This figure shows a large range of thickness ratio in which the C curve increases with increasing thickness ratio. This is directly related to the fact that the C curve for the simply supported plate shows a sharp minimum at this orientation for this aspect ratio, which means that when the active plate is free to rotate at the edges C is very low for this case. Since the cover plates and sidewalls stiffen the edge of the active region,

they serve to prevent the free rotation at the edges. Hence, for this case C can increase with increasing $\hat{h}/h$, i.e., stiffness and can also decrease with increasing $\hat{h}/h$ because the free edge is not necessarily an absolute minimum. Figure 7 indicates that the case in which the aspect ratio a/b = 2.0 and the ratio of the thickness of a cover plate to the active plate $\hat{h}/h = 2.0$ is the best for minimizing the influence of fabrication imperfection on normal acceleration sensitivity. Similar calculations for the square plate are shown in Fig.8. It can be seen from the figure that for the square plate the influence of a mispositioning of the mode center is truly abominable. This emphasizes the importance of the anisotropy of the support configuration in minimizing the influence of fabrication errors. The results of Figs.5 - 8 are summarized in Figs.9 and 10 for C and C_{f_2}, respectively.

Figures 11 and 12 show the mispositioning coefficients as a function of the size of the openings in the cover plates for two different thickness ratios. The figures show that there is not much change for a given thickness ratio $\hat{h}/h$, but that the C_{f_i} coefficients for $\hat{h}/h = 2.0$ are about one third of the values for $\hat{h}/h = 1.0$. The C curve is different in the two figures in accordance with Fig.9.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No.DAAL03-88-K-0065 and the National Science Foundation under Grant No. MSS-8908188.

References

1. Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators with the Mode Shape Displaced with Respect to Rectangular Supports," J. Appl. Phys., 69, 2862 (1991).
2. H.F. Tiersten and Y.S. Zhou, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," Proceedings of the 44th Annual Symposium on Frequency Control," U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No.90CH2818-3, 461 (1990)
3. H.F. Tiersten and Y.S. Zhou, "On the Influence of a Fabrication Imperfection on the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," to be published in the Proceedings of the 5th European Frequency and Time Forum, EFTF91, Besançon, France (1991).
4. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," J. Appl. Phys., 64, 4334 (1988).
5. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," J. Appl. Phys., 67, 60 (1990).
6. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
7. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).
8. H.F. Tiersten and Y.S. Zhou, "The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Due to its Thickness Asymmetry," Proceedings of the 45th Annual Symposium on Frequency Control (these proceedings).
9. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., 9, 1445 (1968).
10. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4.
11. L. Brand, Vector and Tensor Analysis (Wiley, New York, 1947), p.222
12. For more detail see Y.S. Zhou and H.F. Tiersten, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Quartz Cover Plates Supported by Clips," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
13. L.V. Kantorovich and V.I. Krylov, Approximate Methods of Higher Analysis (Interscience Publ. Inc., New York and P. Noordhoff Ltd., Groningen, The Netherlands, transl. by C.D. Benster from 3rd Russian ed., 1964), pp.258-260, 272-273, 279-281.
14. L. Collatz, The Numerical Treatment of Differential Equations, transl. by P.C. Williams (Springer-Verlag, Berlin, 1960), 2nd ed., pp.202-207 and 213-216.
15. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).
16. R. Bechmann. "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).
17. R.N. Thurston. H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

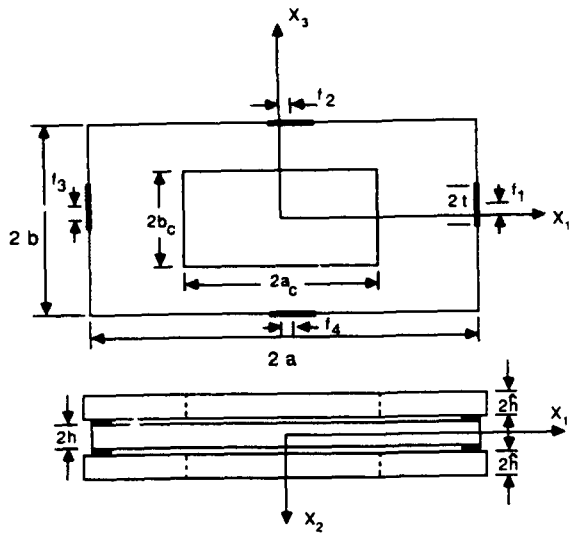


Figure 1 Plan View and Cross-Section of Box Structure Showing Active Plate, Cover Plates and Sidewalls

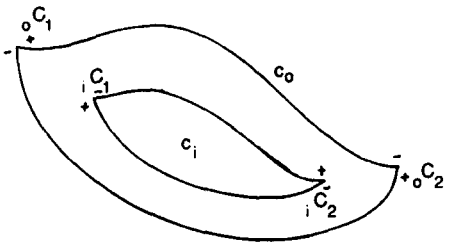


Figure 2 Plan View of Asymmetric Configuration with Arbitrary Curved Edges and Corners for Derivation of Variational Principle

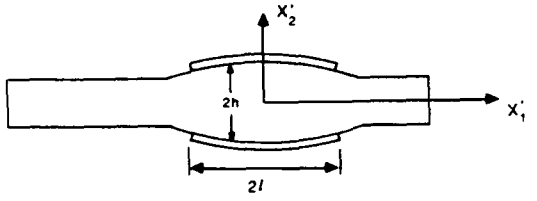


Figure 3 Cross-Section of Biconvex Resonator

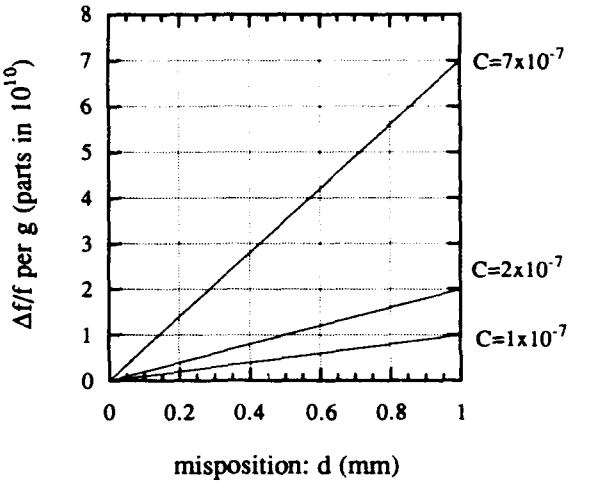


Figure 4 Acceleration Sensitivity Versus Mispositioning Dimension for Different Values of the Mispositioning Coefficient

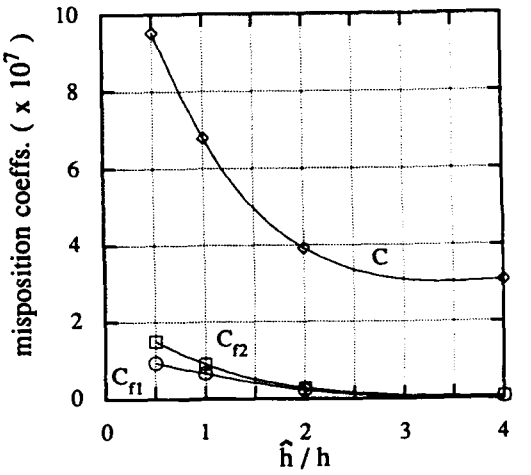


Figure 5 Mispositioning Coefficients Versus Thickness-Ratio for Aspect Ratio a/b = 1.5, where ab = 100 m^2 is Maintained in all Calculations, i.e., for Figs.5-12

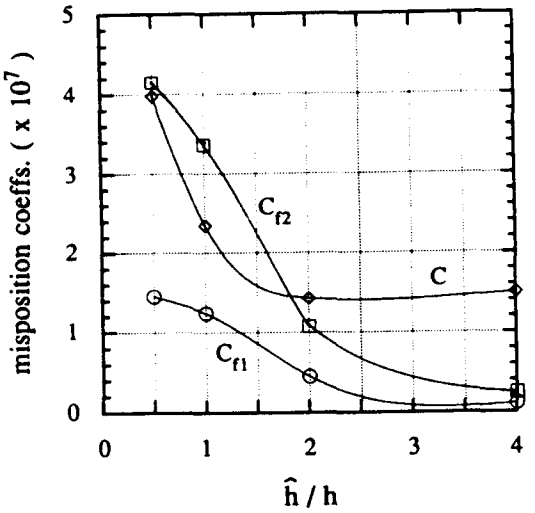


Figure 6 Mispositioning Coefficients Versus Thickness Ratio for a/b = 1.8

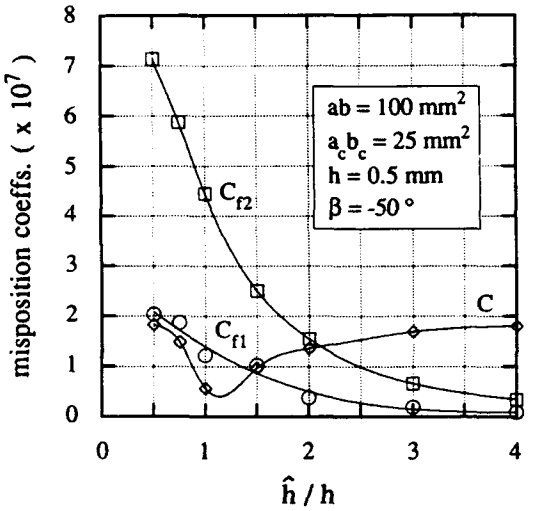


Figure 7 Mispositioning Coefficients Versus Thickness Ratio for a/b = 2.0

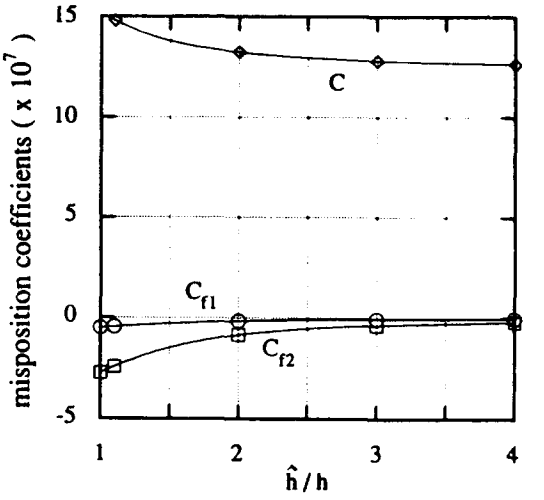


Figure 8 Mispositioning Coefficients Versus Thickness-Ratio for a/b = 1.0

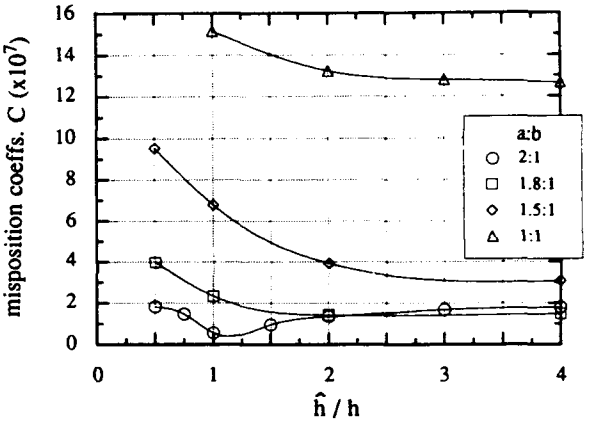


Figure 9 Coefficients for Mispositioning of Mode Center Versus Thickness-Ratio from Figs.5 - 8

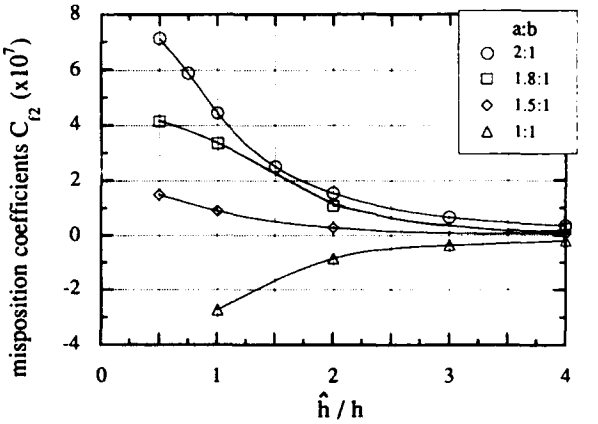


Figure 10 Coefficients for Mispositioning of Clips Versus Thickness-Ratio from Figs.5 - 8

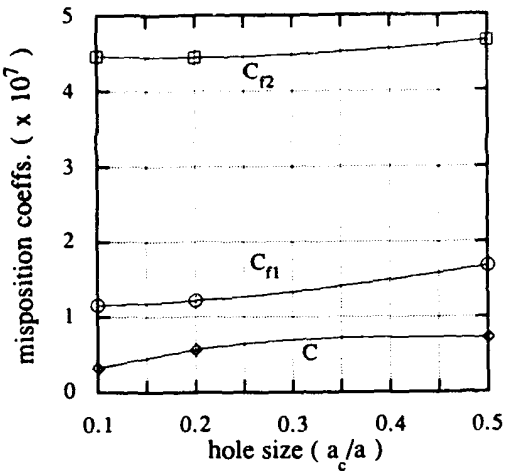


Figure 11 Mispositioning Coefficients Versus Opening Ratio for Thickness-Ratio $\hat{h}/h = 1.0$

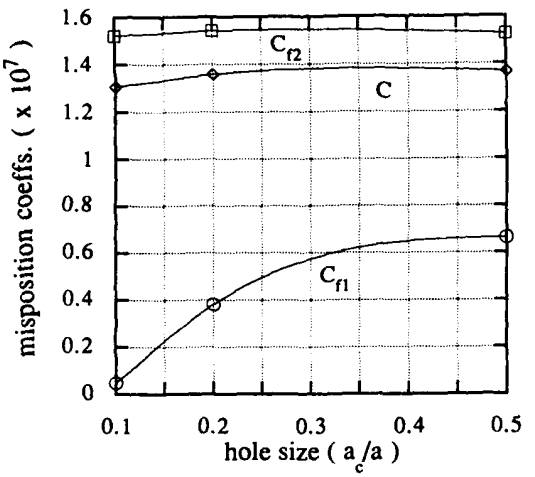


Figure 12 Mispositioning Coefficients Versus Opening Ratio for Thickness Ratio $\hat{h}/h = 2.0$

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY SHIFT CALCULATIONS FOR QUARTZ RESONATORS

LARRY D. CLAYTON AND ERROL P. EERNISSE
Quartztronics, Inc.
1020 Atherton Drive, Bldg. C
Salt Lake City, UT 84123

ABSTRACT

A new solution technique has been developed to calculate frequency shifts for quartz resonators arising from mechanical stresses. This solution technique utilizes finite element analysis as an initial step to calculate mechanical stress distributions in quartz resonators. Output from the finite element solution is used in a recently developed program to calculate resonator frequency shifts as the final step. Frequency shifts are calculated via numerical integration of the perturbation integral derived by Tiersten. The solution technique is general in that any combination of mount and resonator geometries may be modeled. Any crystallographic orientation may be chosen and any load or combination of loads may be applied to the resonator. The frequency perturbation calculation includes movement of the mode to any position in the general vicinity of the resonator center. Experimental results for AT- and SC-cut resonators subjected to diametric force and inertial loading verify the accuracy of frequency shifts calculated using the new solution technique. Minimum acceleration sensitivities corresponding to optimum mode locations and mount orientations were calculated for ideal SC-cut crystal units. Simple mount imperfections were modeled to determine what affect these might have upon resonator performance.

INTRODUCTION

Numerous theoretical and empirical studies have been performed over the past thirty years investigating the effects of static and dynamic mechanical stresses upon the performance of quartz crystal resonators [1-30]. Resonator stresses investigated in these studies arise from many loading conditions including diametric force pairs [1-9], surface tractions generated by thin metal films [10], thermal effects [11-15], and time-varying inertial loads generated by environmental vibrations [16-30]. Of particular concern to investigators has been the effect these and other load conditions have upon the accuracy and stability of the resonator frequency. A new solution technique has been developed to calculate resonator frequency shifts arising from virtually any combination of loading and boundary conditions.

Mechanical stresses induced by external actions and distributed throughout the resonator body are computed using finite element analysis (FEA). The finite element method is a numerical technique for formulating and solving the governing differential equations for continuum problems [31-35]. Finite element models may be created which accurately characterize the mechanical behavior of quartz resonators and the structures which support them. Models may be built which include full anisotropic material properties, differences in thermal expansion, interfacial and geometric effects. Static or dynamic loads may be concentrated at a point, distributed over a surface or throughout a volume.

A program has been written which retrieves element stresses and nodal displacements from output files generated for finite element models of quartz resonators. This program computes resonator frequency shifts for the finite element solution via numerical integration of the perturbation equation derived by H.F. Tiersten [36]. Recent studies suggest that the location of the mode with respect to stresses distributed throughout the resonator as well as distortions in mode shape affect resonant frequency shifts [21,22,24,25,27,28]. The formulation of the perturbation integral used in the new solution technique incorporates movement of the mode center location with respect to resonator stress distributions. Details of the new solution technique are presented in Appendix A.

Independent experimental results were sought to verify the accuracy of frequency shifts computed using the frequency perturbation program and finite element models of quartz resonators. Frequency shifts measured for resonators subjected to diametrically opposed in-plane forces and for crystal units subjected to steady-state acceleration were selected for this purpose. Force-frequency coefficients calculated for AT- and SC-cut resonators subjected to diametric forces at varying azimuth angles were compared with published experimental measurements [5,6,and7]. The minimum acceleration sensitivity predicted for the optimum off-center mode location was compared to optimized acceleration sensitivities achieved experimentally for AT-cut crystal units [21].

Acceleration-induced frequency shifts were also

CH2965-2/91/0000-309 $1.00 © 1991 IEEE

calculated for SC-cut crystal units. Calculations were performed for variations in the orientation of a four-point mount with respect to the crystallographic axes of the SC-cut resonator. Minimum acceleration sensitivities and the corresponding optimum mode locations were derived for each mount orientation. Finite element models of ideal mount and resonator geometries were used for these calculations.

Manufacturing tolerances will result in imperfect mount and resonator geometries. Finite element models of an SC-cut crystal unit were modified to approximate simple mount imperfections. Optimum mode locations which minimize acceleration sensitivity were calculated to determine what effect imperfect mount geometries might have upon acceleration-induced frequency shifts. These calculations revealed that mount imperfections are an important factor contributing to increased acceleration sensitivity and reduced precision for manufactured crystal units.

FORCE-FREQUENCY COEFFICIENTS CALCULATED FOR AT- AND SC-CUT QUARTZ RESONATORS

The solution technique described in Appendix A has been used to calculate the force-frequency effect for AT- and SC-cut quartz resonators. Force-frequency coefficients for AT- and SC-cut resonators have been measured in the past by numerous investigators [1-9]. Force-frequency coefficients for quartz resonators are computed using [5]

$$K_f = \frac{D t_o}{F N_o} \frac{\Delta f}{f_o} \qquad (1)$$

where D is the resonator diameter, t_o is the maximum thickness at its origin, and N_o its frequency constant. Fractional frequency shifts $\Delta f/f_o$ in Equation 1 arise from in-plane diametric forces of magnitude F concentrated at the resonator circumference. Prior investigations have focused upon measuring what effect varying the azimuthal angle ψ locating the angular orientation of the diametric forces has upon K_f.

Finite element models of AT- and SC-cut quartz resonators subjected to diametric forces at varying azimuth angles have been created. Frequency shifts computed using the new solution technique for each orientation were converted to force-frequency coefficients for comparison to experimental results. Crystallographic orientations used in these calculations for the AT- and SC-cuts are $\phi = 0°; \theta = 35.25°$ and $\phi = 22.4°; \theta = 34.3°$, respectively. Both models are of plano-convex resonators with a diameter of 1.397cm (.550") and a radius of curvature of 141.333cm (3/8 diopter). A thickness of .0688cm (.027") at the center of each resonator produces a fundamental C-mode frequency of 2.4MHz (N_o = 1651 Hz-m) for the AT-cut and 2.636MHz (N_o = 1813 Hz-m) for the SC-cut.

Figure 1 depicts the deformed shape of the FEA model for the SC-cut resonator subjected to 100 Dyne diametric forces at an azimuthal angle of $\psi = 90°$. The cross-hatched region at the center of the model is the portion of the active region over which numerical integration of the perturbation integral was performed. Force-induced frequency shifts were calculated for the fundamental and third overtone of the C-mode for each resonator. Azimuthal angles ranged from 0° to 90° for the AT-cut resonator and from 0° to 180° for the SC-cut resonator. Figure 2 depicts variations in the force-frequency coefficient with azimuth angle measured for the AT-cut resonator C-mode [5,6]. A similar curve [7] for the SC-cut C-mode appears in Figure 3. Experimental results in Figures 2 and 3 are for resonators at room temperature (25°C).

Force-frequency coefficients calculated for both resonators are plotted for comparison with the experimental results in Figures 2 and 3. The agreement between the calculated and measured force-frequency coefficients is excellent for the AT-cut resonator. Force-frequency coefficients calculated for the SC-cut resonator are in good agreement with the measured results. Small discrepancies might be attributed to slight differences in the orientations of the SC-cut resonators as modeled and tested.

ACCELERATION SENSITIVITY CALCULATED FOR AT-CUT CRYSTAL UNITS

Having verified that the solution technique accurately calculates frequency shifts for AT- and SC-cut resonators loaded diametrically, acceleration-induced frequency shifts were studied next. Mount Reactions to either steady-state or time-varying inertial loads produce mechanical stresses distributed throughout the resonator. Mechanical stresses in the active region change the resonant mode frequencies of quartz crystal oscillators. Frequency changes in proportion to the inertial load magnitude and depends upon the resonator orientation relative to the acceleration direction. Acceleration sensitivity of quartz resonators is typically expressed as a vector quantity known as the gamma vector $\vec{\Gamma}$. Gamma vector components are fractional frequency shifts produced by unit acceleration in three orthogonal directions. Fractional frequency shifts arising from acceleration in any direction are computed as the scalar product of the gamma vector with the acceleration

vector

$$\frac{\Delta f}{f_o} = \vec{\Gamma} \cdot \vec{a} \qquad (2)$$

where f_o is the frequency for zero acceleration. Gamma vectors for precision resonators typically have magnitudes ranging from 10^{-9} to 10^{-10} per g.

Independent experimental results reported by Smythe and Horton [21] were selected to verify the accuracy of acceleration-induced frequency shifts computed using the new solution technique. Smythe and Horton describe a multi-electrode technique for altering mode shape and location using a simple external circuit adjustment to spatially control mode excitation levels. They were able to selectively null individual components of $\vec{\Gamma}$ and minimize gamma vector magnitudes $|\vec{\Gamma}|$ by iteratively varying the relative excitations of multi-electroded resonators. They reduced acceleration sensitivities by factors ranging from 2.3 to 5.3 and achieved a minimum sensitivity of $1.46X10^{-10}$/g.

A finite element model of the crystal units tested [21] has been created to investigate acceleration effects. The model, depicted in Figure 4, is of a 30MHz, 3rd, AT-cut disc resonator supported by four Houston Electronics M-704 clip mounts equally spaced around its circumference. The mounts are aligned with the crystallographic X- and Z'-axes ($\psi = 0°$) which correspond to the global X- and Z-axes of the model. Material properties for the AT-cut (i.e., $\phi = 0°; \theta = 35.25°$) were entered into the model of a plano-plano resonator with a diameter of 0.787cm (0.31"). A constant resonator thickness of .0165cm (.0065") produces a fundamental C-mode frequency of 10.046MHz (N_o = 1657.65 Hz-m) for this resonator. Gravitational acceleration (i.e., self weight) was applied to the model along three mutually orthogonal axes. Figure 5 depicts the deformed shape of the resonator model subjected to gravitational acceleration in the global Z-direction.

Gamma vector components were calculated for the C-mode (3rd overtone) located at the center of the resonator and offset ± .05cm in the X- and Z-directions. Gamma vector components sum to form the gamma vector

$$\vec{\Gamma} = \Gamma_i \hat{i} + \Gamma_j \hat{j} + \Gamma_k \hat{k} \qquad (3)$$

and are used to calculate the gamma vector magnitude

$$|\vec{\Gamma}| = \sqrt{\Gamma_i^2 + \Gamma_j^2 + \Gamma_k^2} \qquad (4)$$

Off-center gamma vector components were used to linearly extrapolate the optimum mode location which minimizes $|\vec{\Gamma}|$.

To simulate the scatter in $|\vec{\Gamma}|$ measured by Smythe and Horton for crystal units prior to optimization, gamma vector components computed for ± .05cm offsets in the X- and Z-directions were extrapolated to ± 0.1cm offsets. The scatter in $|\vec{\Gamma}|$ calculated for movement of the mode within a 0.2cm circle centered at the resonator origin is depicted in Figure 6. The measured scatter [21] for crystal units prior to minimization of $|\vec{\Gamma}|$ is also depicted in Figure 6 for comparison. Evidently, manufacturing tolerances resulting in imperfect support, resonator, and electrode geometries in combination with material imperfections give rise to variations in $|\vec{\Gamma}|$ which may be approximated by random movement of the mode 0.1cm off-center. By iteratively adjusting the spatial excitation of the mode, the scatter and acceleration sensitivities for the experimental crystal units were significantly reduced. Optimized $|\vec{\Gamma}|$ for selected crystal units plotted in Figure 6 range from $1.46X10^{-10}$/g to $3.41X10^{-10}$/g. Considering unavoidable differences between the crystal units and the idealized FEA model and discretization errors arising from the approximate solution technique, experimental optimums are in surprisingly good agreement with the predicted optimum of $1.7X10^{-10}$/g.

ACCELERATION SENSITIVITY CALCULATED FOR SC-CUT CRYSTAL UNITS

Additional acceleration sensitivity calculations were performed for SC-cut crystal units using the new solution technique. The crystal unit modeled for these calculations is a 10MHz, 3rd, SC-cut resonator supported by four "V" clip-mounts distributed at equal angular intervals around its circumference. A drawing of this crystal unit appears in Figure 7.

A finite element model of the SC-cut crystal unit created for gamma vector calculations appears in Figure 8. Gamma vector components were computed for mount orientations ranging from $\psi = 0°$ to $\psi = +90°$ in 15° increments. Figure 9 depicts the sense of the mount orientation angle ψ with respect to the crystallographic axes of the SC-cut resonator (X", Z"). Variations in mount orientation were entered into the model via changes in the material properties of the elements modeling the resonator. Gamma vector components for orientations ranging from $\psi = -90°$ to $\psi = -15°$ were inferred from components for the positive orientations using symmetry. Frequency shifts were calculated for the C-mode (3rd overtone) located at the center of the resonator and off-center ± 0.1cm in the X- and Z-directions for each mount orientation. Off-center gamma vector components were used to linearly extrapolate optimum mode locations which minimize

$|\Gamma|$. Components of the spatial vector in the optimum direction for each mount orientation were used to extrapolate $|\Gamma|$ for the mode offset 0.1cm towards the optimum location.

Figure 10 depicts variations in $|\Gamma|$ with mount orientation for the mode at the center of the resonator, off-center 0.1cm in the optimum direction, and at the optimum location. Moving the mode off-center 0.1cm towards the optimum reduces acceleration sensitivity for all mount orientations. Moving the mode further to the optimum location for each mount orientation results in additional reductions in $|\Gamma|$. The largest reduction in $|\Gamma|$ occurs for $\psi = 0°$ and by symmetry $\psi = \pm 90°$. Moving the mode 0.23cm to the optimum location for this orientation reduces the gamma vector by a factor of 374 from $4.49X10^{-10}$/g to $.012X10^{-10}$/g. Movement of the mode this distance is not practical using selective mass loading [22,24,25] or excitation techniques [21], but, any movement towards the optimum will result in improvements.

Figure 11 depicts the variation in the scatter of $|\Gamma|$ with mount orientation for the mode off-center ± 0.1cm in the X- and Z-directions. Figure 11 also depicts variations in $|\Gamma|$ with ψ for the mode located at the resonator center and off-center 0.1cm towards the optimum. This figure illustrates the range of acceleration sensitivities which might occur in manufactured crystal units for various mount orientations. This scatter, produced by movement of the mode within a 0.2cm circle centered at the resonator origin, characterizes variations in acceleration sensitivity which could arise from manufacturing limitations and material imperfections.

Acceleration sensitivity calculations described to this point have been based upon finite element models of ideal mount and resonator geometries. Manufacturing tolerances result in imperfect mounts and resonators. Careful inspection of crystal units manufactured with "V" clip mounts as well as other mounts confirmed that significant deviations from ideal mount geometries do occur. To investigate what effect mount imperfections might have upon acceleration sensitivity, the FEA model of the SC-cut crystal unit was modified. Both mounts along the X-axis were skewed so that the attachment point of one mount was raised .015" with respect to its ideal position while the attachment point for the opposite mount was lowered .015". The same modification was made to mounts along the Z-axis for a second model depicted in Figure 12. The mounts were aligned with the crystallographic axes of the SC-cut resonator (i.e., $\psi = 0°; \pm 90°$) for both models. Prior calculations indicated that the largest reductions in $|\Gamma|$ are achieved for this mount orientation.

Gamma vector components were calculated for the mode located at the resonator center, off-center ± 0.1cm along the X- and Z-axes, at the optimum location, and offset 0.1cm and 0.2cm in the optimum direction. Figure 13 depicts the scatter in $|\Gamma|$ for the mode off-center ± 0.1cm and the resonator supported by ideal mounts and mounts skewed in the X- and Z-directions. Figure 13 also depicts the minimum acceleration sensitivities calculated for each model and acceleration sensitivities for the mode moved 0.1cm and 0.2cm in the optimum direction. The minimum $|\Gamma|$ computed for the ideal model ($.012X10^{-10}$/g) increases by a factor of 290 to $3.48X10^{-10}$/g for the model of the resonator supported by mounts skewed along the X-axis. The minimum $|\Gamma|$ increases by a factor of 81 to $0.97X10^{-10}$/g for the model with mounts skewed along the Z-axis. Imperfect mounts also resulted in increased scatter and significant increases in acceleration sensitivities for the other mode locations.

Calculations for the SC-cut crystal unit with imperfect mounts indicate that mode position is not the only critical factor controlling the acceleration sensitivities of manufactured crystal units. Equally important are alterations in resonator stress distributions which arise from imperfect and inconsistent mount geometries.

CONCLUSION

A new solution technique has been presented which accurately computes quartz resonator frequency shifts arising from virtually any combination of loading and boundary conditions. This solution technique combines finite element analysis with numerical integration of the frequency perturbation integral derived by H.F. Tiersten [36]. The inherent strength of this solution technique stems from the capability of the finite element method to accurately model the mechanical response of structural systems. Models may be built which characterize anisotropic material properties, ideal and imperfect resonator and mount geometries, interfacial effects, and virtually any form of external loading. Frequency shifts may be evaluated using the new solution technique for the normalized mode shape located anywhere in the general vicinity of the resonator center.

Independent experimental results confirmed the accuracy of frequency shifts calculated using the new solution technique. Calculated variations in the force-frequency coefficient with azimuth angle ψ for AT- and SC-cut resonators agreed closely with measured results [5,6,7]. The minimum acceleration sensitivity predicted for AT-cut crystal units modeled after those tested by Smythe and Horton [21] is in good agreement with

optimized acceleration sensitivities achieved by iteratively varying the relative excitations of multiple electrodes. Scatter in the acceleration sensitivity calculated for the mode off-center ± 0.1cm approximately characterizes the measured scatter for AT-cut crystal units prior to optimization [21].

Additional gamma vector calculations identified critical factors affecting the acceleration sensitivity of manufactured crystal units. Gamma vector components were calculated for SC-cut crystal units with various mode locations and mount orientations. Calculations identified mount orientations and mode locations for which the largest reductions in acceleration sensitivity are achieved. Solution results predict that a minimum $|\Gamma|$ of $.012X10^{-10}/g$ will occur for $\psi=0°;\pm 90°$ and the mode offset 0.23cm to the optimum location. These results were based upon finite element models of an ideal crystal unit. Modifying the finite element model to include simple mount imperfections resulted in significant increases in acceleration sensitivity and scatter. Manufacturing tolerances resulting in imperfect mount and resonator geometries are unavoidable. Inconsistencies in support geometry and spacing alter ideal resonator stress distributions. Altered mode shapes and locations arise from imperfect resonator geometries. The combined effects of altered stress distributions and mode shapes are critical to controlling the consistency and performance of manufactured crystal units.

APPENDIX A
SOLUTION TECHNIQUE

Steady-state and time-varying external actions are balanced by internal stresses distributed throughout the elastic continuum of a quartz resonator in equilibrium. External actions include concentrated loads, distributed surface tractions, and body forces which act over the entire resonator volume. Mechanical stresses distributed throughout the body of the quartz resonator cause changes in the resonant frequencies of its vibration modes. Frequency changes arise from the combined effects of geometric and material nonlinearities. The shape of the resonator changes slightly under loading. Changes in shape are characterized by a nonlinear strain-displacement relationship referenced to a fixed Cartesian system and known as the Green-Lagrange strain tensor. Small changes in a resonator's shape alter its stiffness and characteristic frequencies slightly. Frequency changes also arise from nonlinear stress-strain relationships characterized by the third-order elastic coefficients of quartz.

Frequency shifts for quartz resonators subjected to external actions are best calculated using the perturbation integral derived by H.F. Tiersten [36]. The difference between the unperturbed ω_n and the perturbed ω eigenfrequencies is expressed as

$$\Delta\omega = \omega_n - \omega = \frac{H_n}{2\omega_n} \qquad (A.1)$$

where H_n is the perturbation integral. The perturbation integral, simplified for purely elastic nonlinearities, is given by [24]

$$-H_n = \int_V \hat{C}_{L\gamma M\alpha}\, g^n_{\alpha,M}\, g^n_{\gamma,L}\, dV \qquad (A.2)$$

where $g^n_{\alpha,M}$ and $g^n_{\gamma,L}$ are spatial gradients of the mass normalized unit eigenvectors for the nth odd harmonic of the resonator pure thickness mode triad [37]. The effective elastic coefficient $\hat{C}_{L\gamma M\alpha}$ is a tensor of the form

$$\hat{C}_{L\gamma M\alpha} = T_{LM}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}E_{KN} + C_{L\gamma KN}\omega_{\alpha,K} + C_{LKM\alpha}\omega_{\gamma,K} \qquad (A.3)$$

where the second rank tensors T_{LM} and E_{KN} contain stress and strain components and $\omega_{\alpha,K}$ and $\omega_{\gamma,K}$ are deformation gradients. The constitutive tensor with second-order elastic components $C_{L\gamma KM}$ relates stress to strain for linear elastic materials. The sixth rank tensor $C_{L\gamma M\alpha KN}$ contains third-order nonlinear elastic coefficients [38] and $\delta_{\gamma\alpha}$ is the Kronecker delta.

Since the last two terms on the right hand side of Equation A.3 contract to scalar quantities upon substitution into Equation A.2 (i.e., all indices are paired) repeated indices may be rearranged without affecting the summation. Rearranging the indices and exploiting the symmetry of $C_{LKM\alpha}$, these terms may be combined and Equation A.3 rewritten as

$$\hat{C}_{L\gamma M\alpha} = T_{LM}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}E_{KN} + 2C_{L\gamma KN}\omega_{\alpha,K} \qquad (A.4)$$

The deformation gradient $\omega_{\alpha,K}$ is a combination of the symmetric strain tensor

$$E_{\alpha K} = E_{K\alpha} = \frac{1}{2}(\omega_{\alpha,K} + \omega_{K,\alpha}) \qquad (A.5)$$

and the anti-symmetric rotation tensor

$$\Omega_{K\alpha} = \frac{1}{2}(\omega_{\alpha,K} - \omega_{K,\alpha}) \qquad (A.6)$$

Replacing $\omega_{\alpha,K}$ by $E_{\alpha K}+\Omega_{K\alpha}$, the effective elastic

coefficient tensor becomes

$$\hat{C}_{L\gamma M\alpha} = T_{LM}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}E_{KN} + 2C_{L\gamma KN}E_{\alpha K} + 2C_{L\gamma KN}\Omega_{K\alpha} \quad (A.7)$$

Strain is related to stress through the second-order compliance tensor S_{KNLM} as follows

$$E_{KN} = S_{KNLM}T_{LM} \quad (A.8)$$

which yields

$$\hat{C}_{L\gamma M\alpha} = (\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}S_{KNIJ} + 2C_{L\gamma KN}S_{\alpha KIJ})\,T_{IJ} + 2C_{L\gamma KN}\Omega_{\alpha K} \quad (A.9)$$

upon substitution into Equation A.7.

The effective elastic coefficient $\hat{C}_{L\gamma M\alpha}$ given by Equation A.9 is used in the formulation of an algorithm to numerically integrate the perturbation integral. Substituting Equation A.9 into Equation A.2 the perturbation integral becomes

$$-H_n = \int_V [\,(\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}S_{KNIJ} + 2C_{L\gamma KN}S_{\alpha KIJ})\,T_{IJ} + 2C_{L\gamma KN}\Omega_{\alpha K}]\; g^n_{\alpha,M}g^n_{\gamma,L}dV \quad (A.10)$$

which is the form used in the frequency perturbation program. Piecewise integration of the perturbation integral is performed numerically as follows

$$-H_n = \sum_{N=1}^{NEL} [\,(\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + C_{L\gamma M\alpha KN}S_{KNIJ} + 2C_{L\gamma KN}S_{\alpha KIJ})\,T_{IJ}(N) + 2C_{L\gamma KN}\Omega_{\alpha K}(N)\,]\; g^n_{\alpha,M}(N)\; g^n_{\gamma,L}(N)\; \Delta V(N) \quad (A.11)$$

using single quadrature points located at the centroids of elements in the finite element model of the resonator.

The summation in Equation A.11 is performed over NEL, the number of elements comprising a significant portion of the active region of the resonator mode. Components of the stress tensor $T_{IJ}(N)$, computed at the center of each element, are retrieved from finite element output files. Components of the antisymmetric tensor $\Omega_{\alpha K}(N)$, for rigid body rotations about the centroid of each element, are computed using nodal displacements retrieved from the FEA solution. Element centroids and volumes $\Delta V(N)$ are computed using the coordinates for nodes attached to each element used in the summation. Solution results are retrieved from output files created by SAP90$^+$, a finite element program marketed by Computers and Structures of Berkeley, California [34].

Spatial dependence of the thickness mode displacements components is given by [37]

$$g_i^n = \hat{g}_i^n \sin\left(n\pi\frac{Y''}{t}\right) e^{-\frac{\alpha_n}{2}X''^2} e^{-\frac{\beta_n}{2}Z''^2} \quad (n=1,3,5,\ldots) \quad (A.12)$$

where $\hat{g}_i^n$ are the eigenvector components prior to normalization. Eigenvector components for the odd-harmonics of the thickness mode triad are computed for the doubly-rotated orientation of the resonator crystallographic axes (X'',Y'',Z''). Constants α_n and β_n control spatial decay of the modal amplitude and t is the resonator thickness. Spatial dependence of the vibration displacement components may be rewritten in terms of the global system (X,Y,Z) of the FEA model using the following coordinate transformation

$$\begin{aligned} X'' &= (X-\Delta X)\cos a_n + (Z-\Delta Z)\sin a_n \\ Y'' &= Y \\ Z'' &= -(X-\Delta X)\sin a_n + (Z-\Delta Z)\cos a_n \end{aligned} \quad (A.13)$$

This transformation incorporates off-center translations ΔX and ΔZ of the mode center location with respect to the resonator origin into the frequency perturbation calculation. The transformed coordinates X'' and Z'' are rotated about the Y'' axis by the angle $-a_n$. The rotation angle, a_n, given by

$$a_n = \psi - \beta_n \quad (A.14)$$

is the difference between the orientation of the global FEA system with respect to the crystallographic axes of the resonator and the angle $\hat{\beta}_n$ derived to transform the dispersion relation into a separable form [37].

The spatially transformed thickness mode displacement components become

$$g_i^n = \hat{g}_i^n\, H(X,Y,Z) \quad (A.15)$$

where

$$H = \sin\left(\frac{n\pi Y}{t}\right)\left\{e^{-\frac{\alpha_n}{2}[(X-\Delta X)\cos a_n + (Z-\Delta Z)\sin a_n]^2} + e^{-\frac{\beta_n}{2}[-(X-\Delta X)\sin a_n + (Z-\Delta Z)\cos a_n]^2}\right\} \quad (A.16)$$

Gradients of the spatially transformed thickness mode components are obtained upon differentiating Equation A.16 with respect to the global system coordinates X, Y, and Z.

The mode shape gradients used in the perturbation integral are mass normalized. Eigenvector mass normalization factors are given by [36]

$$NF^2 = \int_V \rho_o\, g_i^n\, g_i^n\, dV \quad (A.17)$$

where g_i^n are the displacement components for the thickness mode shapes from Equation A.15 and ρ_o is the mass density of quartz. Integration of Equation A.17 requires knowledge of the spatial variation of the resonator thickness t. For plano-convex resonators t varies as follows

$$t = t_o\left[1 - \frac{(X^2+Z^2)}{2Rt_o}\right] \qquad (A.18)$$

where t_o is the maximum thickness at the resonator origin and R is its radius of curvature.

The second- and third-order elastic and compliance coefficients and eigenvector components are rotated to the global coordinate system of the FEA model prior to the contraction of $\hat{C}_{L\gamma M\alpha}$ with the mode shape gradients in Equation A.11. Discrete approximations of the mass-normalized mode shape gradients calculated at each quadrature point are

$$g_{a,b}^n(N) = \frac{g_a^b}{NF(N)} H_{,b}(N) \qquad (A.19)$$

Differentiating H(X,Y,Z) with respect to the global coordinates, spatial gradients calculated at the centroid of each element N become

$$H_{,X}(N) = \{-\alpha_n[(X(N)-\Delta X)\cos a_n + (Z(N)-\Delta Z)\sin a_n]\cos a_n + \beta_n[(Z(N)-\Delta Z)\cos a_n - (X(N)-\Delta X)\sin a_n]\sin a_n\} H(N) \qquad (A.20)$$

$$H_{,Y}(N) = \frac{n\pi}{t(N)}\cos\left(\frac{n\pi Y(N)}{t(N)}\right)\frac{H(N)}{\sin\left(\frac{n\pi Y(N)}{t(N)}\right)} \qquad (A.21)$$

$$H_{,Z}(N) = \{-\alpha_n[(X(N)-\Delta X)\cos a_n + (Z(N)-\Delta Z)\sin a_n]\sin a_n - \beta_n[(Z(N)-\Delta Z)\cos a_n - (X(N)-\Delta X)\sin a_n]\cos a_n\} H(N) \qquad (A.22)$$

where

$$H(N) = \sin\left(\frac{n\pi Y(N)}{t(N)}\right)\left\{e^{-\frac{\alpha_n}{2}[(X(N)-\Delta X)\cos a_n + (Z(N)-\Delta Z)\sin a_n]^2} + e^{-\frac{\beta_n}{2}[-(X(N)-\Delta X)\sin a_n + (Z(N)-\Delta Z)\cos a_n]^2}\right\} \qquad (A.23)$$

and

$$t(N) = t_o\left[1 - \frac{X^2(N)+Z^2(N)}{2Rt_o}\right] \qquad (A.24)$$

Consistent with numerical integration of the perturbation integral, the mass normalization factor in Equation A.19 is also integrated using single quadrature points at the element centroids

$$NF^2(N) = \rho_o g_i^n g_i^n \sum_{N=1}^{NEL} H^2(N)\,\Delta V(N) \qquad (A.25)$$

Piecewise formulations for all tensors used in the contraction and numerical integration of the perturbation integral have been presented. Following integration, H_n is used in Equa ion A.1 to compute the resonator frequency shift. Frequency perturbation calculations may be performed for quartz resonators subjected to virtually any combination of loading and boundary conditions. Effects of small off-center mode locations are automatically included in these calculations.

ACKNOWLEDGEMENTS

This work has been supported by the U.S. Army Laboratory Command Electronics Technology and Devices Laboratory under Contract No. DAAL01-90-C-0009.

REFERENCES

1. V. Bottom, "Note on the Anomalous Thermal Effect in Quartz Plates," Amer. Mineralogist, 32, 1947, pp. 590-591.

2. A.D. Ballato, "Effects of Initial Stress on Quartz Plates Vibrating in Thickness Modes," in Proceedings 14th Annual Symposium on Frequency Control, 1960, pp. 89-114.

3. A.D. Ballato and R. Bechmann, "Effect of Initial Stress on Vibrating Quartz Plates," Proceedings IRE, 48, 1960, pp. 261-262.

4. C.R. Mingins, L.C. Barcus, and R.W. Perry, "Reactions of a Vibrating Piezoelectric Plate to Externally Applied Forces," in Proceedings 17th Annual Symposium on Frequency Control, 1963, pp. 51-87.

5. J.M. Ratajski, "Force-Frequency Coefficient of Singly Rotated Vibrating Quartz Crystals," IBM J. Res. Dev., 12, No. 1, January 1968, pp. 92-99.

6. C.R. Dauwalter, "The Temperature Dependence of the Force Sensitivity of AT-cut Quartz Crystals,"in Proceedings 26th Annual Symposium on Frequency Control, 1972, pp. 108-112.

7. A. Ballato, E.P. EerNisse, and T. Lukaszek, "The Force-Frequency Effect in Doubly Rotated Quartz Resonators," in Proceedings 31st Annual Symposium on Frequency Control, 1977, pp. 8-16.

8. E.P. EerNisse, T.J. Lukaszek, and A. Ballato, "Variational Calculation of Force-Frequency Constants of Doubly Rotated Quartz Resonators," IEEE Transactions on Sonics and Ultrasonics, SA-25, No.3, May 1978, p. 132.

9. E.P. EerNisse, "Temperature Dependence of the Force Frequency Effect for the AT-, FC-, SC-, and Rotated X-Cuts," in Proceedings 34th Annual Symposium on Frequency Control, May 1980, pp. 426-430.

10. E.P. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," in Proceedings 29th Annual Symposium on Frequency Control, 1975, pp. 1-4.

11. R. Holland, "Nonuniformly Heated Anisotropic Plates: Mechanical Distortion and Relaxation," IEEE Transactions on Sonics and Ultrasonics, July 1974, pp. 171-178.

12. E.P. EerNisse, "Calculations on the Stress Compensated (SC-Cut) Quartz Resonator," in Proceedings 30th Annual Symposium on Frequency Control, 1976, pp. 8-11.

13. J.A. Kusters, "Transient Thermal Compensation for Quartz Resonators," IEEE Transactions on Sonics and Ultrasonics, July 1976, pp. 273-276.

14. J.A. Kusters, C. Adams, H. Yoshida, and J. Leach, "TTC's-Further Developments and Results," in Proceedings 31st Annual Symposium on Frequency Control, 1977, pp. 3-7.

15. P.E. Morley and K.J. Williamson, "Determination of the Optimum Orientation of an SC-Cut Resonator Using a Pulsed Laser," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 193-200.

16. A.W. Warner and W.L. Smith, "Quartz Crystal Units and Precision Oscillators for Operation in Severe Mechanical Environments," in Proceedings 14th Annual Symposium on Frequency Control, 1960, pp. 200-216.

17. W.L. Smith and W.J. Spencer, "Quartz Crystal Controlled Oscillators," Final Report, Contract DA36-039SC85373, U.S. Army, 1963.

18. R. Besson, et. al., "Design of Bulk Wave Quartz Resonator Insensitive to Acceleration," in Proceedings 33rd Annual Symposium on Frequency Control, 1979, pp. 337-345.

19. R.L. Filler, J.A. Kosinski, and J.R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," in Proceedings 37th Annual Symposium on Frequency Control, 1983, pp. 265-271.

20. R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 35, No.3, May 1988, pp. 297-305.

21. R.C. Smythe and W.H. Horton, "Adjustment of Resonator G-Sensitivity by Circuit Means," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 437-443.

22. E.P. EerNisse, R.W. Ward, and O.L. Wood, "Acceleration Induced Frequency Shifts in Quartz Resonators," in Proceedings 43rd Annual Symposium on Frequency Control, 1989, pp. 388-395.

23. D.V. Shick and H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," in Proceedings 43rd Annual Symposium on Frequency Control, 1989, pp. 405-411.

24. E.P. EerNisse, L.D. Clayton, and M.H. Watts, "Distortions of Thickness Shear Mode Shapes in Plano-Convex Quartz Resonators with Mass Perturbations," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 37, No. 6, Nov. 1990, pp. 571-576.

25. E.P. EerNisse, et. al., "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 37, No. 6, Nov. 1990, pp. 566-570.

26. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," J. Appl. Phys., 67, No. 60, 1990.

27. P.C.Y. Lee and X. Gyo, "Acceleration Sensitivity of Crystal Resonators Affected by Mass and Location of Electrodes," in Proceedings 44th Annual Symposium on Frequency Control, May, 1990, pp. 468-473.

28. Y.S. Zhou and H.F. Tiersten, "On the Influence of

A Fabrication Imperfection on the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," in Proceedings 44th Annual Symposium on Frequency Control, May, 1990, pp. 452-467.

29. P.C.Y. Lee and K-M Wu, "The Influence of Support-Configuration on the Acceleration Sensitivity of Quartz Resonator Plates," in Proceedings 31st Annual Symposium on Frequency Control, June 1977, pp. 29-34.

30. P.C.Y. Lee and K-M Wu, "Effects of Acceleration on the Resonance Frequencies of Crystal Plates," in Proceedings 30th Annual Symposium on Frequency Control, June 1976, pp. 1-7.

31. O.C. Zienkiewicz, The Finite Element Method, 3rd ed., McGraw-Hill, 1977.

32. K.J. Bathe, Finite Element Procedures in Engineering, Prentice-Hall, 1982.

33. R.D. Cook, Concepts and Applications of Finite Element Analysis, 2nd ed., John-Wiley and Sons, 1981.

34. SAP90^{+}, Computer and Structures, Inc., 1918 University Avenue, Berkeley, California.

35 L.D. Clayton and P.S. Evans, Finite Element Analysis on Desktop Computers: A Detailed Comparison of Alternatives, 2nd ed., Micro-Analytical Publications, 1990.

36. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., **64**, 1978, pp. 832-837.

37. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., **79**, 1986, pp. 1811-1826.

38. R.N. Thurston, H.J. McSkimmin, and P. Andreatch, Jr., "Third-Order Elastic Coefficints of Quartz," J. Applied Physics, **37**, No.1, Jan. 1966, pp. 267-275.

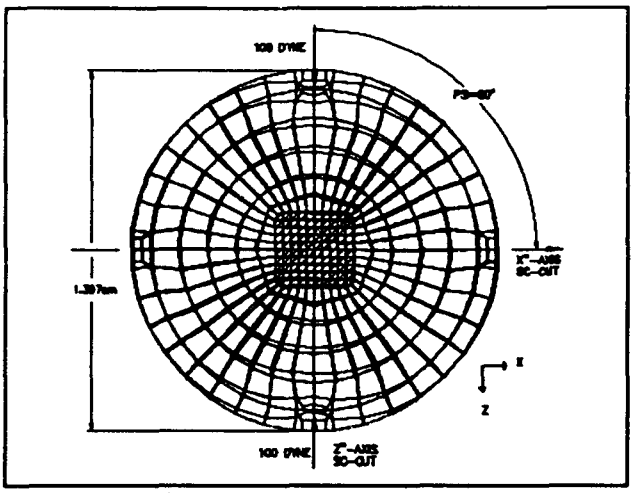


Figure 1. Deformed Shape of the SC-Cut Resonator Model Subjected to Diametric Forces at $\psi = 90°$

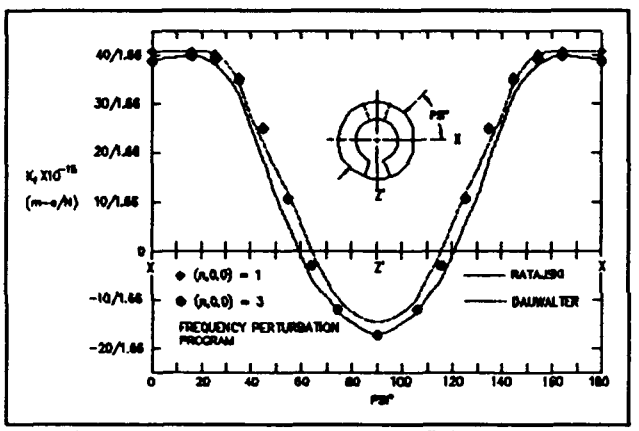


Figure 2. Force-Frequency Coefficient Versus Azimuth Angle for the AT-Cut Resonator

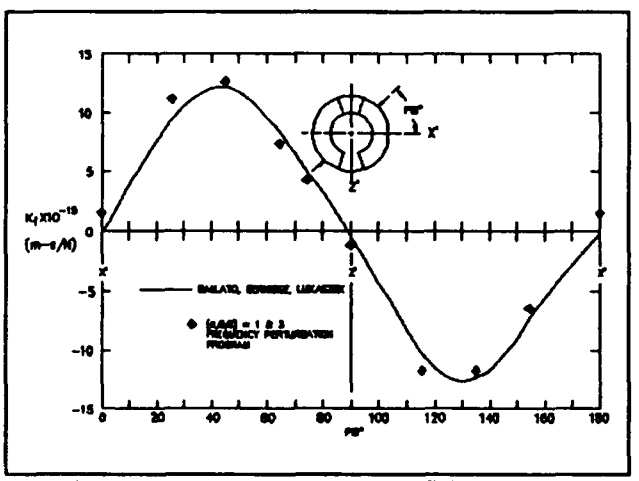


Figure 3. Force-Frequency Coefficient Versus Azimuth Angle for the SC-Cut Resonator

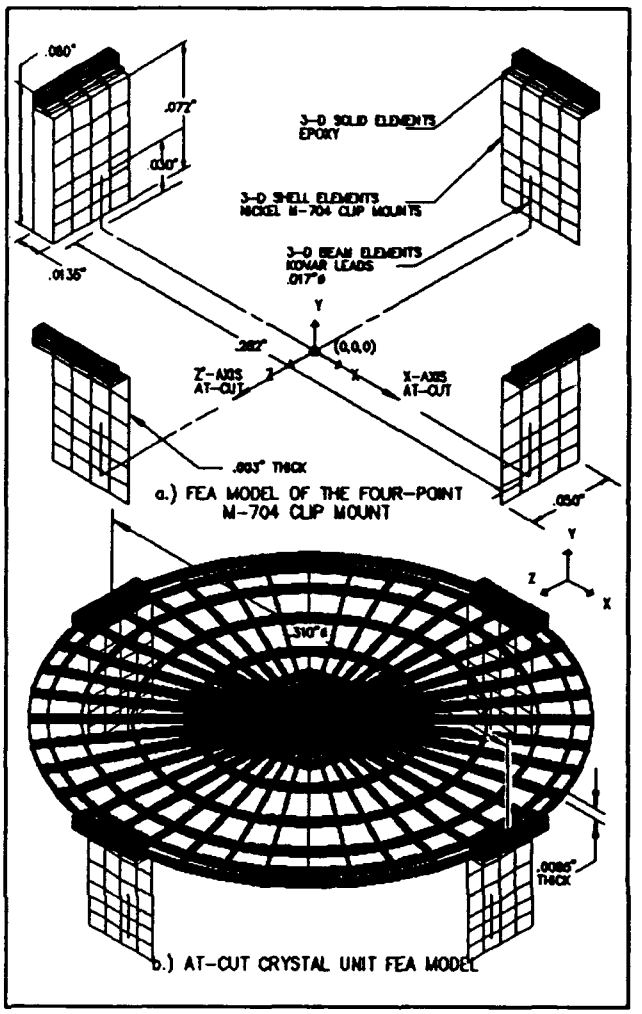


Figure 4. Finite Element Model of the Smythe and Horton AT-Cut Crystal Unit

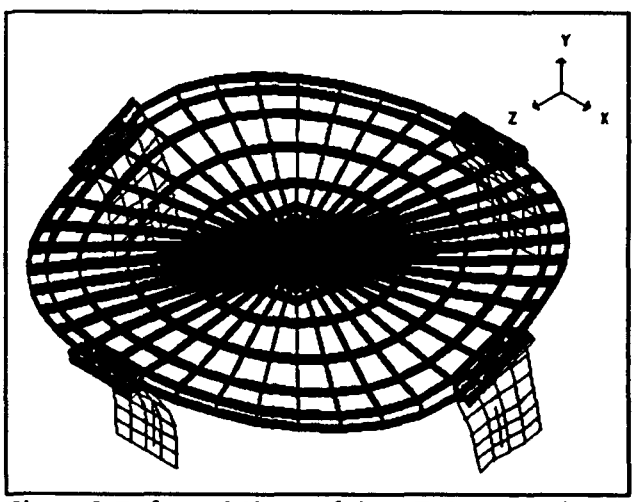


Figure 5. Deformed Shape of the AT-Cut Crystal Unit Subjected to Z-Axis Gravitational Loading

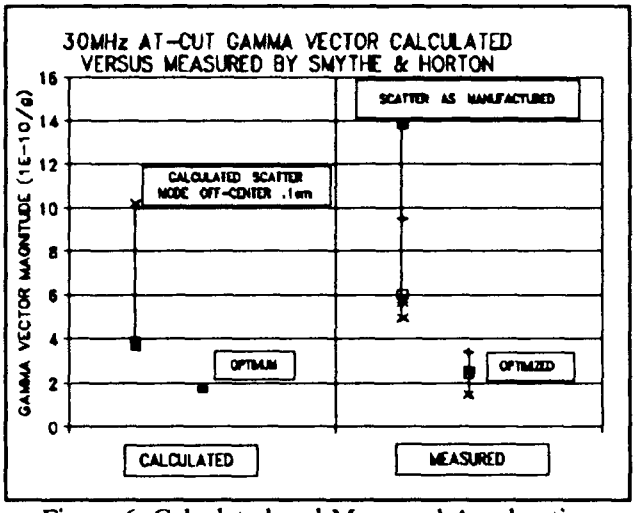


Figure 6. Calculated and Measured Acceleration Sensitivities for the AT-Cut Crystal Unit

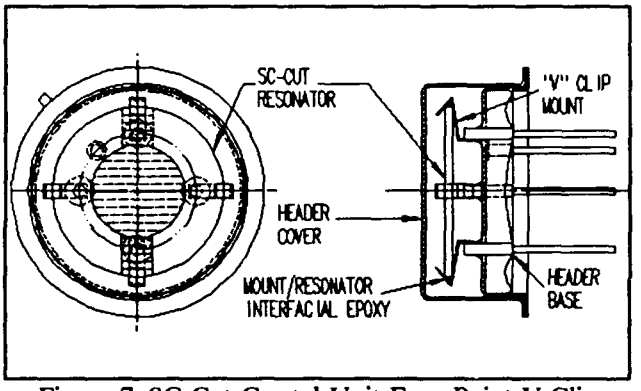


Figure 7. SC-Cut Crystal Unit Four-Point V-Clip Mount Package

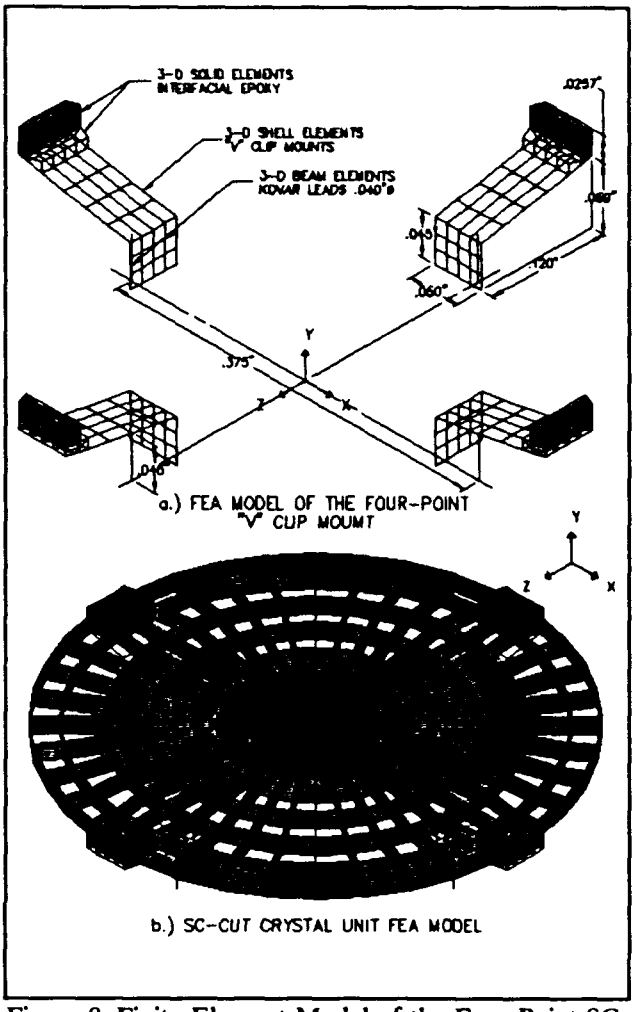


Figure 8. Finite Element Model of the Four-Point SC-Cut Crystal Unit

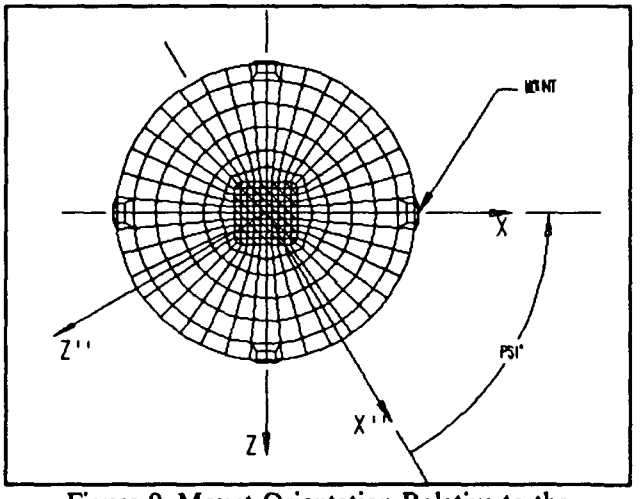


Figure 9. Mount Orientation Relative to the Crystallographic Axes of the SC-Cut Resonator

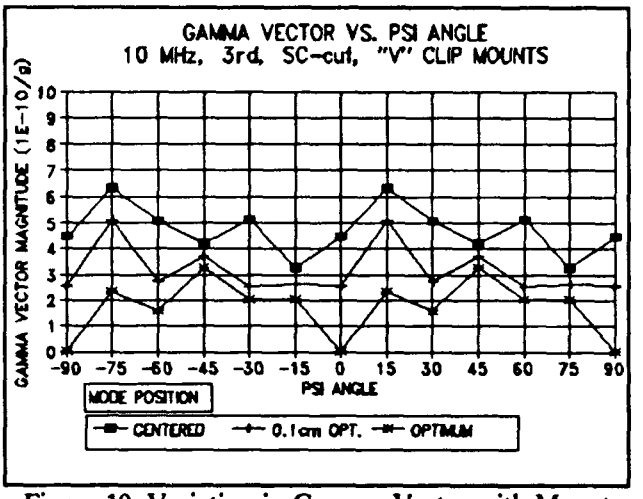


Figure 10. Variation in Gamma Vector with Mount Orientation for the SC-Cut Crystal Unit

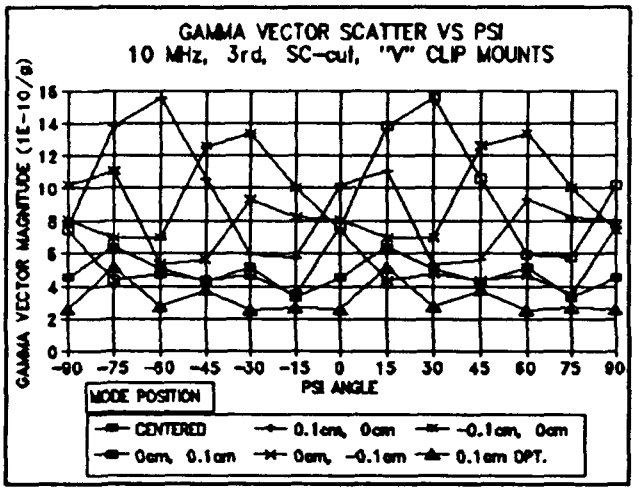


Figure 11. Gamma Vector Scatter Versus Mount Orientation for the SC-Cut Crystal Unit

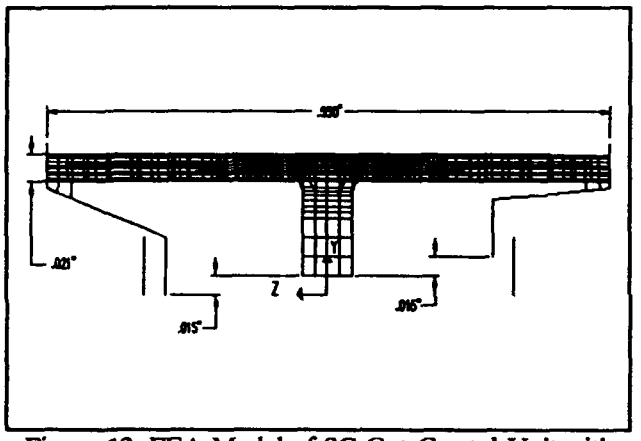


Figure 12. FEA Model of SC-Cut Crystal Unit with Imperfect Mounts Along the Z-Axis

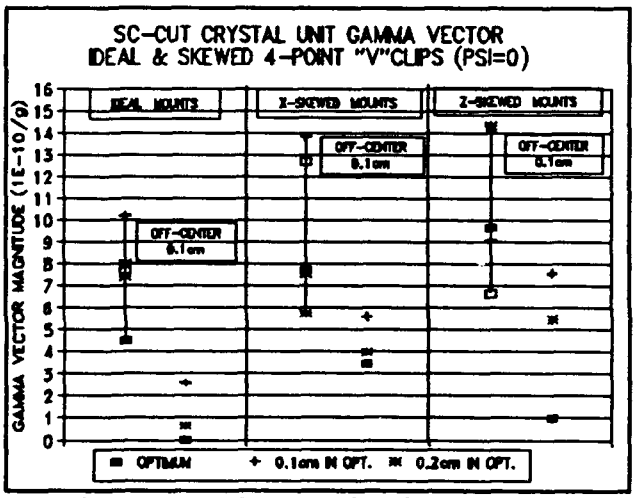


Figure 13. SC-Cut Crystal Unit Acceleration Sensitivity for Perfect and Imperfect Mounts

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SAW OSCILLATORS WITH LOW VIBRATION SENSITIVITY

T. E. Parker, J. A. Greer, and G. K. Montress

Raytheon Research Division
131 Spring Street
Lexington, MA 02173 USA

ABSTRACT

Surface Acoustic Wave (SAW) oscillators with vibration sensitivities well below $1x10^{-9}$ fractional change in frequency per g have been demonstrated. Data is presented which indicates the need for a rigid support structure under an inherently low vibration sensitivity "All Quartz Package" (AQP) SAW device in order to achieve low vibration sensitivity at the oscillator level. An alumina stiffener supporting a hybrid circuit SAW oscillator provides this essential rigidity. Vibration sensitivity data for twenty-six oscillators in a 1.22 inch (3.10 cm) square hybrid circuit package mounted on a 0.30 inch (7.6 mm) thick alumina stiffener yielded an average value for the magnitude of the vibration sensitivity vector, Γ, of $3.3x10^{-10}/g$ at a vibration frequency of 500 Hz. It also appears that some degree of vibration isolation at the AQP SAW device level may be achieved with the proper choice of material used to mount the AQP SAW device into the hybrid circuit package.

I. Introduction

It has been shown both theoretically [1]-[3] and experimentally [4], [5] that quartz Surface Acoustic Wave (SAW) devices can exhibit acceleration (vibration) sensitivities well below $1x10^{-9}$ fractional change in frequency per g. A uniformly supported SAW device on a rigid surface, as shown in Fig. 1(a), is theoretically capable of exhibiting an acceleration sensitivity in the low $10^{-11}/g$ range for acceleration in the direction normal to the plane of the SAW substrate [1]. However, the arrangement in Fig. 1(a) is not practical in many cases since the anisotropic thermal expansion of quartz requires that the uniform support be provided by a soft material which will serve to relieve thermally induced stresses. Unfortunately, all known practical soft support materials also outgas significant amounts of material that will contaminate the exposed surface of the SAW device and result in excessive long-term frequency drift. Therefore, this is not an acceptable approach for high performance oscillator applications that require a long-term frequency stability in the ±1 ppm/year range.

A better approach is to use the "All Quartz Package" (AQP) [6], as shown in Fig. 1(b), which has exhibited excellent long-term frequency stability [5]. Here, the active SAW surface is safely sealed between the two identical pieces of quartz by the glass frit. Any suitable soft material can be used to support the AQP since outgassing is no longer a problem. In practice it has been found that a lower vibration sensitivity can be obtained if the AQP is mounted ***upside down*** [4], as shown in Fig. 1(b). This has been explained by the fact that bending stresses for two perpendicular in-plane directions cause opposite frequency shifts for vibration normal to the plane of the SAW substrate [2]. Theoretical predictions

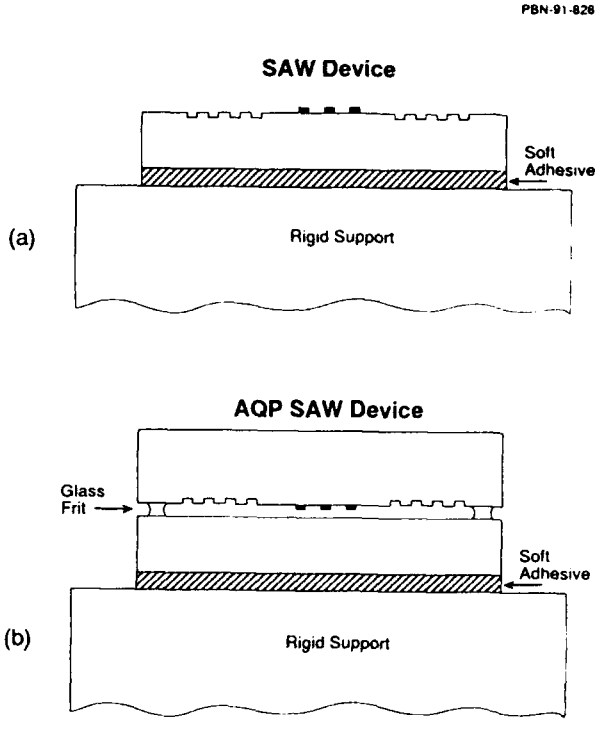


Figure 1. (a) Unpackaged SAW device with uniform soft mounting material on a rigid surface. (b) An AQP SAW device mounted ***upside down*** with a uniform soft mounting material on a rigid surface.

CH2965-2/91/0000-321 $1.00 © 1991 IEEE

indicate that values for γ_1 (acceleration normal to the plane of the SAW substrate) in the low 10^{-11} fractional change in frequency per *g* range should be obtainable for a proper frit geometry [2]. For acceleration in either of the two in-plane directions (γ_2 is parallel to the transducer fingers and γ_3 is parallel to the direction of acoustic propagation) the sensitivity should also be in the low $10^{-11}/g$ range [3]. Once again, a rigid support surface is assumed, along with a means to uniformly transmit acceleration forces to the SAW device. Bending will occur in the support structure if it is not sufficiently rigid, and these bending forces will be transmitted to the SAW device. This will increase the device's vibration sensitivity to levels above the theoretically predicted values.

Our objective was not only to develop SAW devices with low vibration sensitivity, but also to demonstrate high performance SAW oscillators with low vibration sensitivity. This low vibration sensitivity is very important if low noise oscillators are to operate satisfactorily in high vibration environments. Our approach to obtaining low vibration sensitivity SAW oscillators was to use an AQP SAW device and hybrid circuit technology. The small size and the available flat surface on the base of a hybrid circuit package makes obtaining a rigid structure much easier. Figure 2 shows the basic structure of our low vibration sensitivity hybrid circuit SAW oscillator. In this paper we will not attempt to address the topic of vibration isolation at the oscillator level.

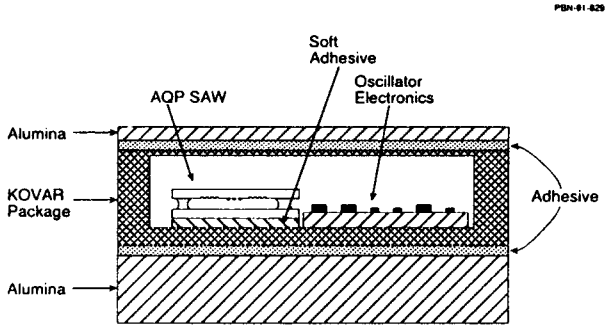


Figure 2. Cross-sectional diagram of a hybrid circuit AQP SAW oscillator with low vibration sensitivity.

The oscillator electronics consist of conventional hybrid circuit components on a 0.020 inch (0.51 mm) thick alumina substrate in a feedback loop oscillator configuration [7]. Careful selection and mounting of the electronic circuit components is required in order to minimize their potential contributions to the oscillator's overall vibration sensitivity. The AQP SAW resonator is mounted adjacent to the oscillator circuitry with a soft adhesive and both are contained in a KOVAR package. Alumina stiffeners are used on both the top and bottom of the KOVAR package in order to prevent low frequency resonances and to provide a sufficiently rigid structure. The bottom stiffener is much thicker than the top stiffener.

The results of measurements on more than thirty oscillators in the 700 MHz to 1 GHz range are discussed in this paper. More detailed oscillator dimensions will be given as specific examples are presented.

II. Measurement Procedures

Most of our evaluations have been performed at the oscillator level since our interest was in low vibration sensitivity SAW oscillators. Furthermore, SAW devices with relatively low group delays (less than two microseconds) are difficult to measure at the device level since flexing of the RF cables leading up to the shaker table causes interfering phase shifts. It has proven very difficult to get accurate vibration sensitivity data in the low $10^{-10}/g$ range on short delay devices with measurements performed on just the SAW device. In an oscillator circuit the cable flexing problem is significantly reduced since there is only one RF cable leading from the shaker table and its varying impedance can be isolated from the oscillator loop by the use of buffer amplifiers and a power divider or coupler. Therefore, even when our primary interest was only certain aspects of the AQP SAW device, the measurements were generally made with the SAW device mounted in a hybrid oscillator circuit.

All measurements of vibration sensitivity were made with sinusoidal acceleration at a peak level of about 3 *gs* on a small shaker table, as shown in Fig. 3. The configuration pictured here is for measuring γ_1 (vibration sensitivity normal to the plane of the SAW substrate). The other two vibration sensitivity components are measured by rotating the 2.5 inch (6.35 cm) aluminum cube on which the oscillator is mounted. The data collection system is computer controlled and automatically makes measurements at 261 vibration frequencies from 50 Hz to 7 kHz. The vibration induced FM sideband level is measured with custom software on a Hewlett-Packard 3047A Phase Noise Measurement System using a 500 nsec coaxial cable delay line as part of a discriminator circuit. Though the discriminator technique is not the most sensitive phase noise measurement system, it does require only one oscillator and was more than adequate for this purpose. The system is capable of measuring vibration sensitivities as low as 1×10^{-10} per *g* at 50 Hz and the system floor gradually drops to about 3×10^{-12} per *g* at 7 kHz. A small accelerometer is located next to the oscillator and its output is passed through a signal conditioning amplifier and measured by a digital voltmeter. The Hewlett-Packard 3582A Spectrum Analyzer in the 3047A Phase Noise Measurement System is capable of making relative phase measurements and this feature is also used to provide relative phase data between the output of the accelerometer and the output of the FM discriminator. This phase data is very useful for evaluating phase lag between the SAW device and the motion of the shaker table. Figure 4 shows a block diagram of the measurement system.

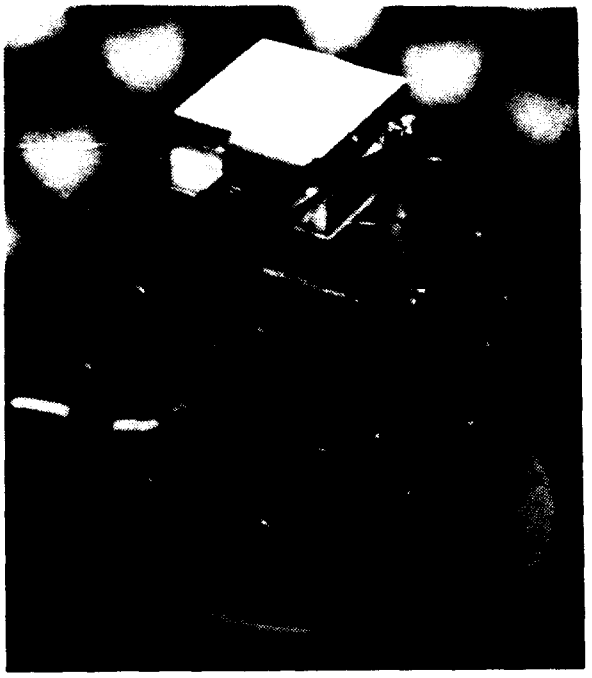


Figure 3. Hybrid circuit AQP SAW oscillator with alumina stiffener mounted on the shaker table. The aluminum cube permits vibration sensitivity measurements for three mutually perpendicular axes.

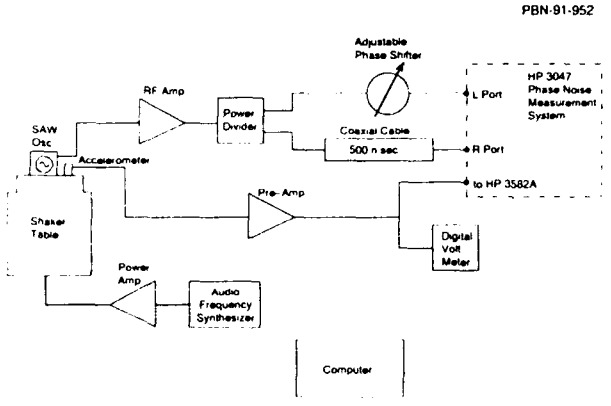


Figure 4. Block diagram of system used to measure the AQP SAW oscillators' vibration sensitivities.

III. Mechanical Design of the SAW Oscillator

The need for a stiffener underneath the hybrid circuit package is illustrated by the data in Fig. 5. Here the value for γ_1 (vibration in a direction normal to the plane of the SAW substrate) is plotted as a function of vibration frequency for three different stiffener thicknesses. In these measurements the alumina stiffener is supported only under its four corners as it sits on the cube shown in Fig. 3, and is therefore free to flex in the middle. A curve for uniform support under the entire stiffener is also shown for comparison. The four corner mount simulates one of the least rigid mounting geometries that might be encountered in actual use, whereas uniform support on the aluminum cube represents the most rigid mount available. These measurements were made on a 2.0 x 2.0 inch (5.1 x 5.1 cm) square hybrid circuit SAW oscillator operating at 700 MHz. No cover was used on the hybrid circuit. The measurements shown in Fig. 5 were made before the automated measurement system was available so there are only approximately 22 data points in each curve.

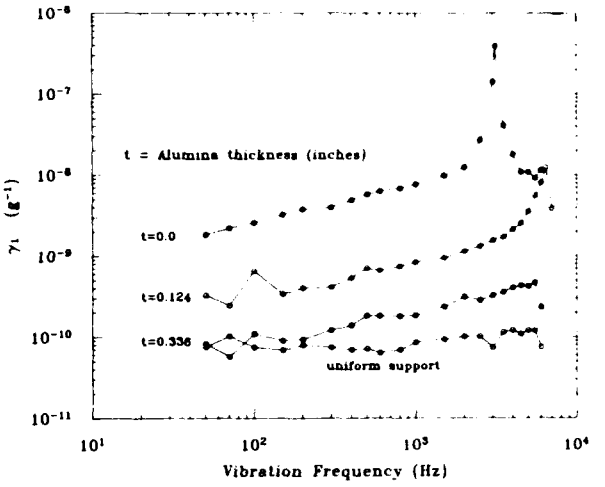


Figure 5. Measured values of γ_1 as a function of vibration frequency for three different stiffener thicknesses using a four corner mount. Data for a uniform mount is also shown.

The upper curve (t=0.0 inches) was made with no alumina stiffener and a resonance in the 0.06 inch (1.5 mm) thick bottom of the KOVAR package is clearly seen at 3.1 kHz. At the resonant frequency the vibration sensitivity peaks near $1x10^{-6}$ per g. Also note that the value of γ_1 is increasing with vibration frequency for vibration frequencies well below the resonance frequency. With a 0.124 inch (3.15 mm) thick alumina stiffener the overall vibration sensitivity is significantly reduced and the resonance frequency is pushed up to almost 7 kHz. The vibration sensitivity at the peak of the resonance is now only slightly greater than $1x10^{-8}$ per g. For t= 0.336 inches (8.53 mm) the resonance is no longer visible and the overall sensitivity is now well below $1x10^{-9}/g$. The curve for uniform support shows that this oscillator is capable of $1x10^{-10}/g$ under ideal conditions.

Figure 6 shows a plot of γ_1 at a vibration frequency of 1 kHz for the oscillator in Fig. 5 for five different stiffener thicknesses, and for the uniform support (the data point located at 2 inches). The improvement with increased stiffener

thickness is quite dramatic, and the vibration sensitivity has nearly reached its minimum value for a stiffener thickness of 0.448 inches (1.14 cm). Alumina is a very good choice for the stiffener material since it is significantly stiffer than KOVAR, yet is a good match to KOVAR in thermal expansion. The latter attribute is important because it makes fastening the alumina to the KOVAR package a relatively simple task.

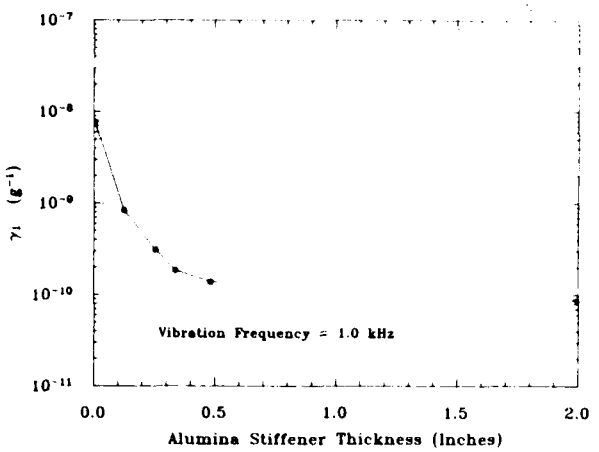


Figure 6. Measured values of γ_1 at 1.0 kHz vibration frequency as a function of stiffener thickness. Uniform support is represented by the data point at 2.0 inches.

The deflection at the center of the square stiffener due to its own mass under acceleration is approximately proportional to the length of a side to the fourth power and inversely proportional to the thickness squared. The data in Fig. 6 does not decrease quite as fast as $1/t^2$, but the vibration sensitivity does depend strongly on the stiffener thickness. Also, the fourth power dependence on the length or width of the stiffener means that a smaller hybrid circuit package can use a significantly thinner stiffener for a given vibration sensitivity level, whereas a larger hybrid circuit package would require a thicker stiffener.

The oscillator package design used for most of the measurements that will be discussed in this paper is shown in Fig. 7. The KOVAR package is 1.22 inches (3.10 cm) square and has a wall thickness of 0.06 inches (1.52 mm). The base of the package is 0.075 inches (1.91 mm) thick and the overall height is 0.273 inches (6.93 mm). A field replaceable SMA connector is mounted on one side on a 0.080 inch (2.0 mm) thick boss-out. The package can be hermetically sealed with a welded KOVAR cover of 0.020 inch (0.51 mm) thickness. The KOVAR package is mounted on a 0.30 inch (7.6 mm) thick alumina stiffener and a 0.015 inch (0.38 mm) thick alumina stiffener is used on the cover. This package and the two alumina stiffeners were shown in Fig. 3. Standard adhesives such as epoxies or thermoplastics can be used to attach the alumina stiffeners to the KOVAR package. The SAW resonator is contained in a 0.40 inch (1.0 cm) by 0.50 inch (1.27 cm) AQP of 0.072 inch (1.83 mm) thickness which is mounted ***upside down*** directly on the base of the KOVAR package. In most cases the AQP SAW device was fastened to the hybrid circuit package with a .010 inch (0.25 mm) thick piece of an acrylic sheet adhesive manufactured by 3M Company. The sheet adhesive is intended to provided a uniform, but pliable support under the AQP SAW device.

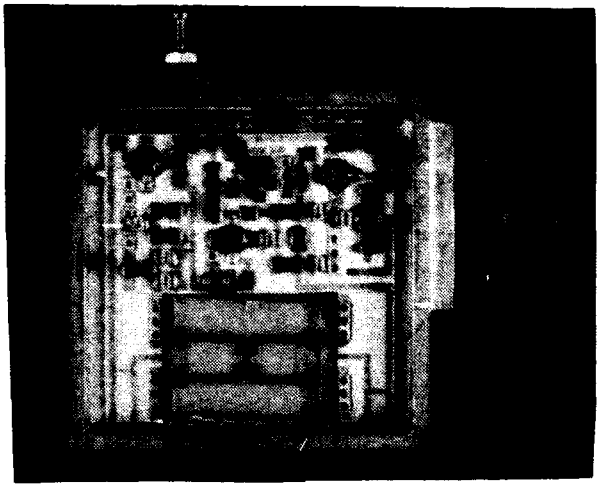


Figure 7. Photograph of the 1.22 inch square hybrid circuit AQP SAW oscillator used for most of the vibration sensitivity measurements.

IV. Measurement Results

Figure 8 shows the results of a typical measurement of γ_1 versus vibration frequency for the 1.22 inch (3.10 cm) square oscillator with a uniform mount under the stiffener. The AQP SAW was mounted with a uniform layer of acrylic sheet adhesive. The peak near 6 kHz is caused by a cover resonance. With no cover stiffener this resonance drops to about 2.5 kHz, and with a .025 inch (0.64 mm) thick alumina cover stiffener it is above 7 kHz. The smaller peak near 3 kHz is from a resonance caused by the RF connector. At low vibration frequencies cable flexing can cause interference and this is at least partially responsible for the ragged nature of the data below about 200 Hz. The overall vibration sensitivity level is in the mid to low $10^{-10}/g$ range, but there is a significant decrease in the value of γ_1 with increasing vibration frequency. The reason for this tilt is discussed below.

Figure 9 shows the vibration sensitivity for all three axes for the same device as in Fig. 8. The magnitude of the vibration sensitivity vector, Γ, is also shown as the solid curve in Fig. 9. In general γ_1 is the largest component, with γ_2 and γ_3 usually falling in the low 10^{-10} to upper $10^{-11}/g$ range.

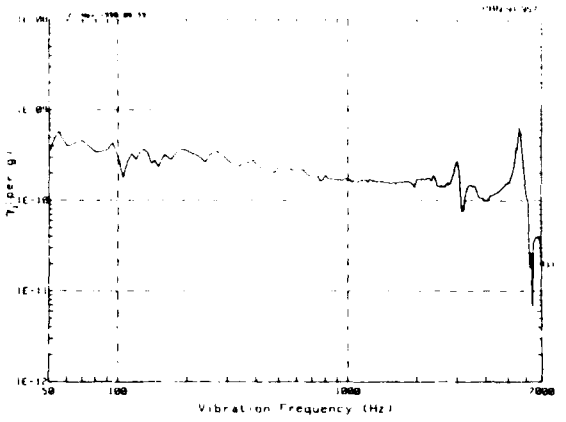


Figure 8. Typical plot of γ_1 versus vibration frequency for the 1.22 inch square hybrid circuit AQP SAW oscillator with uniform mount. The AQP SAW device was uniformly supported with acrylic sheet adhesive.

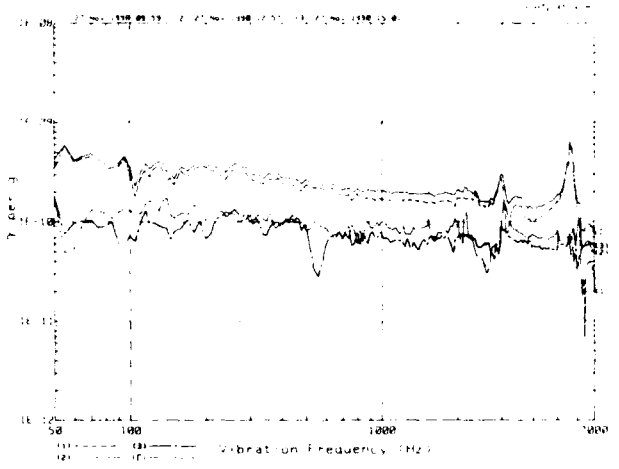


Figure 9. Vibration sensitivity for all three axes (γ_1, γ_2, and γ_3), and the magnitude of the vibration sensitivity vector, Γ, for the same oscillator as was used in Fig. 8.

Figure 10 shows the relative phase shift between the discriminator output and the output of the accelerometer as a function of vibration frequency for the γ_1 measurement on the same device used in Figs. 8 and 9. If the SAW device were moving in unison with the shaker table, one would expect the phase shift to be either 0 or 180°, depending on whether the resonator frequency shifts up or down for a given stress pattern. However, it is clear from the data in Fig. 10 that there is a significant deviation from 180°. The rapid phase shifts near 3 and 6 kHz are caused by the resonances mentioned earlier, but the general decrease in phase from about 160° to approximately 110° was a puzzle at first. However, the data in Figs. 11 and 12 has helped to clarify the situation.

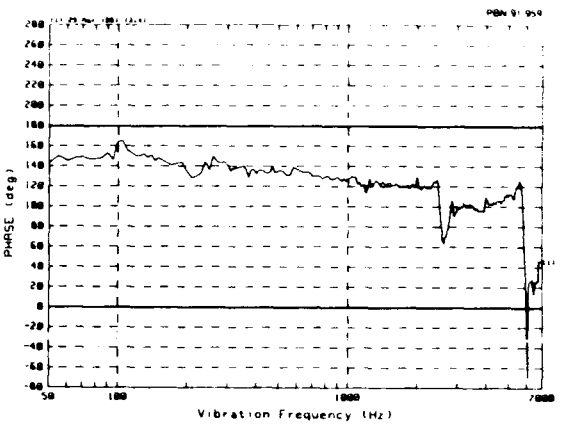


Figure 10. Relative phase shift between γ_1 and the output of the accelerometer mounted on the shaker table, as a function of vibration frequency, for the same oscillator as used in Figs. 8 and 9.

Figure 11 shows γ_1 versus vibration frequency for a 1.22 inch square hybrid circuit AQP SAW oscillator similar to the one used to generate the data in Figs. 8, 9, and 10. This oscillator was uniformly supported on the aluminum cube in the same fashion as the other oscillator, but the AQP SAW device was mounted in the KOVAR package with a uniform layer of silicone rubber rather than the acrylic sheet adhesive. Figure 12 shows the relative phase for γ_1. Two characteristics of the data shown in Figs. 11 and 12 are very different from that shown in Figs. 8 and 10. In Fig. 11 the value of γ_1 is essentially independent of vibration frequency, although the cover resonance is still present, and a very small RF connector resonance is also visible around 2 kHz. The raggedness at low frequencies is also evident due to the cable flexing, but the overall downward tilt of Fig. 8 is not present. In Fig. 12 the phase is now almost exactly 180°, except in the vicinity of the resonances and at frequencies above about 3 kHz. (The latter is caused by a phase distortion in the FM measurement system.) The tilt and phase shifts evident in Figs. 8 and 10, respectively, are clearly dependent on the material used to mount the SAW device.

Our preliminary interpretation of the data in Figs. 8 and 10, in comparison to the data in Figs. 11 and 12, is that the SAW device mounted on the acrylic sheet adhesive (which is very soft and pliable) exhibits a highly damped mechanical resonance. The tilt in the vibration sensitivity and the decrease in phase that are evident in Figs. 8 and 10 appear to result from the onset of a significant level of vibration isolation caused by the damped resonance of the mass of the SAW device on the soft adhesive. This phenomena is not evident in Figs. 11 and 12 because the silicone rubber is harder and less lossy than the acrylic adhesive. Thus, the silicone rubber provides no isolation between the SAW device and the KOVAR package. The somewhat higher value of γ_1 in Fig. 11, as

compared to that in Fig 9 (even at the lowest frequency), may be at least partially due to the fact that the silicone rubber was not uniformly distributed under the entire SAW device, and that a small area in one corner was not supported. An investigation of different mounting materials is currently underway in an attempt to better understand their influence on AQP SAW device vibration sensitivity.

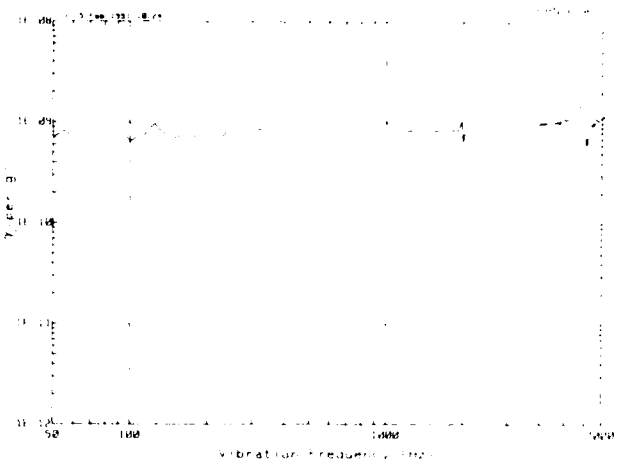


Figure 11. Vibration sensitivity, γ_1, versus vibration frequency for an oscillator with a uniform mount, but in which the AQP SAW device was uniformly supported with silicone rubber.

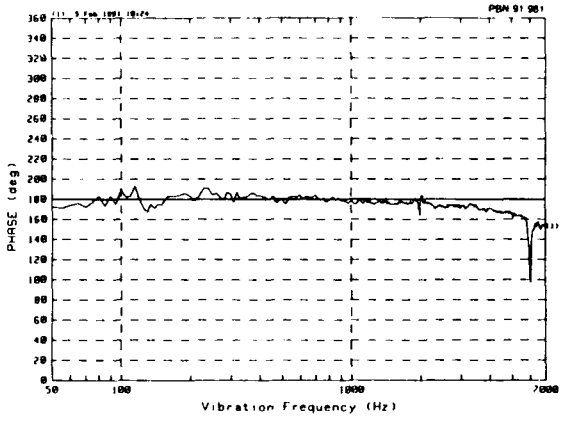


Figure 12. Relative phase shift between γ_1 and the output of the accelerometer mounted on the shaker table as a function of vibration frequency, for the same oscillator as used in Fig. 11.

Table I summarizes the measured vibration sensitivities obtained from twenty-six hybrid circuit SAW oscillators. All were in 1.22 inch square KOVAR packages and the AQP SAW resonators were all mounted with the .010 inch thick acrylic sheet adhesive. The average value for the magnitude of the vibration sensitivity vector, Γ, at a vibration frequency of 500 Hz was 3.3×10^{-10} fractional change in frequency per g. The standard deviation was 1.4×10^{-10} per g, with maximum and minimum values of 6.2×10^{-10} per g and 1.2×10^{-10} per g, respectively. About 75% of the oscillators exhibited a vibration sensitivity with a downward tilt similar to that present in Fig. 9. Therefore, the average value of Γ at vibration frequencies below 500 Hz is somewhat higher than $3.3 \times 10^{-10}/g$, and somewhat lower for vibration frequencies above 500 Hz. All of these oscillators were mounted with uniform support on the aluminum cube, and therefore the measured vibration sensitivities represent the best achievable values. At this time it is not yet clear whether the observed oscillator-to-oscillator variation in vibration sensitivity orginates in the AQP SAW, the acrylic adhesive mount, the oscillator electronics and housing, or possibly some combination of all of the above. However, remounting entire oscillators on the ceramic and on the aluminum cube has shown good reproducibility. The adhesives used to mount the KOVAR package to the ceramic stiffener and the stiffener to the aluminum cube are relatively hard materials and no evidence of resonances has been observed in these mounts.

TABLE I
Summary of Measurement Results

- Twenty-six Hybrid Circuit SAW Oscillators Tested
- AQP SAW Mounted with 0.010 inch thick Acrylic Sheet Adhesive
- Uniform Support under Oscillator Package
  - Average Γ (at 500 Hz) = $3.3 \times 10^{-10}/g$
  - Standard Deviation = $1.4 \times 10^{-10}/g$
  - Maximum Γ = $6.2 \times 10^{-10}/g$
  - Minimum Γ = $1.2 \times 10^{-10}/g$

In order to determine how much degradation in vibration sensitivity would occur in a less than ideal mounting situation, four 1.22 inch square oscillators were measured with a four corner mount on the aluminum cube rather than the uniform mount. The two mounting configurations are illustrated in Figs. 13(a) and 13(b). The 0.30 inch (7.6 mm) thick alumina stiffener was used under the KOVAR package. Figure 14 shows γ_1 for both the uniform mount (light line) and the four corner mount (heavy line) for one of these oscillators. There is clearly some degradation with the four corner mount, but γ_1 is still in the low $10^{-10}/g$ range. Measurements on the other oscillators gave similar results, indicating that the four corner mount causes γ_1 to increase by roughly a factor of two. This increase is most pronounced at higher vibration frequencies. Measurements of γ_2 and γ_3 have not as yet been made with a four corner mount.

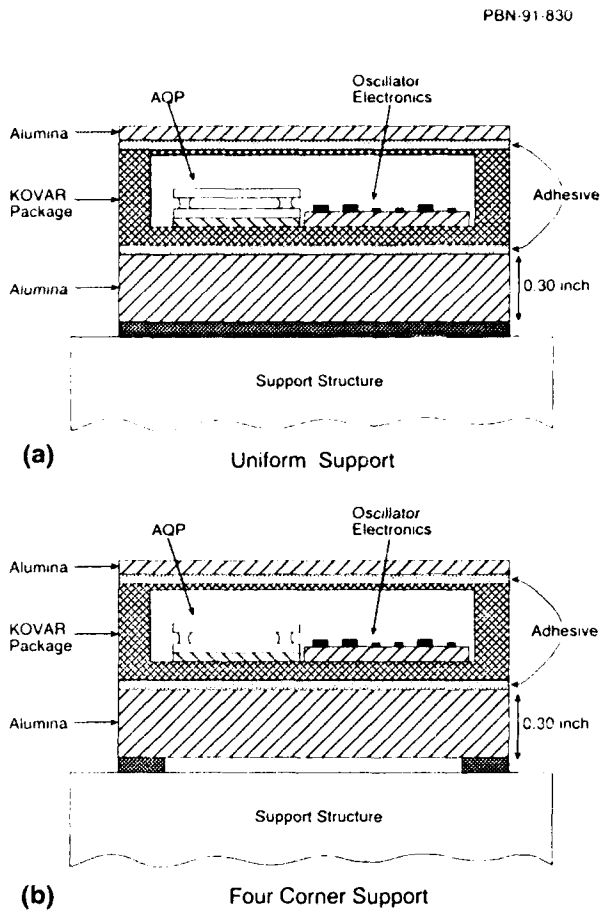


Figure 13. Cross-sectional diagrams of oscillators with: (a) uniform mount, and (b) four corner mount.

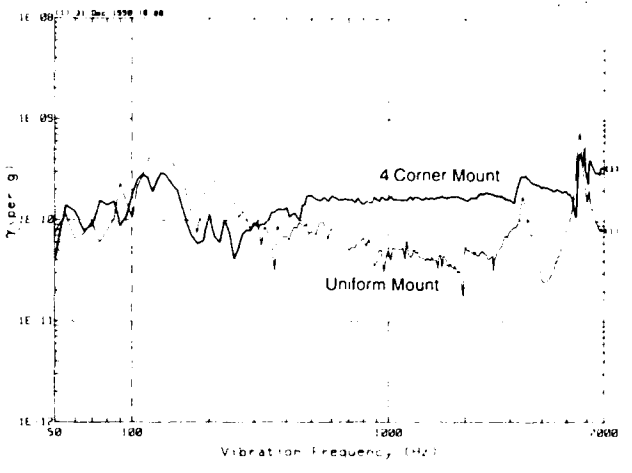


Figure 14. Comparison of γ_1 versus vibration frequency for an oscillator with either uniform mount or four corner mount.

It has been shown in the literature that the vibration sensitivities of both SAW and BAW (bulk acoustic wave) oscillators can exhibit a dependence on the oscillator frequency, if the operating frequency of the oscillator is adjusted relative to the frequency of the quartz resonator [8]. To determine if this was occurring in our devices, the operating frequency of several oscillators was varied by changing the phase shift inside the oscillator loop. Figure 15 shows γ_1 (at a vibration frequency of 1 kHz) as a function of oscillator frequency for one of the oscillators. The oscillator frequency was varied over the entire 1 dB bandwidth of the SAW resonator and it is clear that there is no significant dependence on oscillator frequency. Similar data at other vibration frequencies and on other oscillators gave the same results.

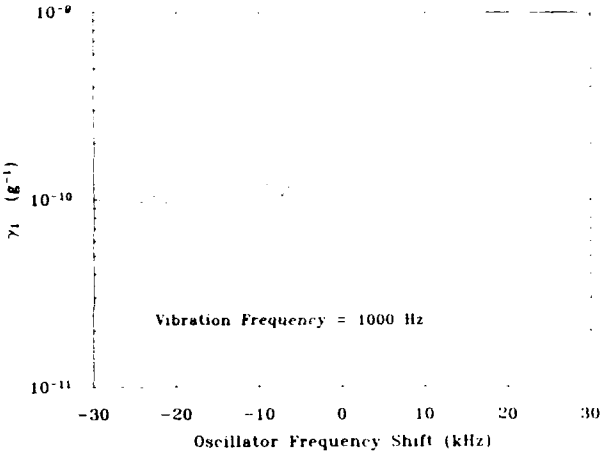


Figure 15. Vibration sensitivity, γ_1, versus oscillator frequency. The oscillator frequency was varied by adjusting the phase shifter in the feedback loop.

Finally, a promising direction for future development is the use of smaller AQP SAW devices. Figure 16 shows a comparison between the standard 0.4 inch (1.02 cm) by 0.5 inch (1.27 cm) AQP used for most of the SAW oscillators discussed so far and a new smaller AQP of 0.16 inch (4.1 mm) by 0.33 inch (8.5 mm) which contains the same SAW resonator design as the larger AQP. Both packages have the same overall thickness of 0.072 inches (1.8 mm). Figure 17 shows the magnitude of the vibration sensitivity vector, Γ, versus vibration frequency for an oscillator using one of these devices. Except for cable flexing problems at low vibration frequencies, the vibration sensitivity of this device is essentially at the $1x10^{-10}/g$ level. Similar results have been obtained several times, indicating that a factor of two to three improvement over the current results with the larger AQP may be possible.

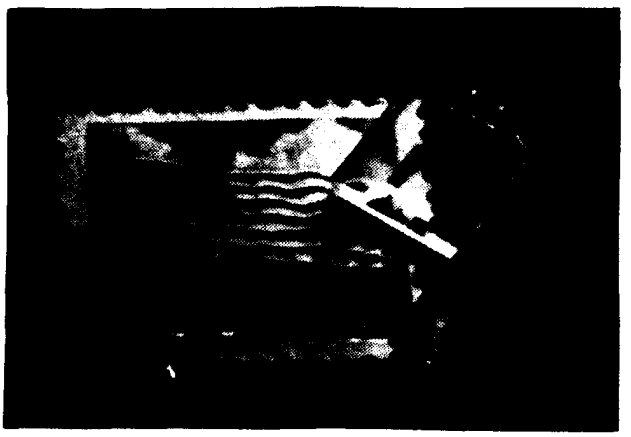


Figure 16. Photograph showing size comparison between a standard AQP SAW resonator device and the new smaller AQP SAW device.

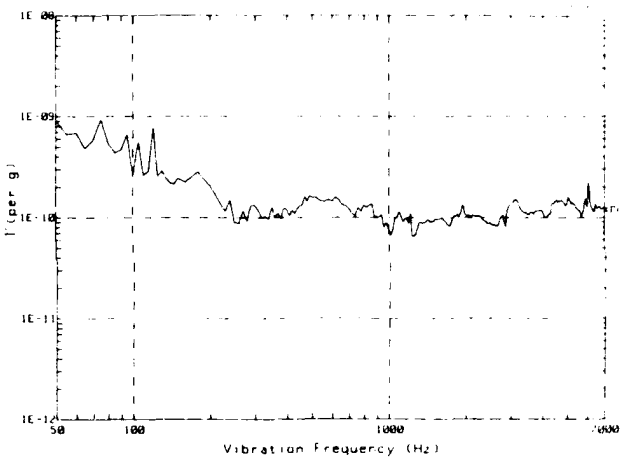


Figure 17. Magnitude of the vibration sensitivity vector, Γ, versus vibration frequency for a hybrid circuit AQP SAW oscillator using the new smaller AQP SAW device shown in Fig. 16.

V. Summary and Conclusions

The data presented in this paper demonstrates that SAW oscillators with vibration sensitivities well below $1x10^{-9}$ per *g* can be reliably fabricated if a sufficiently rigid mounting structure is used. It has also been found that the adhesive on which the AQP SAW is mounted can play a significant role in the vibration sensitivity of the oscillator. However, the details of this process are not yet fully understood. The observed levels of vibration sensitivity are still about an order of magnitude higher than theoretically predicted values, and it has not yet been established why this is so. Several possible reasons are variations in the frit geometry, uncontrolled nonuniformities in the material used to support the AQP, as well as other unidentified sources of vibration sensitive phase shifts in the oscillator circuitry. Methods to achieve improved device performance continue to be investigated, including the use of smaller AQPs and more suitable mounting materials.

In addition to achieving further reductions in the vibration sensitivity of SAW oscillators, other characteristics such as size, weight, producibility and long-term frequency stability must also be considered. Intentionally nonuniform supports for the AQP can yield low vibration sensitivities and may even improve reproducibility. Also, optimizing nonuniform supports for the entire oscillator may permit a reduced thickness for the alumina stiffener. Furthermore, it may be possible to reduce the overall size of the stiffened oscillator by perfecting a balance between a stiffened oscillator package and a more rigid AQP. Even with all of the requirements for low vibration sensitivity, it is still necessary that all of the materials and components be compatible with good long-term frequency stability. For the material under the AQP it is particularly important that it not cause slowly changing static stresses on the SAW device. In some circumstances this has been a problem with the acrylic sheet adhesive, and aging studies are currently underway. The process of obtaining high performance hybrid circuit AQP SAW oscillators with the lowest possible vibration sensitivity has proven to be very complex, and it is now increasingly clear that the investigation has just begun.

References

[1] H. F. Tiersten, D. S. Stevens, and R. D. Weglein, "Acceleration Sensitivity of Surface Acoustic Wave Resonators," *Applied Physics Letters*, Vol. 48, No. 2, pp. 97-99, 13 January 1986.

[2] H. F. Tiersten and D. V. Shick, "An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Acoustic Wave Resonators with Substrate Extending Beyond the Supports," in *Proceedings of the 2nd European Frequency and Time Forum*, EFTF 88, p. 753.

[3] H. F. Tiersten and D. V. Shick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Acoustic Wave Resonators with Interior Rectangular Supports," in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 396-404, IEEE Catalog Number 89CH2690-6.

[4] J. A. Greer and T. E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator," in *Proceedings of the 42nd Annual Symposium on Frequency Control*, 1988, pp. 239-251, IEEE Catalog Number 88CH2588-2.

[5] T. E. Parker and G. K. Montress, "Frequency Stability of High Performance SAW Oscillators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Vol. 1, pp. 37-45, IEEE Catalog Number 89CH2791-2.

[6] T. E. Parker, J. Callerame, and G. K. Montress, "A New All Quartz Package for SAW Devices," in *Proceedings of the 39th Annual Symposium on Frequency Control*, 1985, pp. 519-525, IEEE Catalog Number 85CH2186-5.

[7] T. E. Parker and G. K. Montress, "Precision Surface Acoustic Wave Oscillators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 3, pp. 342-364, May 1988.

[8] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, and R. McGowan, "Electronic Desensitization of Resonators to Accelerations," in *Proceedings of the 44th Annual Symposium on Frequency Control*, 1990, pp. 444-451, IEEE Catalog Number 90CH2818-3.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A LOW NOISE VIBRATION ISOLATED AIRBORNE RADAR SYNTHESIZER

A. Vulcan and M. Bloch

Frequency Electronics, Inc.
Mitchel Field, NY 11553

ABSTRACT

A Ku-Band synthesizer is described which incorporates both direct and indirect techniques to generate exceptionally pure signals with low phase jitter and spurious content. Fast switching speed and phase noise of -120 dBc/Hz from 10 kHz to 40 MHz is obtained. The radar operates in a fighter aircraft environment that has a high level of random vibration and static G loading, and the mechanical design incorporates vibration isolation techniques. The master oscillator, from which the various fixed and steppable frequency outputs are derived, uses a low G sensitivity SC cut crystal. Ku-Band phase noise performance data is presented and the correlation between anticipated and measured performance is discussed.

PERFORMANCE CHARACTERISTICS

The synthesizer design emphasizes low phase noise performance over severe environmental conditions while maintaining electrical characteristics which are compatible with high resolution airborne multi-mode radar missions. The performance requirements are below:

- Outputs
  - Agile Ku-Band, Fixed L-Band, and Fixed VHF
- Tuning Range
  - 750 MHz
- Step Size
  - 1 MHz or 50 MHz
- Switching Speed
  - 1 Microsecond
- Warmup
  - 2 Minutes to ±1PPM Accuracy
- Spurious Signals
  - -60 dBc
- Ku-Band Phase Noise During Vibration
  - -120 dBc/Hz at 30 kHz
  - -60 dBc/Hz at 100 Hz
- Temperature Range
  - -55°C to +75°C

The major design features of the synthesizer are:

- Direct Synthesis
  - High Speed
- Dual Conversion
  - Low Spurious Signals
- Hybrid RF Subassemblies
  - Small Size, Low Weight
- Thick Film Hybridized Oscillator
  - Fast Warmup, Low Power
- Low Aging SAW VCO
  - High Stability FA Mode

The synthesizer reference source is an ovenized Butler oscillator, using an SC cut crystal. The oscillator/buffer amplifier and oven controllers are hybridized in order to minimize size and weight. The master oscillator is suspended in a mechanical isolation structure which attenuates low frequency vibration inputs while encountering 6 G static loads in any axis.

Direct synthesis is used to enhance speed and minimize system phase noise. A high degree of filtering is used within the synthesizer to keep the spurious signals at extremely low levels. VHF crystal filters, as well as helical filters at UHF frequencies, and cavity devices at Ku-Band are incorporated.

The size and weight of the synthesizer is minimized through the use of thick and thin film hybrid components

CH2965-2/91/0000-330 $1.00 © 1991 IEEE

and the use of integrated subassemblies to a maximum extent. The various functions are grouped and packaged in separate mechanical chassis to facilitate maintenance and reduce crosstalk.

ELECTRICAL DESIGN

Figure 1 is a Block Diagram of the Ku-Band Synthesizer. The master oscillator synthesizes a Ku-Band output using dual upconversion with internal signals derived from comb generators and switchable filter banks. The second mixer is fed from either a coherent fixed frequency signal or a multiplied SAW VCO output to provide fine frequency tuning steps. The vibration isolated master oscillator subassembly consists of the hybridized OCXO, ×3 multiplier, and crystal filter. The OCXO direct output drives a comb generator whose multiples from 18 to 23 are fed to a six channel switchable filter bank with spacings of approximately 50 MHz. Six-pole helical filters and high isolation pin diode switches ensure that adjacent comb lines are attenuated by 75 dB. The ×3 output from the master oscillator is crystal filtered so that the VHF output has a noise floor of -170 dBc/Hz. This signal is doubled, filtered, and fed to a step recovery diode comb generator, whose sixth through tenth harmonics are selected. The spacing of these lines is six times the OCXO frequency. This filter bank uses four-pole cavity filters to provide three switchable LO's from the eighth through tenth harmonic. The sixth harmonic is used for STALO 2 and the seventh feeds the second LO ×5 multiplier via a 1 × 2 diode switch. A third switchable filter bank consisting of three filters, each covering approximately 250 MHz, selects the appropriate first mixer output while rejecting the LO and lower sidebands. The seventh harmonic output from the filter bank is multiplied by five to generate the fixed local oscillator signal for the second mixer. The second LO can also be obtained from a multiplied digitally controlled SAW VCO. In this frequency agile mode, the synthesizer Ku-Band output is non-coherent since the 10 GHz second LO is derived from the SAW, which is not locked to the master oscillator. The VCO is tuned in thirty incremental steps of 25 kHz yielding channels with 1 MHz spacings covering ±15 MHz at the system output. The SAW and ×4 multiplier are mounted on a temperature controlled thermal plate, and a D to A converter and PROM select the appropriate frequency. The second mixer output is filtered with a seven-pole cavity filter and amplified to +20 dBm with a GaAs FET amplifier.

VIBRATION CONSIDERATIONS

The electrical and mechanical design of the synthesizer is optimized for low vibration induced phase noise. The following measures are taken to ensure that this goal is met.

- The vibratory phase noise performance must be determined only by the isolated master oscillator assembly whose characteristics are predictable and measurable. Consequently, the mechanical resonance and vibration susceptibility of the various circuit elements which cause random phase modulations must be reduced to levels which are below the multiplier enhanced master oscillator noise at that point in the system. This is accomplished by encapsulating critical circuits and using thick and thin film RF assemblies whose resonances are well above the vibration frequencies.

- The crystal oscillator is mounted on a three-axis spring dampened vibration isolator, shown in Figure 2, which reduces G levels at mission critical offset frequencies above 500 Hz. The anti-vibration mount design employs a system of looped, stainless steel, multi-strand cables to achieve a high level of isolation within a small space. Each end of the oscillator package is supported by a pair of flexed cable loops and a pair of cantilevered cable posts. Isolation is achieved in three mutually perpendicular axes either by flexure of the cable loops or by deflection of the cantilever posts. The isolator acts as a second order low pass mechanical filter with a roll-off of 10 to 15 dB per octave.

- An SC-cut crystal with G-sensitivity of less than 2.5×10^{-10}/G in the worst axis is used in the oscillator.

Since the phase noise under vibration is directly proportional to the crystal G sensitivity, every crystal manufactured for use in the synthesizer is subjected to a sine vibration sweep in the 3 orthogonal axes to characterize the devices. In the operational environment, the worst vibration is encountered in the aircraft vertical axis and consequently, the best crystal axis from the standpoint of minimum vibration sensitivity is aligned with this axis. This is done within the oscillator assembly during the alignment procedure. Each completed oscillator is again characterized for both sine and random vibration spectra response after alignment is completed. This hard mounted test of the oscillator ensures that no internal resonances exist and phase noise contributions due to factors besides the crystal are negligible.

The single sideband phase noise is related to frequency deviation Δf and modulation frequency f_m by:

$$L(f) = 20 \log \left(\frac{\Delta f}{2 f_m} \right) \text{ dBc/Hz} \qquad (1)$$

In the presence of random vibration γ (in g^2/Hz)

$$\Delta f = \sqrt{2\gamma} \times f_0 \times S$$

Where f_0 is oscillator frequency and S is oscillator vibration sensitivity in parts per G.

$$\text{Thus, } L(f) = 20 \log \left(\frac{\sqrt{2\gamma} \times f_0 \times S}{2 f_m} \right) \text{ dBc/Hz} \qquad (2)$$

In the synthesizer, vibration induced oscillator phase noise is enhanced by 50 dB when translating from the OCXO output to 14 GHz. Since L(f) during vibration is determined only by S of the crystal, a resonator selected for good static phase noise is not necessarily the best for a specific application where the vibration induced component is significant.

Figure 3 is the random vibration profile which is applied to the synthesizer. A 0.016 g^2/Hz peak occurs at 11 Hz with energy present up to 2 kHz.

Five crystals were selected for use in synthesizer master oscillators. The G sensitivity of each is itemized in Table 1 for each axis. The Y axis is perpendicular to the plane of the quartz resonator. The Z axis is parallel to the surface and runs through the active electrodes.

In order to evaluate a worst case scenario, crystal number 4 was used in the synthesizer for which data is presented.

TABLE 1. CRYSTAL G-SENSITIVITY $\times 10^{-10}$/G

UNIT NO.	X-AXIS	Y-AXIS	Z-AXIS	RSS
1	1.27	1.13	1.42	2.21
2	0.57	0.45	0.46	0.85
3	1.00	0.47	0.47	1.18
4	0.90	1.13	2.25	2.67
5	1.43	1.79	0.23	2.30

PERFORMANCE DURING VIBRATION

Figure 4 is the static phase noise measured at the Ku-Band output. The 60 Hz and 120 Hz spikes are AC line induced in the test equipment. The noise floor at 100 kHz offset is -124 dBc/Hz which is equivalent to -164 dBc/Hz at the crystal filter output of Figure 1. Figures 5, 6, and 7 show the Ku-Band phase noise in axes X, Y, and Z respectively with the system subjected to the random vibration spectrum of Figure 3. Overlayed on each set is the static L(f) and equation (2) plotted for $S = 3 \times 10^{-10}$/G. It is seen that for axis X and Y, the noise is significantly less than the predicted value since the actual value of S is less in these axes. The Z-axis phase noise of Figure 7 clearly shows the mount resonance at 120 Hz and the rolloff at higher frequencies. The static and dynamic curves coincide above 2 kHz indicating that vibration induced phase modulations are below the static levels at higher offsets, and that mechanical package resonances, which often generate harmonic noise, have little effect.

Based upon the described measurements the following conclusions are drawn:

	AXIS X	Y	Z
• MEASURED L(f) AT 20 Hz – dBc/Hz	-54	-50	-41
• CALCULATED CRYSTAL S – PARTS PER 10^{-10}	0.57	0.90	2.55
• MEASURED CRYSTAL S – PARTS PER 10^{-10}	0.90	1.13	2.25
• MOUNT RESONANCE – Hz	75	120	120
• MOUNT ROLL-OFF – dB/OCTAVE	9	14	10

The values of S are calculated from equation (2) using the measured L(f) values at 20 Hz and $\gamma = 0.016$ g^2/Hz. Considering that the phase noise with a flat random vibration input would decrease at 6 dB per octave, a rough calculation of the mount attenuation can be made by noting the roll-off between 150 and 300 Hz and subtracting 6 dB from this value.

Roll-off is seen to be in the 9 to 14 dB per octave range, which is within the expected range for the vibration isolator used in the synthesizer. Excellent correlation is seen between the values of S measured on the crystal resonator itself and that derived from Ku-Band system data.

MECHANICAL DESIGN

The synthesizer, which consists of seven separate modules, is packaged in a carrier measuring 8.0 inches long × 6.5 inches wide × 4.0 inches high and weighs 12 pounds. The carrier assembly, as well as the individual modules, are milled out of single aluminum blocks to ensure structural rigidity and minimize performance-degrading mech-

anical resonances. Figures 8 and 9 depict the configuration.

The synthesizer mounting surface in the radar is the thick metal plate on which the master oscillator is mounted. Thus, the random vibration spectrum is transmitted directly to the oscillator without any intervening structural members having unpredictable dynamic characteristics.

CONCLUSIONS

A Ku-Band synthesizer having low vibration induced phase noise is described. Electrical and mechanical design parameters are selected to optimize performance when the synthesizer is subjected to simultaneous static and dynamic G forces.

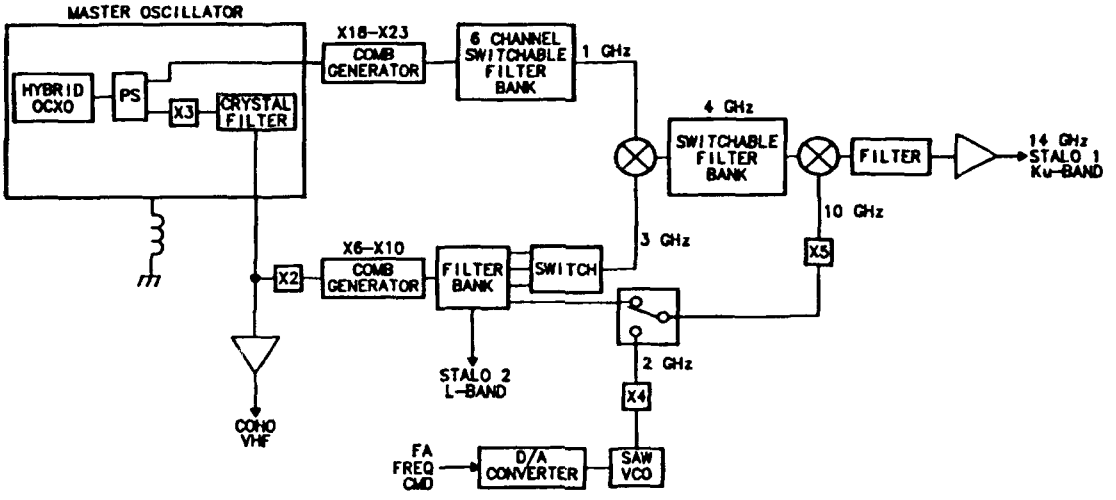


Figure 1. Block Diagram of Ku-Band Synthesizer

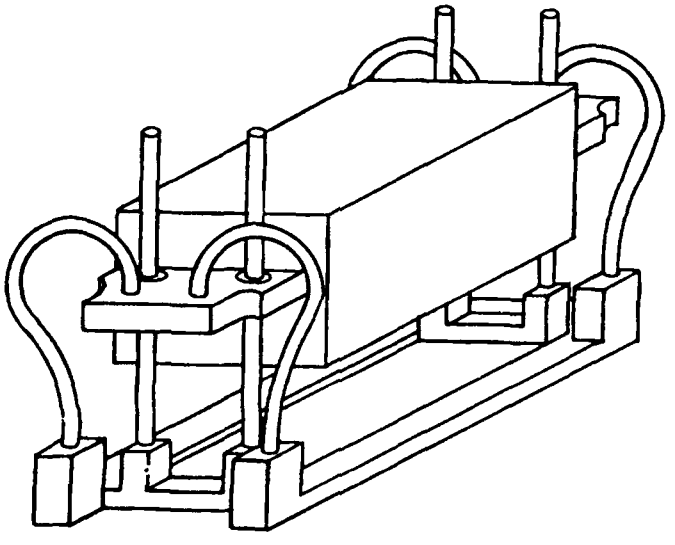


Figure 2. Vibration Isolated Oscillator

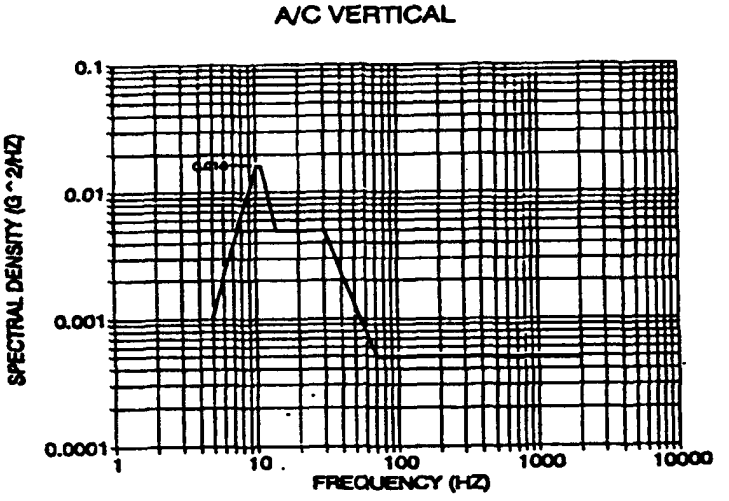


Figure 3. Performance Vibration Level

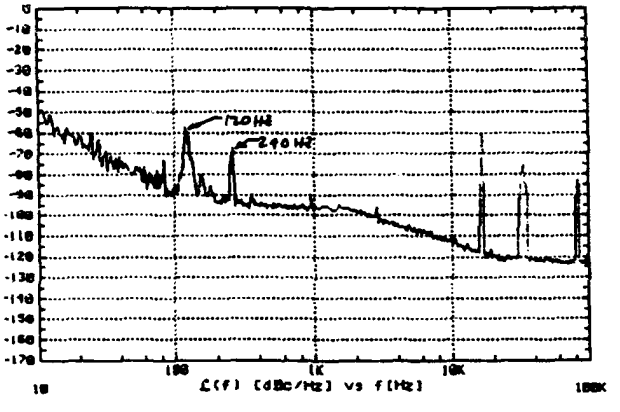


Figure 4. Static Phase Noise

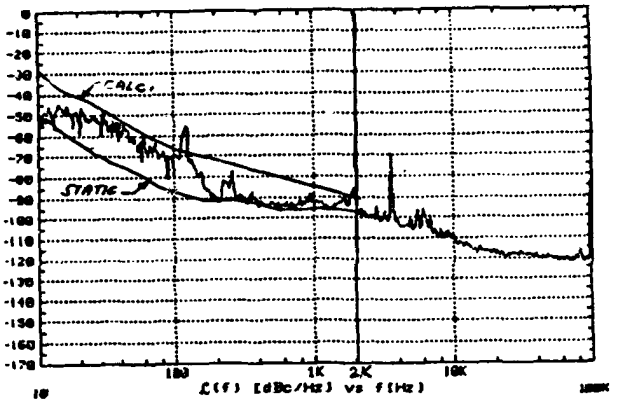


Figure 6. Y-Axis Phase Noise

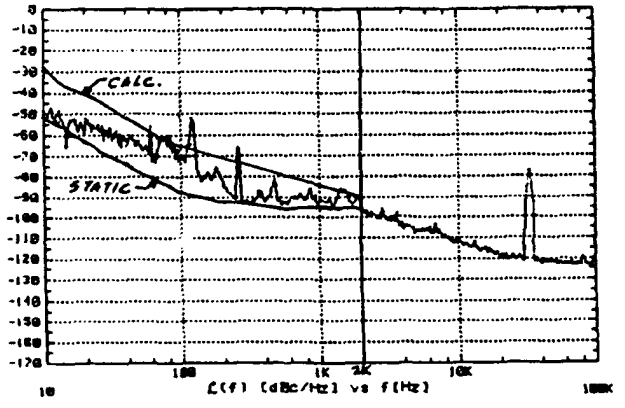


Figure 5. X-Axis Phase Noise

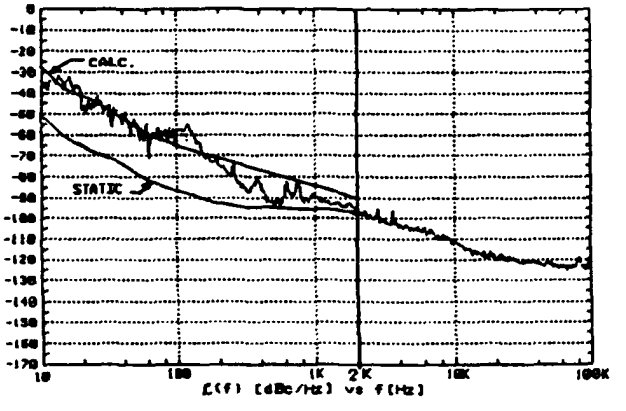


Figure 7. Z-Axis Phase Noise

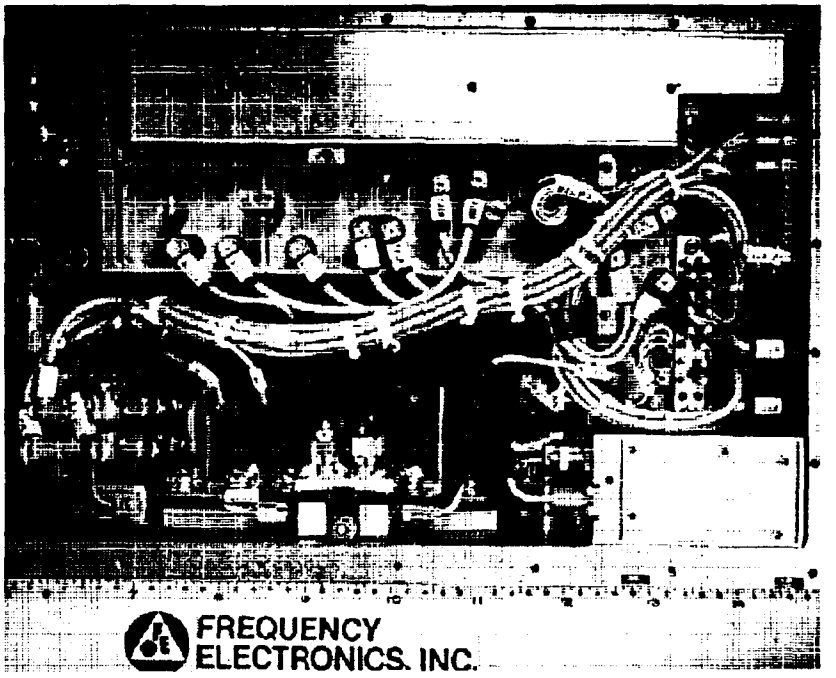


Figure 8. Assembly Configuration — Right Side View

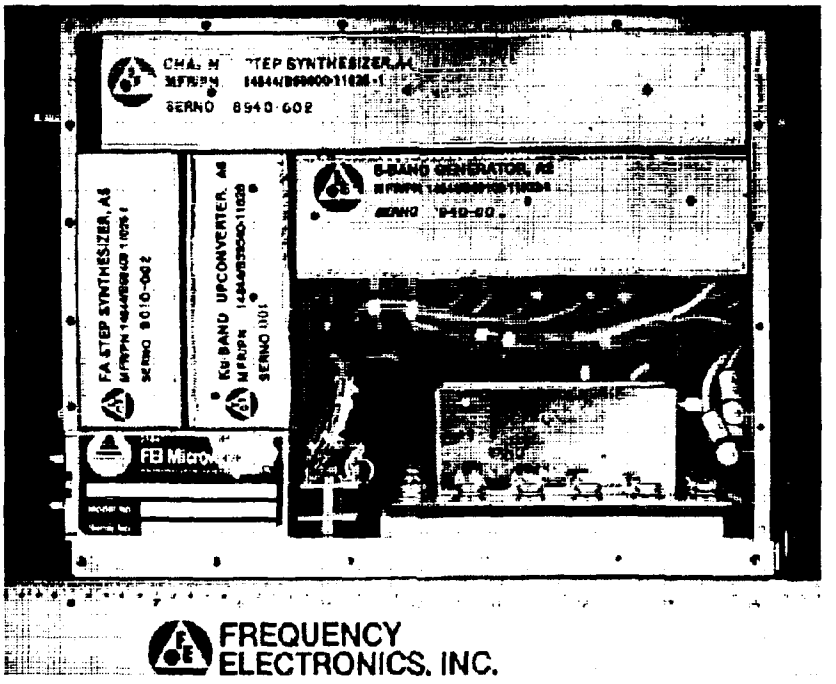


Figure 9. Assembly Configuration — Left Side View

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

COMPUTATION OF OSCILLATOR OPEN-LOOP BODE PLOTS

Donald E. Phillips
Collins Avionics & Communications Division
Rockwell International Corporation
Cedar Rapids, IA 52498

Introduction

This paper presents a method for computing the open-loop gain and phase response of oscillators using two successive SPICE runs. A current balance method is described which insures accuracy. Some graphical results are discussed.

Open-loop response computation and plotting is a standard procedure in the design of negative feedback systems. It indicates the degree of stability and other aspects of transient and steady-state performance. It is quite straight forward to compute the product of all the transfer functions around a feedback loop, especially when these functions are well buffered. Several ingenious measurement methods are in use.

Knowing the oscillator open-loop characteristics would help in the analysis and design of high-performance oscillators. However, these positive feedback circuits present a real challenge. Especially at the higher frequencies, the active device is bilateral; its input impedance depends upon the output load, and its output impedance depends upon the source impedance which is reflected back as a load on the output. It has been commonly thought to be impossible to analyze this condition.

Two problems that exist with closed-loop computation do not occur with open-loop computation:

1) The open-loop analysis is in the frequency domain, where nothing "happens" as it would in the time domain (such as oscillation). Only steady-state conditions are recognized, and these only from a specified driving signal which is made small enough so that all numbers are within the computer's dynamic range.

2) The open-loop response is not multiplied up to high values at resonance like the closed-loop response. The expected values at resonance are 0 Db magnitude and 0 degrees phase.

The real problem is in producing a drive signal which does not alter the loop impedance loading and feedback conditions. Before discussing the current balance method, the difficulties with conventional methods will be described.

Oscillator Loop with Test Circuit

Attempts have often been made to break the feedback connection for the analysis or measurement (Figure 1). It should be understood that the desired open-loop response is the total complex gain around the loop when it is closed. If the measurement or analysis technique alters the loop impedance loadings and feedback conditions, whether the loop is open or closed, the resulting response is inaccurate.

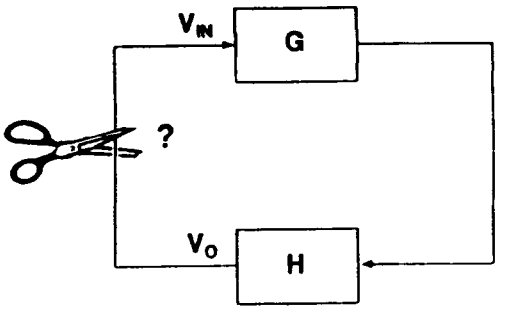


Fig. 1. Open Loop Transfer Function

Figure 2 shows an oscillator loop with a series test circuit, with designators as used in this paper as follows:

Zin	Active device input impedance
Zs	Source impedance to the device input
Vin	Input voltage to the device
Vo	The loop output voltage up to the device input
Test Circuit	Means of signal injection into the loop for analysis

CH2965-2/91/0000-336 $1.00 © 1991 IEEE

Since there is bilateral signal flow, that is, additional feedback around the loop in the opposite direction of the arrows, the impedances Zs and Zin are not buffered, but affect each other. For accurate computation of (Vo/Vin)(s), the following three requirements must be met:

1) Driving signal Vin applied to Zin
2) Zero series impedance both ways
3) Infinite shunt impedance
4) Zero voltage fed back to Zs, so that only Vo from the oscillator loop is present

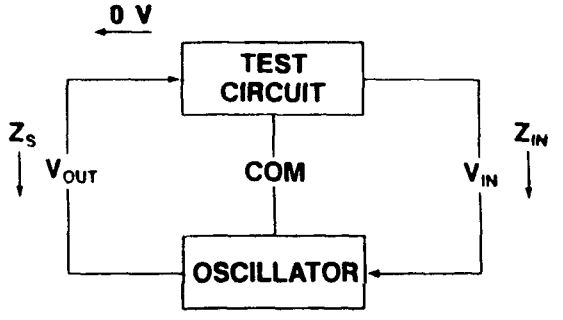


Fig. 2. Open Loop Test Circuit.

Inaccurate Methods

The conventional methods of adding a driving signal generally fall into one of the following:

1) Breaking the loop (Figure 3) to treat the open-loop response as a single gain block with output isolated from the input. This not only makes it difficult to maintain d.c. bias conditions, but the impedance loading and feedback between input and output are no longer present. Even if dummy impedances are used as shown, the altered feedback conditions might make the resulting "measurement" inaccurate.

2) Leaving the loop closed while adding a series voltage source (Figure 4) or shunt current source (Figure 5). While not disturbing the impedances and feedback, a voltage is fed back to the source side of the generator which adds to the loop output voltage Vo at this point and can cause serious errors.

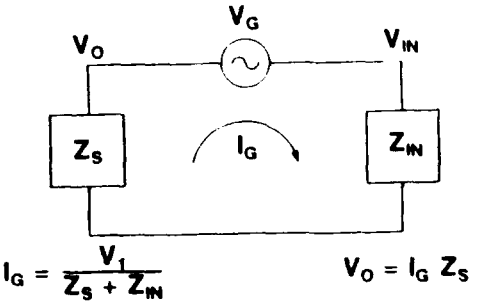


Fig. 3. Voltage Source Error.

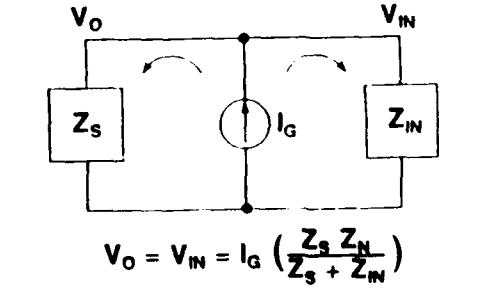


Fig. 4. Current Source Error.

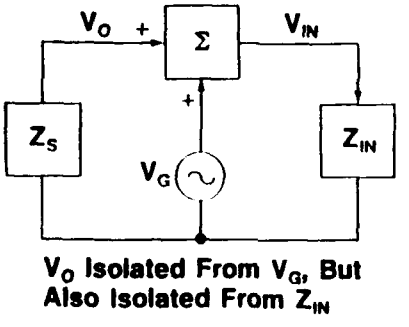


Fig. 5. Voltage Adder Error.

3) Adding a driving signal through a summing circuit (Figure 6). While this prevents a false voltage fed back to Vo, it also isolates Zin and Zs from each other, also causing errors as in 1 above.

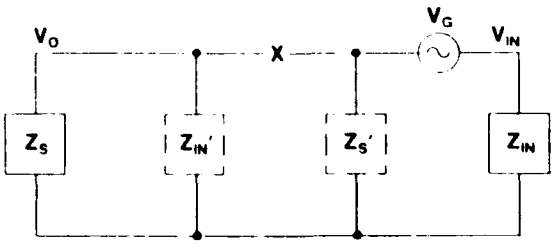


Fig. 6. Dummy Impedances Error.

An Accurate Method

This method introduces a current balance technique which allows a driving signal to be entered into the loop without disturbing the impedance loadings and feedback, and nulls out the unwanted voltage from the measurement of the loop output Vo.

As illustrated in Figure 7, a voltage generator Vg is the Basic driving signal to produce a desired input voltage at Zin. The resulting current Ig flowing through the entire oscillator loop produces an undesired voltage at Zs.

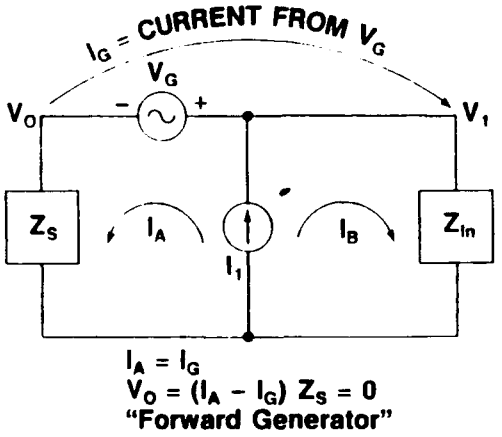


Fig. 7. Error-Free Circuit.

An additional current generator I1 is added which produces two currents, Ia through Zs, and Ib through Zin. The magnitude and phase of the two generators are made such that the current Ia is equal to Ig in magnitude and phase. Flowing in opposite directions, the net driving current through Zs is now zero, and the net current through Zin is larger producing the total Vin. Therefore, the true open-loop function is complex Vo/Vin at each frequency. Note that the impedances and generator values must be computed at each frequency.

By using a frequency analysis program such as SPICE, most of the analysis can be automated. Basically, two successive runs are needed: First, to find the impedances Zs and Zin in the closed-loop configuration; second, using these impedances as models, to find the open-loop response. While by no means a simple operation, it could be highly automated with suitable associated software. I have already automated some of it within the second SPICE run by using circuit analysis techniques to perform some of the computation.

Detailed Explanation

A more detailed description of the process using SPICE starts with a first analysis to find Zs and Zin at the frequencies desired for the analysis. These impedances are stored as SPICE models. In SPICE, impedances are determined by finding the associated voltages and currents. As shown in Figure 8, currents are found by voltage generators used as ammeters, whose generator voltages must be set to zero; otherwise the voltage differences would cause erroneous impedance values. The driving signal for this run must come from a current generator I1. The impedance values for each frequency are printed out to a file.

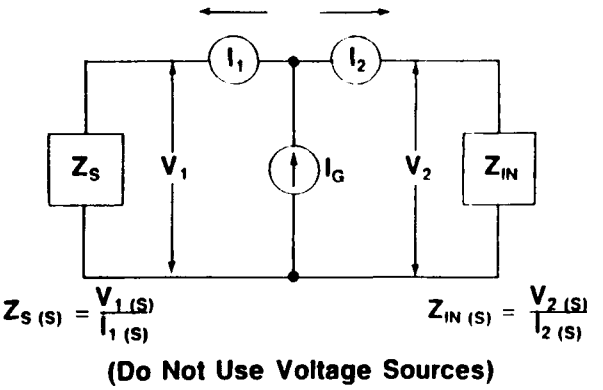


Fig. 8. Determining Zs, Zin.

These impedances can be plotted and are interesting to study. It should be remembered that they are active impedances, and will change when other circuit values are altered.

These impedances can be plotted and are interesting to study. It should be remembered that they are active impedances, and will change when other circuit values are altered.

Then a second SPICE analysis is performed using the Zs and Zin models to derive the generator relationships. I prefer to set I1 at an arbitrary value (such as 1 microampere for small signal analysis), and then compute the required voltage by using additional circuits simultaneously during the second run. As shown in Figure 9, the first circuit computes a current Ia, which when used in the second circuit, computes the value of the voltage Vg to be used in the oscillator test circuit. Special computer runs have tested the degree of balance; the undesired signal component of Vo was attenuated by 280 dB.

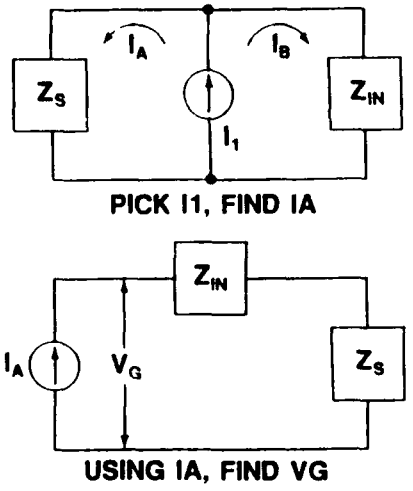


Fig. 9. Computing Vg.

Analyzing a FET Oscillator

Figure 10 is a simplified 200 MHz oscillator using a U310 FET. Measured common source s-parameters were used in the analysis. The resulting open-loop Bode plot is shown in Figure 11. The phase shift around the loop passes through zero degrees at the frequency of oscillation, according to Barkhausen's law, and the gain is higher than unity at that point. In a real oscillator, the gain would be greater than one to start, and then reach limiting where the gain would be very slightly less than unity to allow for the thermal noise driving the circuit.

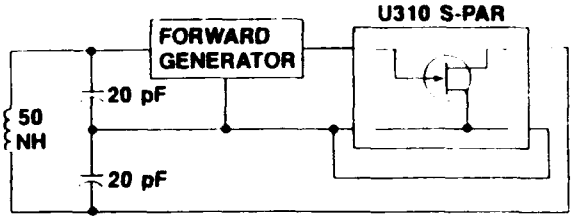


Fig. 10. FET Oscillator.

One of the primary values of this analysis is to compute the oscillator loaded Q, which is a basic parameter in stability and noise analysis. This is the Q of the resonant tank circuit loaded by the output port, the active device, bias circuits, etc. It is not the multiplied Q as seen in the closed loop response, which is very much higher than the loaded Q. For low phase noise close to the carrier, a high loaded Q is necessary. The usual determination of loaded Q is by the carrier offset frequency where the phase noise crosses the additive noise. This computer analysis technique allows loaded Q to be predicted before a circuit has been built. The calculation of loaded Q is based on the slope of phase vs frequency through the zero crossing at resonance.

If the frequency Fo is determined at 0 degrees phase and another frequency F1 at some phase Φ near 0, the loaded Q

$$Q = \frac{F_o \tan \phi}{2 (F_o - F_1)}$$

Using the open-loop phase values derived from the SPICE analysis of the U310 oscillator example (Figure 11), the loaded Qs for different sets of values are:

Tank L	Tank C (2)	Q
40 nH	25 pF	18
25 nH	40 pF	53
10 nH	100 pF	202

This illustrates the strong effect of the L/C ratio on loaded Q, especially if the tank losses are low as in this example where the tank is lossless. According to

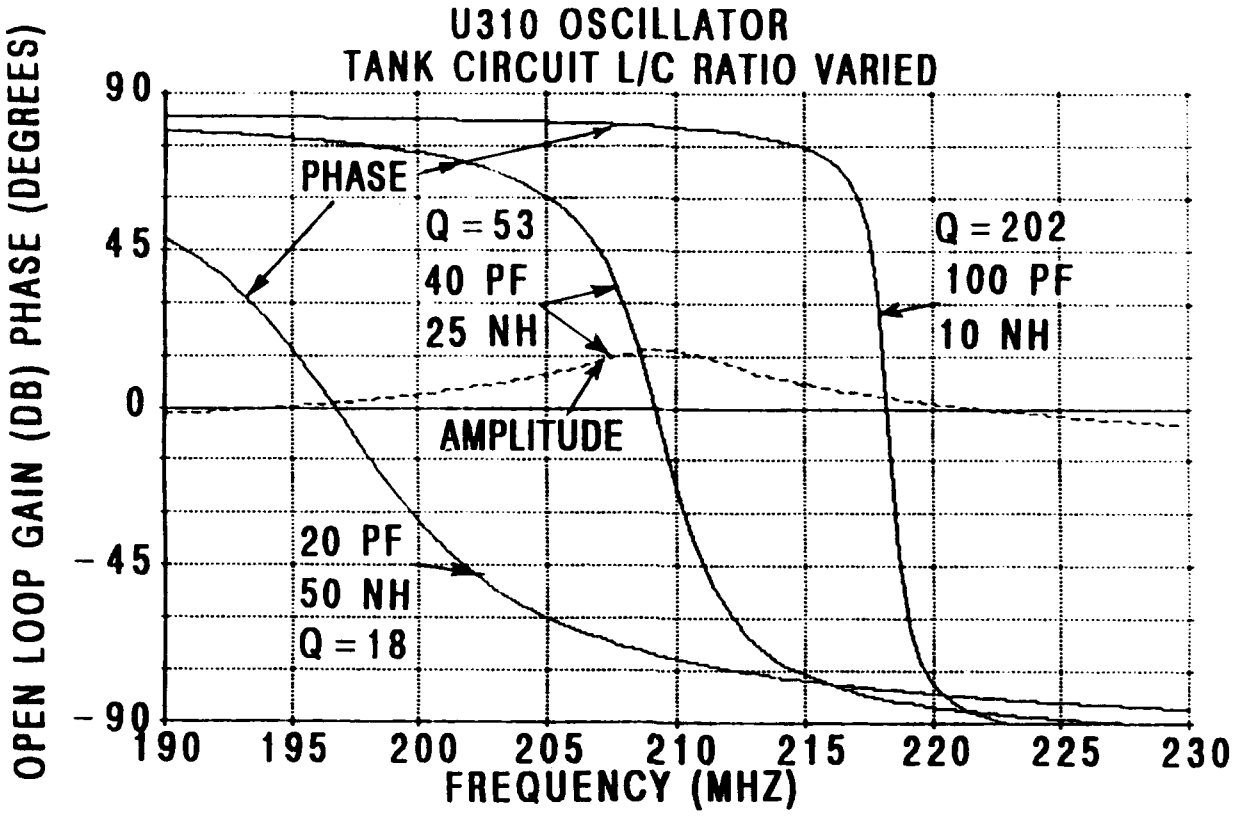


Fig. 11. Open-Loop Bode Plots.

this computation, the loaded Q increases faster than the L/C ratio decreases, perhaps because of the interactive impedances between the transistor and the tank circuit.

Conclusions

An analysis technique was demonstrated to accurately compute oscillator open-loop gain. The phase slope is related to the loaded Q, which can be used to predict some aspects of the oscillator noise characteristic. This analysis was done using small-signal parameters in SPICE. More work needs to be done to extend this technique to large-signal analysis using programs such as SPECTRE or LIBRA, to analyze other oscillator circuit configurations, and to compare these results with whatever measured data is most relevant, especially noise levels.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH PRECISION NONLINEAR COMPUTER MODELLING TECHNIQUE FOR QUARTZ CRYSTAL OSCILLATORS

R. Brendel*, F. Djian* and E. Robert**

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32,avenue de l'Observatoire - 25000 Besançon - France
** Centre National d'Etudes Spatiales - Dept DRT/TIT/TF - avenue E. Belin - 31055 Toulouse Cedex - France

Abstract

This paper presents a full nonlinear modelization technique for quartz crystal oscillators. The model computes both imaginary and real part of the nonlinear complex polynomial characteristic, giving the frequency and amplitude of the oscillation. It is shown how a computer aided symbolic calculation can be used to obtain error free polynomial coefficients. The nonlinear behaviour of the active device is described by using nonlinear admittance parameters whose value is a function of the signal level. These functions have been obtained by using an electrical simulator commercially available, and the results are compared with experimental data. Modelization of the oven has also been performed, and this paper presents the program used to obtain the thermal transfer function of the oven.

Introduction

The two main purposes the simulators can serve are either design or analysis. When developing a new oscillator, its type and structure are usually determined by some mandatory requirement like frequency, power consumption, signal level, stability, etc. Then, the designer has to find the optimum value of the circuit components. This task do not need a high precision simulator since it has only to give the designer the rate of variation and the extreme values of the most important parameters so that he can check that the device will work within the specifications under all conditions.

On the other hand, if one has to model an existing oscillator in order to analyze its behavior in a particular environment or to evaluate the frequency shift produced by a modification of some component or other, a high precision simulator is required. This is why a new approach has been developed to settle a high precision nonlinear simulator intended to analyze without ambiguity the frequency variations measured by Doppler effect within the scope of orbitographic missions. The expected goal is to have a model as accurate as possible enabling the simulation of environmental conditions like temperature, magnetic field, etc., so that their effect can be removed from the useful signal.

Although general purpose electrical simulator like SPICE [1] or ASTEC [2] can be used for the modelling of the oscillators, they are not well adapted to the case of quartz crystal oscillator because of the very high quality factor of these devices. In fact, the frequency and amplitude of oscillations can be obtained only through a transient analysis when the steady state is reached. It is well known that the time required to reach the oscillation steady state from the start or after a perturbation is typically of the order of magnitude Q/f_0 where Q is the quality factor and f_0 the oscillator frequency. For example, a 10 MHz quartz crystal oscillator whose quality factor is about 10^5 needs about 0.01 s to 1 s to reach the steady state that is 10^5 to 10^7 periods, if we would evidence relative frequency changes $\Delta f/f \approx 10^{-6}$ we should be compelled to choose calculation time step ΔT so that $\Delta T/T = \Delta f / f$ where $T = 1/f_0$ i.e. 10^6 points per cycle so the steady state would need 10^{11} to 10^{13} calculation steps. A program like SPICE running on a PC/286 computer would required 10^{10} to 10^{12} s of CPU time to perform this task, in other words : 3,000 to 300,000 years, even with a large computer one million times faster, the calculation time remains today prohibitive.

Over a few years, some programs intended to oscillator C.A.D. have been described in the literature. Most of them are built on the solution of the Barkhausen criterion and are usually restricted to a particular oscillator structure or its derivatives for example Colpitts, Pierce or Clapp oscillators. Some of these programs have an oscillator library.

The oscillator simulators can be classified into three categories :

- The purely linear simulators [3, 4], they only gives the root of the imaginary part of the characteristic polynomial by assuming the amplifier parameters are constant. The gain condition is restricted to an inequality : the small signal amplifier gain has to be large enough to insure the oscillation can start, obviously, these programs cannot give the amplitude of oscillations.

- The frequency-linear amplitude-nonlinear simulators. Usually these simulators calculates the root of the imaginary part of the complex characteristic by setting *a priori* the gain margin to an acceptable value in order to work out the oscillation frequency [5]. The amplitude of the signal

CH2965-2/91/0000-341 $1.00 © 1991 IEEE

is separately calculated by determining the type of limitation the active element is subjected to. For example base-emitter or base-collector cutoff [6]. Nevertheless, these simulators do not take into account the coupling between amplitude and frequency.

- Full nonlinear simulators. Some attempts to elaborate nonlinear models of oscillators have been described. Some of them start from results coming from general purpose electrical simulator either to describe the amplifier behavior before use in the simulator [7, 8] or to simulate the oscillator as a whole, the large transit time problem being overcome by modifying the initial conditions until the steady state amplitude is reached through a successive approximation process [9].

Modelling principles

Many important problems arise when constructing an accurate model for an oscillator. Among them one can quote the large amount of components which leads to a lot of intricate equations difficult to handle, and the intrinsic nonlinear behavior of the device which needs to seek a solution for both the frequency and amplitude of oscillation because they are not independent. To this end the nonlinear behavior of the active component has to be known. This paper describes the different methods used to overcome these difficulties.

Let us consider the simple Colpitts oscillator shown on Fig. 1. It will be used throughout this paper to present the various methods used.

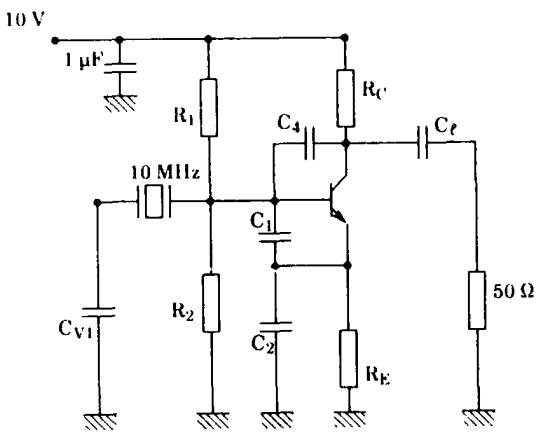


Fig. 1 : Schematic of a simple Colpitts' oscillator

The oscillator (Fig. 1) is conceptually composed of an amplifier and a resonant feedback loop whose main component is often a quartz resonator. The purpose of the amplifier is to sustain the oscillation amplitude to the desired level, it consists of an active component : here a bipolar transistor and its bias circuitry. The derivation of the oscillation equation consists in writing that when the steady state is reached, the resonant circuit feedback to the amplifier input a signal whose amplitude and phase are identical to those which produced it (Barkhausen criterion). This identity leads to a "characteristic equation" which is a complex polynomial whose degree is increasing with the number of components in the circuit. The root of the imaginary part also called "zero phase condition" gives the oscillation frequency. The real part is related to the amplitude condition, the amplifier has to make up for the losses of the remaining circuit. When the oscillations start the amplifier gain is high so the amplitude increases, because the amplifier is a nonlinear device, its gain is decreasing when the amplitude growths and the steady state amplitude is reached when the gain is just large enough to compensate exactly the circuit losses.

Nonlinear description of the active component

When dealing with small signals, the behavior of a bipolar transistor can be represented by a linear equivalent network using small signal parameters. Among the different possible representations, the admittance parameters also called "y parameters" are often used in RF electronics, their value is depending upon three main parameters : the frequency, the bias conditions (operating point) and the temperature but they do not depend on the signal amplitude if it remains small enough. At the opposite, when used in a given oscillator, the small signal approximation can no longer be applied, in return the operating frequency remains constant. At this stage it is possible to carry on with the y parameter representation with the difference that they are now function of the amplitude.

Figure 2 represents the equivalent network of a bipolar transistor using y parameters, the four complex admittances y_n can be put on the form

$$y_n = g_n + pC_n \quad (n = i, f, r \text{ or } o) \tag{1}$$

where g_n is a conductance, C_n a capacitance and p the Laplace's operator. Each of the 8 quantities g_n and C_n depends on the amplitude of the base-emitter voltage v_i, or on the amplitude of the collector-emitter voltage v_o.

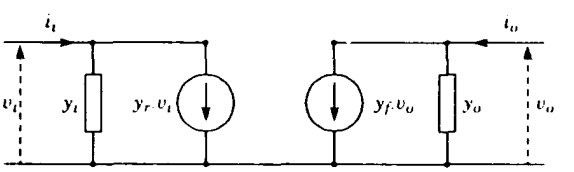


Fig. 2 : Equivalent network of bipolar transistor

There are three methods to determine the way the y parameters are depending on the amplitude : calculation, simulation or measurement. As regard of the complexity of a nonlinear model describing the transistor behavior, calculations would need a lot of approximations which may leads to inaccurate results. The simplest way to obtain these large signal y parameters consists in using an electrical simulator all the more because these programs have today well-supplied component libraries with very sophisticated and accurate models. This method has been used here along with experimental verifications. The simulations have been performed with the program SPICE, now available on a large number of installations. The determination of the y parameters by simulation results from their definition

$$y_i = \left(\frac{i_i}{v_i}\right)_{v_o=0} \qquad y_r = \left(\frac{i_i}{v_o}\right)_{v_i=0}$$

$$y_f = \left(\frac{i_o}{v_i}\right)_{v_o=0} \qquad y_o = \left(\frac{i_o}{v_o}\right)_{v_i=0} \tag{2}$$

The simulator gives the amplitude and phase of the voltages and currents, then the amplitude and phase of y_i may be obtained from the relations

$$|y_i| = \left|\frac{i_i}{v_i}\right|_{v_o=0} = Y_i$$

$$\angle y_i = (\angle i_i - \angle v_i)_{v_o=0} = \Phi_i \tag{3}$$

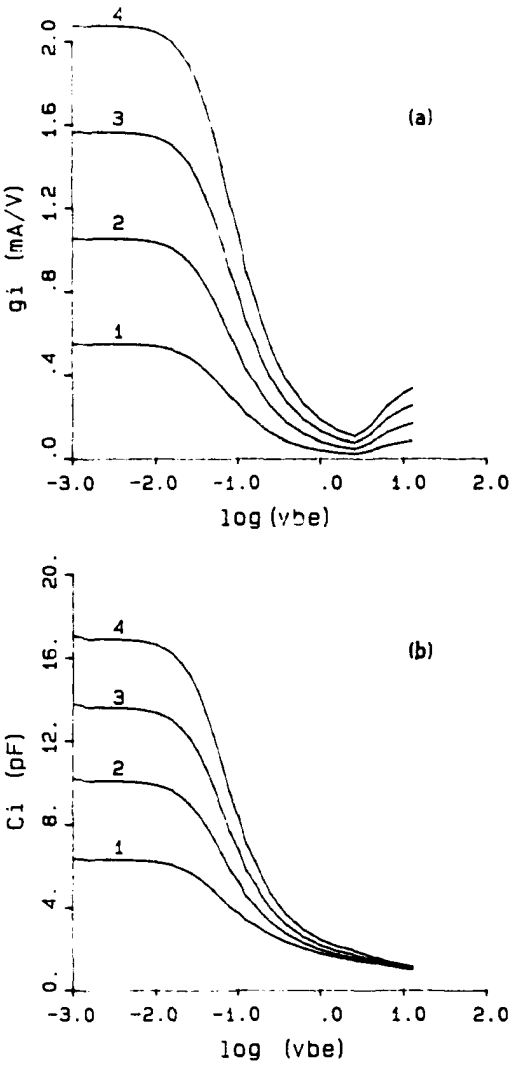


Fig. 3 :
Nonlinear behavior of the parameter y_i, V_{CE} = 3.3 V
(1) I_c = 0.95 mA
(2) I_c = 1.9 mA
(3) I_c = 2.85 mA
(4) I_c = 3.78 mA

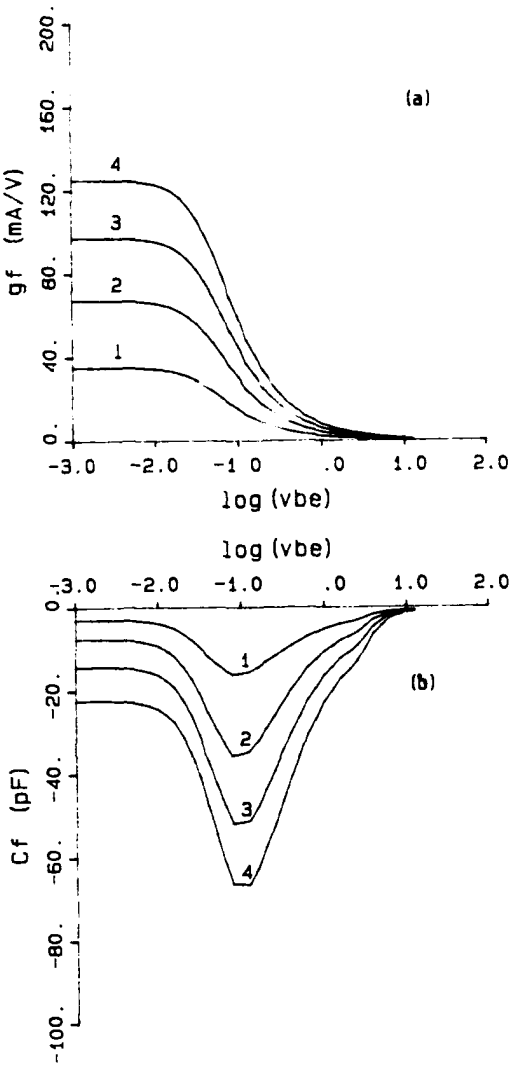


Fig. 4 : Parameter y_f. Same conditions as in Fig. 3

So the conductance g_i and capacitance C_i defined by Eq. 1 can be derived

$$g_i = Y_i \cos\Phi_i$$
$$C_i = \frac{Y_i \sin\Phi_i}{\omega} \qquad (4)$$

The six remaining parameters are obtained in a similar way.

Circuits and method used for the determination of nonlinear y parameters

Eqs. (2) show that two different circuits are needed to determine the y parameters, one for y_i and y_f and another one for y_r and y_o. In both cases, the dc biasing circuit is the same as in the oscillator. In the first case the input signal is applied to the base and the output is short circuited by putting a large capacitor between collector and emitter to insure that $v_o = 0$ while in the second case the input signal is applied to the collector and the transistor input is short-circuited by a capacitor put between base and emitter to insure that $v_i = 0$.

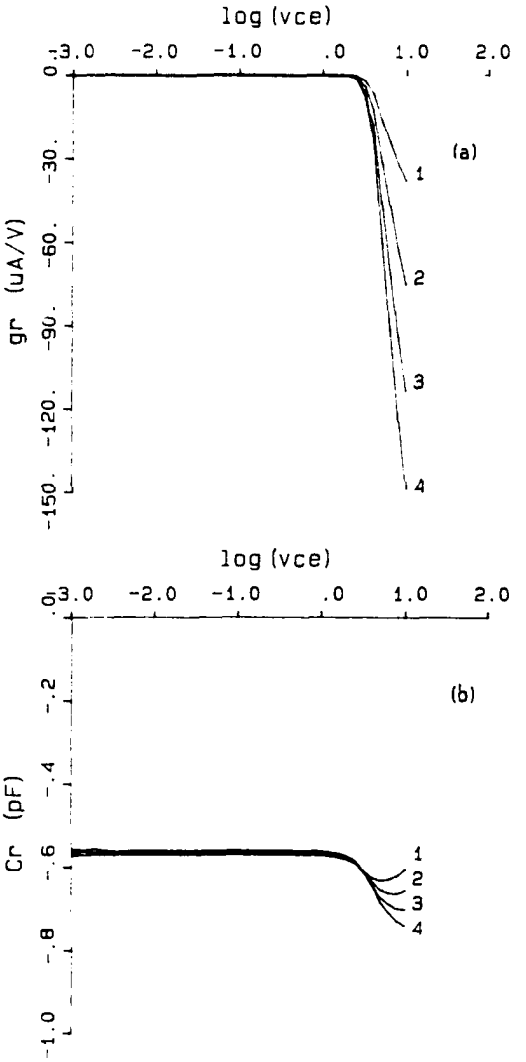


Fig. 5 : Parameter y_r. Same conditions as in Fig. 3

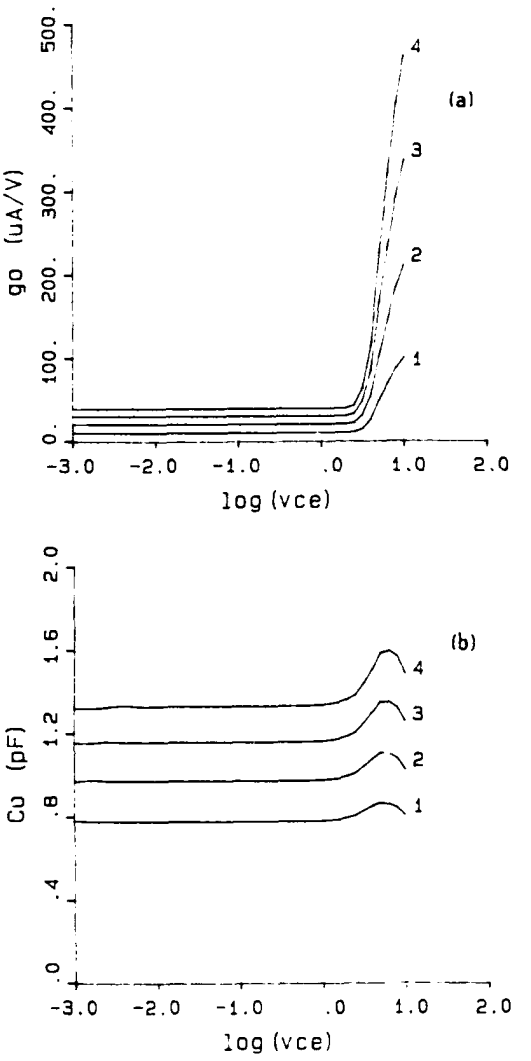


Fig. 6 : Parameter y_o. Same conditions as in Fig. 3

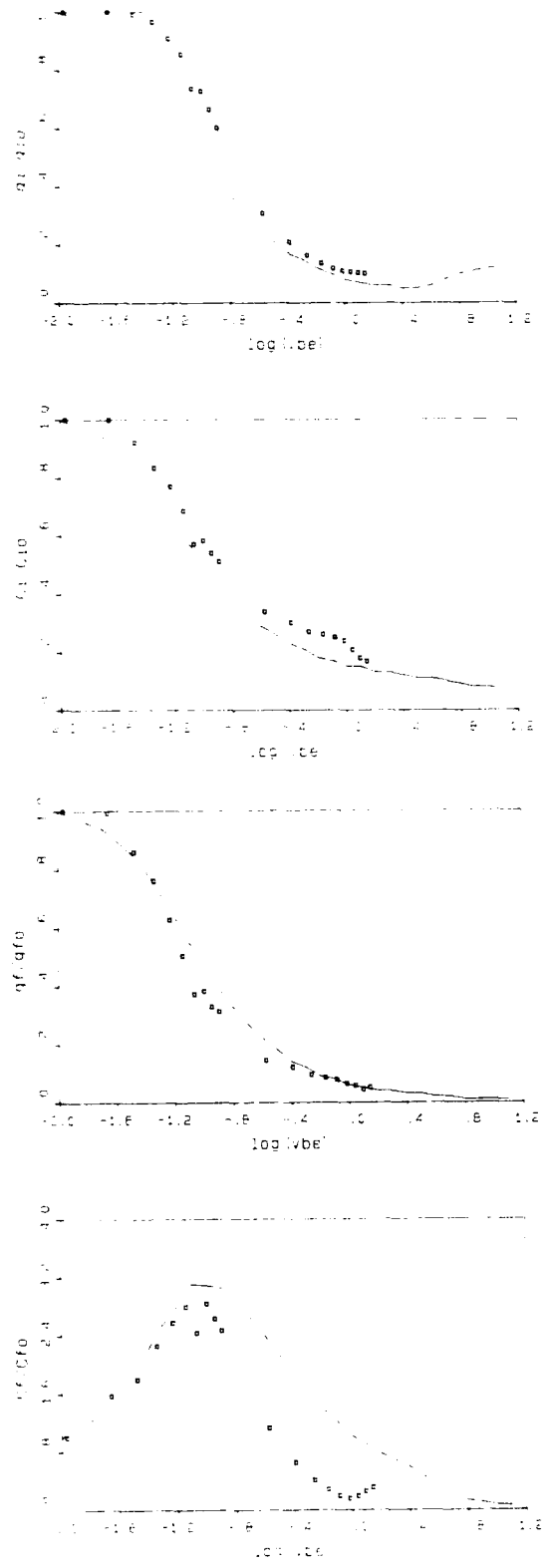


Fig. 7 : Comparison between simulation (solid lines) and experiment (symbols) for parameters y_i and y_f

Simulations have been achieved by using input signals of increasing amplitude. When large amplitudes are used, the signals in the circuit are no longer sinusoidal so the phase differences are more and more difficult to obtain as the amplitude increases. Fortunately, the simulator gives the user the amplitude and phase of the nine first components of the Fourier series from which the parts g_n and C_n of the different parameters can be derived through Eqs. (3) and (4).

Different simulations have been carried out in various biasing and temperature conditions. Figs. 3, 4, 5 and 6 show the results obtained.

The experimental verification of the simulation results has been carried out. The source used was a 10 MHz synthesizer and the measurements were made with a vector voltmeter for input voltage varying from 10 mV to 1 V rms (maximum upper limit for the apparatus).

Only the two parameters y_i and y_f have been measured. Fig. 7 shows the comparison between simulation and experimental data. The curves have been normalized with respect to the small signal value of the parameters.

Basic oscillator equations

Fig. 8 shows the equivalent circuit for the Colpitts oscillator described on Fig. 1.

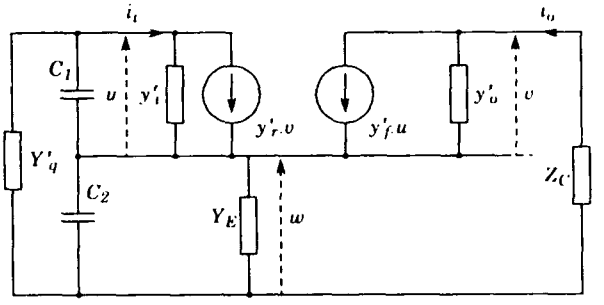


Fig. 8 : Equivalent network of the simple Colpitts' oscillator shown on Fig. 1

The admittance Y'_q is formed by the crystal, the matching capacitor C_{v1} and the bias resistors R_1 and R_2.

The crystal admittance itself can be put under the following form

$$Y_{sq} = \frac{p C_q}{1 + p C_q (R_q + p L_q)}$$

so that (5)

$$Y'_q = \frac{p C_{v1} Y_{sq}}{Y_{sq} + p C_{v1}} + Y_B$$

Y_E is the admittance of the emitter biasing resistor R_E and Z_c is the output impedance which consists of the load impedance R_l, the coupling capacitor C_ℓ and the collector bias resistor R_c.

$$Z_c = \frac{R_c(1 + p R_l C_\ell)}{1 + p C_\ell (R_l + R_c)} \tag{6}$$

The capacitor C_4 has been taken into account by modifying the transistor y parameters

$$y'_i = y_i + y_4 \qquad y'_r = y_r - y_4$$
$$y'_f = y_f - y_4 \qquad y'_o = y_o + y_4 \tag{7}$$

where $y_4 = p C_4$.

The characteristic polynomial can be derived by inspection of Fig. 8. By expressing the input current i_i as a function of the input and output voltages u and v, the input ratio v/u can be written in the form

$$\frac{v}{u} = - \frac{y'_1 + Y_1 + (1 + a_1) Y'_q}{y'_r + a_2 Y'_q} \tag{8}$$

where

$$a_1 = \frac{y_i + y_f + Y_1}{Y_E + Y_2}, \quad a_2 = \frac{y_o + y_r}{Y_E + Y_2}, \quad Y_1 = p C_1, \quad Y_2 = p C_2$$

In the same manner, the output current can be expressed as a function of u and v leading to the output ratio :

$$\frac{v}{u} = - \frac{a_1 + Z_c y'_f}{1 + a_2 + Z_c y'_o} \tag{9}$$

When the steady state is reached, the two ratios (8) and (9) are equal. Replacing all terms by their expression as functions of the circuit parameters and the Laplace's operator p (Eqs. 5, 6, 7) leads to the characteristic polynomial.

$$\sum_{k=0}^{N} a_k p^k = 0 \tag{10}$$

where a_k are function of the circuit parameters and N the polynomial degree which depend on the complexity of the circuit.

Automatic program generation

Even with the simple Colpitts oscillator described above, expressing characteristic polynomial coefficients a_k (Eq. 10) leads to rather long and tedious calculations. When dealing with more sophisticated networks involving mode or overtone selection or electronic pulling circuits, the polynomial degree and the calculation complexity drastically increase with the number of components so the error risk when deriving and coding the polynomial coefficients becomes important.

These difficulties can be avoided by calling on computer aided symbolic calculation. Many softwares permit the user to handle algebraic expressions like polynomial expansion or factorization instead of numerical data. Among the available programs one can quote MATHEMATICA [10], MACSYMA [11] or REDUCE [12]. In addition of their symbolic calculation ability, these programs often allow the user to get results directly expressed in high level computer language. This is the case for the program REDUCE used here which can output results under FORTRAN statement form. Then these instruction code lines can be safely inserted in the simulation program without manual transcription. Among the large number of calculations REDUCE can perform, the most useful features for the present purpose are :

- algebraic calculations
- polynomial manipulations : rearrangement, factorization, expansion, polynomial degree determination
- rational manipulations : munerator and denominator separation ...

Fig. 9 shows an example of a REDUCE session, the numbered lines are typed by the user, statements 1: and 2: assign polynomials to variables x and y, statement 3: ask for all the polynomial coefficients of the product. Note that it is not necessary to know the leading degree of the product, it is calculated by the operator *deg*. The statements 4: and 5: respectively ask for the first derivative of x and for the indefinite integral of y with respect to the variable p.

```
reduce
REDUCE 3.3, 1-Aug-88 ...

1: x := a0 + a1*p + a2*p^2 + a3*p^3 ;

       3        2
X := P *A3 + P *A2 + P*A1 + A0

2: y := b0 + b1*p + b2*p^2 ;

       2
Y := P *B2 + P*B1 + B0

3: for j := 0 : deg(x*y,p) do write coeffn(x*y,p,j) ;

A0*B0

A1*B0 + A0*B1

A1*B1 + A2*B0 + A0*B2

A1*B2 + A2*B1 + A3*B0

A2*B2 + A3*B1

A3*B2

4: z := df(x,p,1) ;

       2
Z := 3*P *A3 + 2*P*A2 + A1

5: int(y,p) ;

      2
 P*(2*P *B2 + 3*P*B1 + 6*B0)
-----------------------------
              6

6: quit ;

Quitting
/
```

Fig. 9 : Sample of a REDUCE session

A common awkwardness when using that kind of program is to input everything it needs, that is the starting basic relations like Eqs. 5, 6, etc. and to ask for the final result Eq. 10 : this is probably the best way to spent a maximum of money and CPU time for a minimum of result or no result at all because of the incredibly large and intricate expressions generated. Rather, the code has to be built step by step : the basic relations are put in the form of rational expressions, the numerator and denominator of which are polynomials whose degree is known, then each polynomial coefficient is expressed in term of the circuit parameters and saved, then they are replaced by an intermediate variable giving the initial rational expressions a simpler form of same degree which will be used for the next step (Eqs. 8 and 9 for example). In the same manner, the resulting rational expression of known degree can be written in terms of a new set of symbolic variables which are function of the former ones. Eventually, the degree of the characteristic polynomial is determined, and the coefficients are expressed in terms of the symbolic variables defined in the previous step. This method presents the advantage that the intermediate variables are always expressed in term of variables previously defined, the circuit parameters being used in the first step. This is important for the high level language coding which thus becomes much more compact and efficient and which can be used without modification in the numerical computation program.

Solving the characteristic polynomial

At this stage, the polynomial coefficients a_k of Eq. 10 are known. By putting $p = j\omega$ this equation can be shared into real and imaginary part :

$$\sum_{k=1}^{N1} \alpha_{2k}\, \omega^{2k} = 0$$

$$\sum_{k=1}^{N2} \beta_{2k+1}\, \omega^{2k+1} = 0 \qquad (11)$$

Each coefficient α and β is expressed as a function of the circuit parameters, some of them (namely y_n) being function of the signal amplitude u (Figs. 3-6). Now the problem is to find the frequency ω_0 and the amplitude u_0 satisfying simultaneously both Eqs. 11. This can be done by analogy with the actual behavior of an oscillator : the amplitude u is given the smallest value u_1 for which the transistor parameters $y_n(u_1)$ are known. then the root of the imaginary part can be calculated giving a first approximation ω_1 of the oscillation frequency. The real part of the characteristic equation (11) is thus calculated, at this step it is likely different from zero, in the second step, the amplitude is increased to a new value $u_2 > u_1$ giving a new set of transistor parameters $y_n(u_2)$ thus a second approximation ω_2 of the oscillation frequency and a second value of the real part likely smaller than the former one. The process is repeated with increasing amplitudes until the sign of the real part changes. The steady state amplitude is then located between two successive values u_n and u_{n+1}. A more accurate solution is obtained by using a bisection process between these values. Eventually, the oscillation amplitude and frequency are determined within a given error range.

Fig. 10 shows the flow chart of the program used.

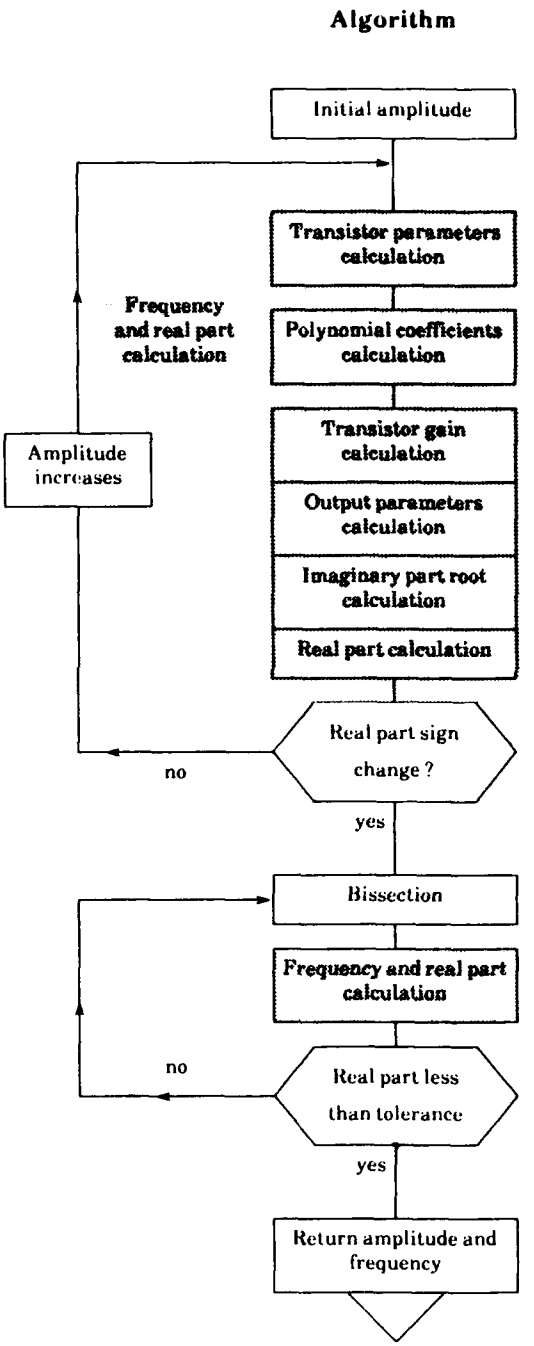


Fig. 10 :
Flow chart of the nonlinear simulation program

Heat transfer principles

Figure 11 shows a simplified form of the oven used for the experimental verifications.

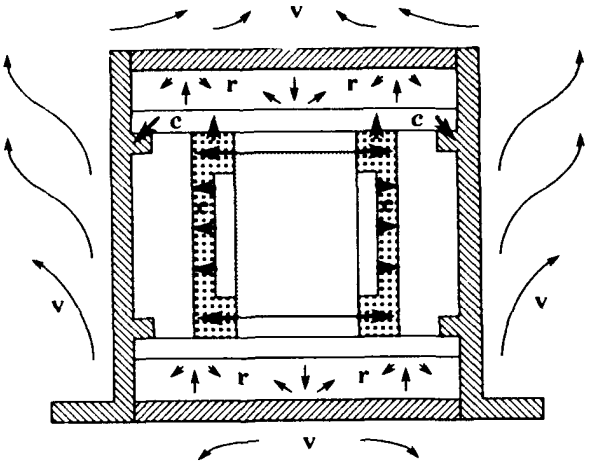


Fig. 11 : Structure of the oven

There are three types of heat transfer mechanisms which have to be considered : radiation (r), convection (v), and conduction (c).

When two surfaces of areas S_1, S_2 and temperature T_1, T_2 are close with each other, they exchange radiated heat flux which can be expressed under the simplified form [13]

$$\Phi = r_{12} (T_1^4 - T_2^4)$$

with

$$r_{12} = \sigma \varepsilon_1 \varepsilon_2 S_1 F_{12}$$

where $\sigma = 5.67\ 10^{-8}$ W.m^{-2}.K^{-4} is the Stefan's constant, ε_1, ε_2 are the emission factors which express that a real body emits only a part of the flux that a black body would emit at the same temperature, F_{12} is a geometrical form factor representing the ratio of the total heat flux emitted by S_1 which is received by S_2. The form factor mainly depends on the geometrical characteristics of the two surfaces. This simplified relation can be refined by taking into account other phenomena like the fact that a part of the received flux is reflected and the reflection can be either specular (polished surfaces) or diffuse (frosted or rough surfaces).

Convection involves heat exchange between a solid side wall and a surrounding fluid [14]. It is described by a law which can be written in the form

$$J = \upsilon_{12} (T_2 - T_1)^{5/4}$$

where J is the surface heat flux (T_2-T_1) the temperature difference between the body surface and the fluid temperature, υ_{12} is a coefficient which depends on the physical conditions of the convection. In the case of free convection of interest here, υ_{12} depends on the nature of the fluid motion which can be either laminar or turbulent. The main features of the fluid motion are described by three parameters :

- Prandtl's number, P_r expresses the ratio between linear momentum and heat transfer abilities. Table 1 shows some typical values of P_r as a function of the temperature when the fluid is air.

T	P_r
0° C	0.712
50° C	0.701
100° C	0.690

Table 1

- Nusselt's number, N_u measures the ratio between heat convection and conduction efficiencies.

- Grashof's number, G_r indicates the nature of the fluid motion : small value means that convection can be limited to heat conduction inside the fluid, medium value means laminar flow, and large value means turbulent flow. Also, geometrical arrangements have to be considered to describe heat convection (for instance the values are different if the wall is vertical or horizontal).

Heat conduction is described by Fourier's law. In the case of constant thermal conductivity, the heat conduction between two nodes at temperatures T_1 and T_2 can be expressed as

$$C \frac{dT}{dt} = C_{12} (T_2 - T_1) \qquad (12)$$

C is the heat capacity of the node of interest, and C_{12} is the conductive coupling coefficient. The coefficients C and C_{12} depend on the geometrical dimensions and on the heat conductivity of the conductor placed between the two nodes.

Thermal modelling of the oven

An accurate description of the thermal behavior of an oven has to take into account the three heat transfer mechanisms previously described : radiation, convection and conduction. In the general case that analysis leads to a set of nonlinear partial differential equations which cannot be solved without approximations and whose coefficients are not well known. On the other hand, heat exchange between the various parts of the oven is often dominated by only one of the three phenomena. The thermal modelling of the oven starts thus with a decomposition of the whole structure into some number of nodes which are considered as isothermal. So the heat exchanges between node i and the other ones are described with a set of differential equations

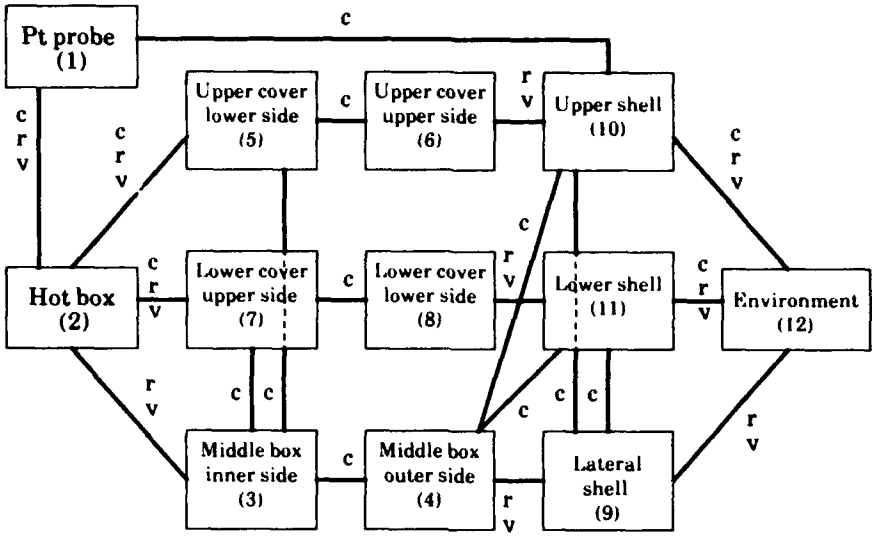


Fig. 12 : Definition of nodes and heat exchanges in the oven

$$C_i \frac{dT_i}{dt} = \sum_{j=1,n} \left| r_{ij}(T_j^4 - T_i^4) + C_{ij}(T_j - T_i) + v_{ij}(T_j - T_i)^{5/4} \right| + P_i \qquad (13)$$

where C_i is the heat capacity of node i, r_{ij}, C_{ij} and v_{ij} are radiative, conductive and convective coupling coefficients respectively, P_i is the thermal internal power dissipated by node i.

In the case of interest, the oven has been divided into 12 nodes whose description is summarized in Table 2.

Node	Designation
1	Inner hot box (Resonator)
2	Hot box
3	Intermediate shell (inner side)
4	Intermediate shell (outer side)
5	Upper cover (lower side)
6	Upper cover (upper side)
7	Lower cover (lower side)
8	Lower cover (upper side)
9	Side shell
10	Upper shell
11	Lower shell
12	Room temperature

Table 2

Node i	Node j	Nonlinear convective coefficient ($W/K^{5/4}$)	Nonlinear radiative coefficient (W/K^4)	Conductive coefficient (W/K)
1	2	0.001	$6.8.10^{-13}$	$1.0.10^{-3}$
1	10			$1.0.10^{-5}$
2	3	0.028	$1.62.10^{-11}$	
2	5	0.008	$1.3.10^{-11}$	0.018
2	7	0.008	$1.3.10^{-11}$	0.018
3	4			2.0
3	5			0.036
3	7			0.036
4	9	0.08	$3.62.10^{-11}$	
4	10			0.05
4	11			0.05
5	6			0.35
6	10	0.007	$2.9.10^{-11}$	
7	8			0.35
8	11	0.007	$2.9.10^{-11}$	
9	10			0.3
9	11			0.3
9	12	0.061	$2.5.10^{-10}$	
10	12	0.015	$1.25.10^{-10}$	0.13
11	12	0.015	$1.25.10^{-10}$	0.13

Table 3 : Thermal coupling coefficients values between nodes

Node 1 corresponds to the oscillator electronics while node 12 is the external medium. Figure 12 shows these nodes and the kind of heat exchanges between them as they result by inspection of the oven structure. Here should be noted that in some case all phenomena have to be considered while in other cases only one or two phenomena occur. Moreover, heat exchanges between some nodes is so small that they can be neglected.

Table 3 shows the values of the different coupling coefficients obtained from calculations or from literature data.

A thermal simulation program has been developed to solve heat transfer equations (13). The set of nonlinear differential equations is solved by using a fourth order Runge Kutta method. This program can be used to get the temperature versus time response of the different nodes of the model. To run the simulation, the program needs the external temperature (either constant or function of time), the temperature setting and the coefficients of the PID controller.

Experimental verifications

The validity of the model previously described has been checked by using the experimental setup shown on Fig. 13. The temperature of the different nodes has been measured by thermocouples and the quartz crystal resonator itself has been replaced by a platinum probe. The cold junction of the thermocouples are gathered on a plate whose temperature is measured by a AD590 monolytic probe. The experiment is driven by a microcomputer which measures the temperatures and stores the results. The temperature can be measured within 0.1°C precision.

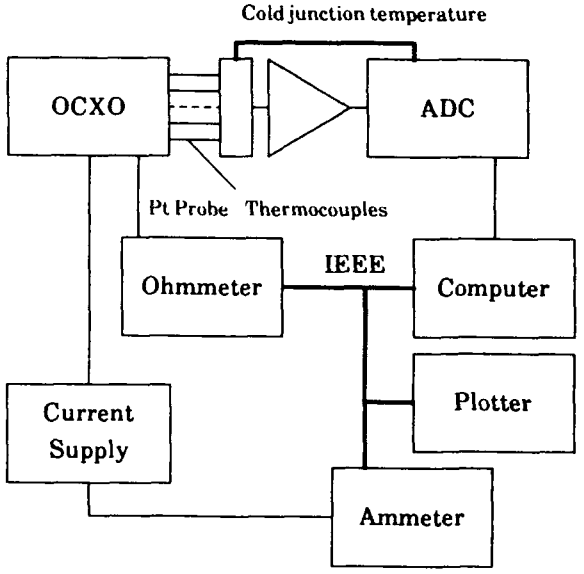
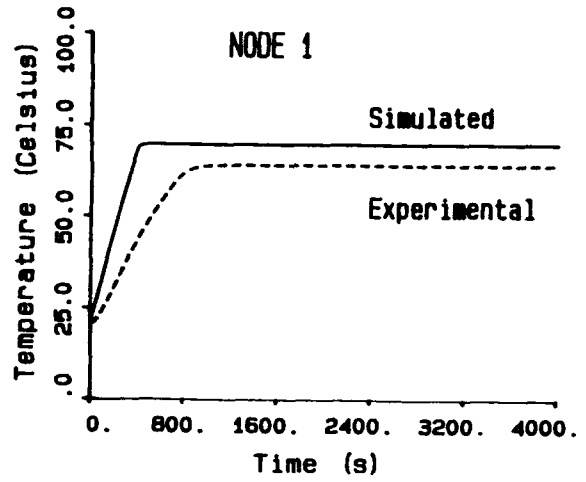
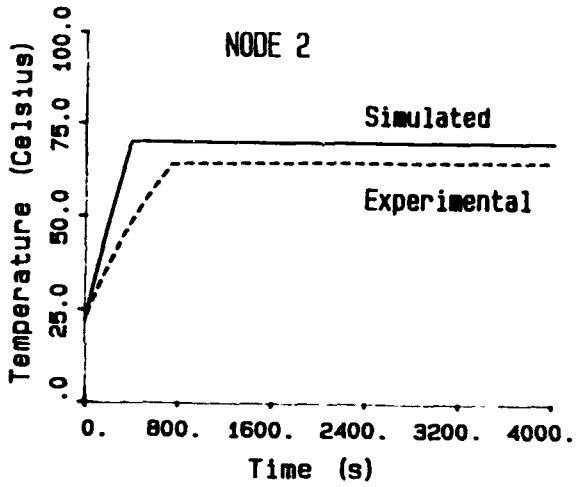


Fig. 13 : Experimental setup
for temperature measurements in the oven

Figure 14 shows the results obtained by simulation and experiment for some selected nodes.





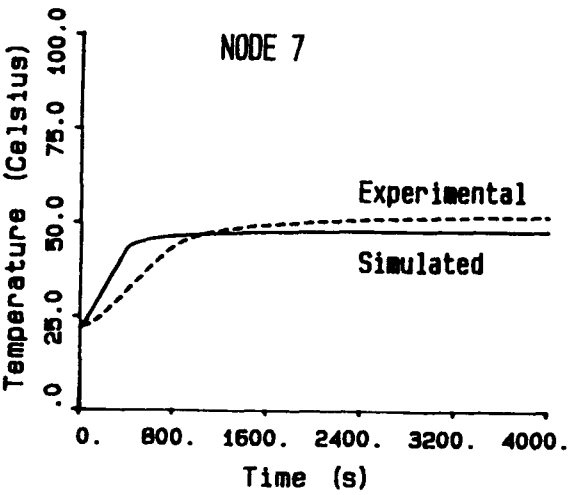


Fig. 14 : Comparison between simulated (solid lines) and experimental (dots) temperature behavior for selected nodes in the oven
node 1 : resonator
node 2 : hot box
node 7 : lower cover lower side

The observed discrepancies are mainly due to the fact that some heat exchange coefficients are very difficult to obtain by a theoretical calculation. Moreover some nodes considered as isothermal may present a marked temperature gradient.

Conclusion

As for most of the oscillator simulation softwares, the present program can be used only for a given oscillator structure and its derivatives obtained by removing some components. However new kernels can be easily generated by using the symbolic calculation program so as to build up an oscillator library covering the most widely used circuits.

Around the kernel, the program has been developed as to offers the user some useful features for the oscillator behavior analysis :

- signal modification calculation when changing one or more component value
- variation calculation by varying a component value between two limits with a given step
- calculation of the signal sensitivity with respect to every circuit component
- component tolerance analysis
- worst case analysis
- temperature effects.

In the future, the program will be completed by implementing other environmental effect like acceleration, magnetic field, noise, etc.

Acknowledgement

This work has been supported by CNES under contract 832 88 5344 00.

Bibliography

[1] L.W. Nagel, "SPICE2 : a computer program to simulate semiconductor circuits", Memorandum ERL-M520, University of California, Berkeley (1975).

[2] M. Heydemann, "ASTEC3 Manuel de référence", Document M208, CISI, Paris (1981).

[3] R.Nardin, J. Ho, "Computer design and analysis for high precision oscillator", Proc. of the 28th AFCS (1974).

[4] Y. Ohata, "New approach to the design of crystal oscillators", Proc. of the 28th AFCS (1974).

[5] A. Benjaminson, "The design and analysis of VHF/UHF crystal oscillator", Proc. 41st AFCS (1987).

[6] B. Parzen, A. Ballato, "Design of crystal and other harmonic oscillators", Wiley Intersciences (1983).

[7] T. Adachi, M. Hirose, Y. Tsuzuki, "Computer analysis of Colpitts crystal oscillator", Proc. 39th AFCS (1985).

[8] J. Goldberg, "A simple way of characterizing high Q oscillators", Proc. 42nd AFCS (1988).

[9] T.M. Hall, "Computer aided design and assembly of oscillator", Proc. 36th AFCS (1982).

[10] S. Wolfram, "Mathematica : System for doing mathematics by computer", Adison Wesley (1988).

[11] R. Bogen *et al.* "MACSYMA reference manual", version ten. The Mathlab Group Laboratory for Computer Science, MIT (1983).

[12] A.C. Hearn, "REDUCE user's manual", version 3.3. The Rand Corporation (1987).

[13] J. Crabol, "Transfert de chaleur", vol. 1 et 2, Masson (1990).

[14] R. Gibert, "Transmission de la chaleur", coll. Génie chimique, vol. 11, Eyrolles (1963).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

OSCILLATOR INSIGHTS BASED ON CIRCUIT Q

Roger L. Clark

Pennsylvania State University - York Campus
1031 Edgecomb Ave
York, PA 17403

(717) 771-4087

Abstract - The concept of negative circuit Q is introduced as a basic oscillator characteristic necessary for signal growth. A limit cycle at steady state causes the circuit Q to increase asymptotically towards infinity and finally transition to positive Q. The introduction of noise forces the loop gain to be less than unity and the steady state Q to be finite and positive. Circuit Q is then principally determined by a limiting relationship on loop gain. Examination of a Norton equivalent circuit for a feedback oscillator indicates that positive and negative Q branches are present, and that the closed-loop dissipation factor is nearly zero. By considering the oscillator to be a tuned noise amplifier, the steady state circuit Q is found to be a determining factor in several areas of oscillator characteristics, including phase noise and frequency modulation. Fundamental limitations on modulation frequency for voltage controlled oscillators are examined, with experimental data indicating the potential for non-symmetrical sidebands for modulation frequencies greater than $f_0/2Q_L$.

I. Introduction

Oscillator design is typically based on either Barkhausen's criterion of unity loop gain and zero phase shift, or negative circuit resistance. It is more common for audio and RF oscillators to be analyzed with the former, while microwave oscillators are often analyzed under the premise of negative resistance, or its equivalent, a reflection coefficient which exceeds unity. Since many feedback paths may be present for microwave oscillators, the negative resistance approach is often preferred.

More recently, Robins presented an analysis of phase noise based on the premise that an oscillator is an extremely high gain amplifier which amplifies the residual noise in the circuit [1]. It is rather interesting in Robins' treatment that although the principle of the oscillator mechanism differs from earlier approaches, the resulting equations are essentially equivalent to those previously developed.

The premise that an oscillator is merely an extremely high gain noise amplifier is further developed in this paper. Some relationships between the negative resistance and feedback models are developed. Starting conditions are first examined and the analysis is then extended to steady state. An implicit assumption is that feedback oscillators can be represented as negative resistance oscillators (see Appendix I).

II. Negative Circuit Q at Start-Up

Nearly every introductory circuit analysis text [2] considers the damped RLC circuit in series and parallel resonant form as shown in Figure 1.

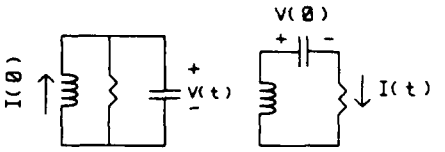


Figure 1. Series and Parallel Resonant Circuits with Initial Conditions

For the series resonant circuit

$$i(t) = e^{-\alpha_s t} \cdot A\cos(\omega_o t) \qquad (1)$$

where $\alpha_s = R/2L$.

The corresponding equation for parallel resonance is given by

$$v(t) = e^{-\alpha_p t} \cdot A\sin(\omega_o t) \qquad (2)$$

and $\alpha_p = 1/2RC$.

By considering energy loss, the corresponding time constants in each case can be written as

$$\alpha_s = \alpha_p = \frac{\omega_o}{2Q} \qquad (3)$$

It is interesting that this is simply the half-bandwidth of the circuit in radians per second. If the net circuit resistance is negative, it follows that the Q will be negative and that the signal envelope will grow in time. This simply infers that the negative resistor is supplying power rather than dissipating power. From the standard definition of Q, namely

$$Q = 2\pi \cdot \frac{\text{energy stored}}{\text{energy lost per cycle}} \qquad (4)$$

it becomes obvious that if AC energy is being acquired, then the net circuit Q must be negative at start-up. Since energy is also growing in a relaxation oscillator, the Q must also be negative, even if there is no discrete frequency with zero phase shift around the loop! A simple example of this would be the common R-C gate oscillator.

It should be noted that the bandwidth of a negative Q circuit is exactly the same as a positive Q circuit. For illustration, Figure 2 shows the frequency response of two circuits with positive and negative values of Q. The resonant frequency was arbitrarily chosen to be 10 MHz, and the simulation was performed with a commercially available circuit simulator [3].

CH2965-2/91/0000-352 $1.00 © 1991 IEEE

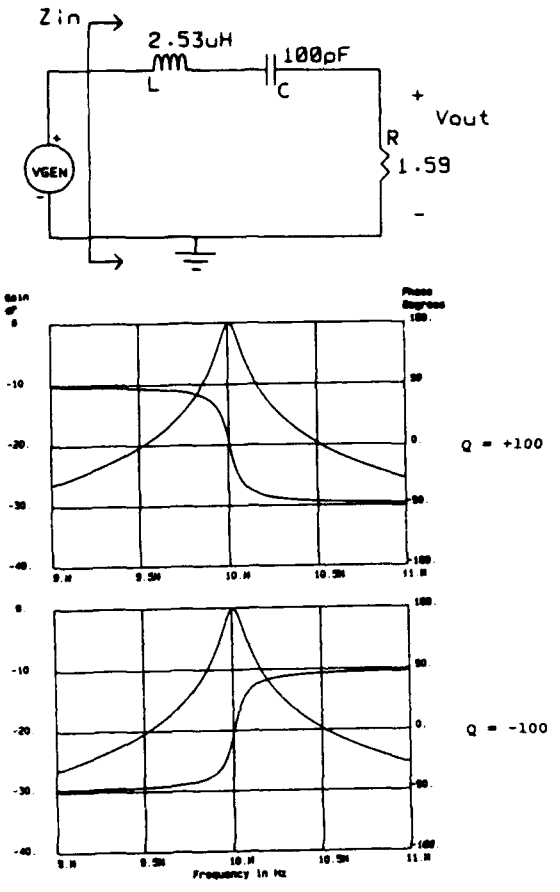


Figure 2. Frequency response for resonant RLC circuit. To obtain the values of Q, the sign of R was changed from positive to negative.

Although the magnitude of the frequency response of the circuits is identical, the time response and impedances of these circuits are quite different. Although this may seem peculiar at first, recall that the differential equation of a simple RLC circuit with a driving source has both a complementary and a particular solution. Phasor based circuit analysis in essence only describes the particular solution, and it is assumed that the complementary solution which describes initial transients has died out. This will clearly not be the case for a negative Q circuit.

Figure 3 shows the transient response for these circuits with a single input pulse. To speed up the simulation times, the pulse width was chosen to be half of the period of the natural resonant frequency of the circuit. Rather than letting numerical noise initiate oscillation, the injection of a narrow pulse into the circuit merely speeds up the cycle initiation.

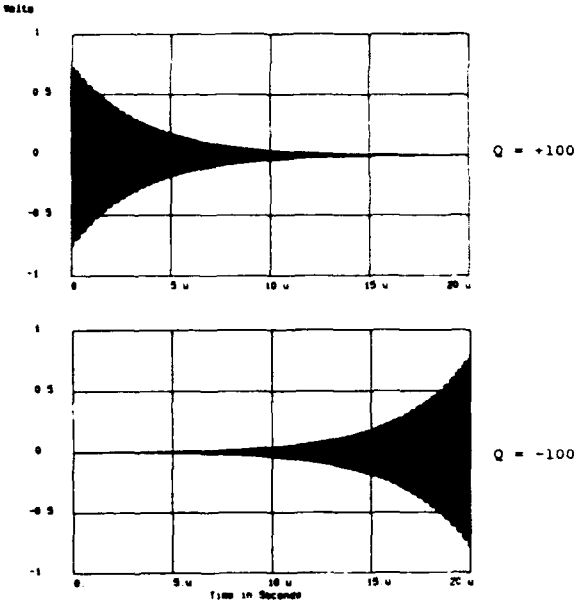


Figure 3. Transient response of positive and negative Q circuits.

Also note that the driving point impedance of a circuit containing negative resistance can traverse out of the usual range of +/- 90 degrees. For a simple resonant circuit, the change from negative to positive resistance changes the impedance angle from zero degrees to 180 degrees at resonance, as shown in Figure 4.

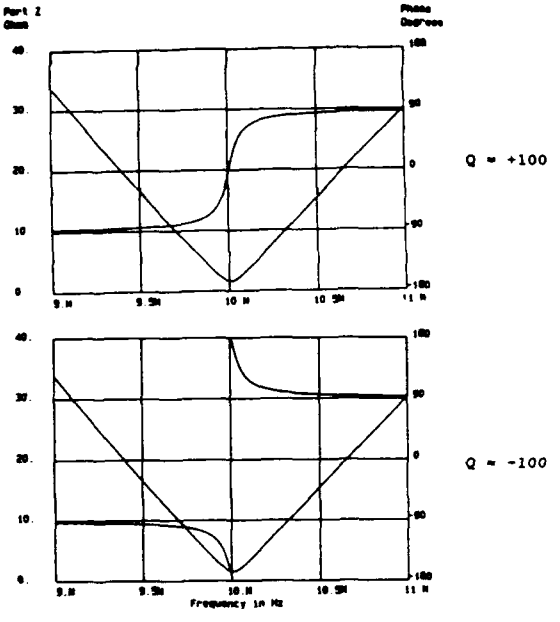


Figure 4. Impedance through series resonant circuit.

Note that all of the above plots were made simply by entering negative resistance values into the circuit simulator, and that similar characteristics can be obtained for any circuit exhibiting negative Q.

In a typical series resonant oscillator, the negative resistance is approximately 2 to 10 times the value of the positive resistance. This condition is required to ensure negative circuit Q and proper start-up. Therefore the closed-loop circuit bandwidth during the starting phase will

generally be of the same order of magnitude as the open-loop bandwidth. It is this finite bandwidth which allows the resonant circuit to acquire the energy needed for signal growth.

A few other observations can be made about the negative Q of a quartz oscillator. It has been empirically found that the maximum Q frequency product of acoustic resonators is defined by a constant. For example, the Q·f product for AT quartz resonators is approximately $16 \cdot 10^{12}$ [4]. From this relationship, the oscillator time constant for series resonant oscillators is found to be

$$\alpha_s = \frac{\pi f^2 (R_1 + R_{ext})}{R_1 \cdot Q \cdot f} \tag{5}$$

where R_1 is the crystal motional resistance and R_{ext} is the total oscillator loop resistance external to the crystal. As the circuit enters a limit cycle, the value of R_{ext} will change so that the rate of growth will become a function of signal amplitude and will be reduced.

Of course, one could apply this equation to parallel resonant oscillators by substituting the positive and negative values of conductance. One implication of this is that if rapid start-up is required, as is often the case with clock oscillators used in digital circuits, it may be advantageous to choose a high frequency for the oscillator and divide the output to obtain the desired frequency.

III. Additive and Multiplicative Noise

Before considering the oscillator at steady state, it will be beneficial to review additive and multiplicative noise, as the fundamental characteristics of the two noise mechanisms are very different [5],[6]. Additive noise has the property that the signal power can be increased without an attendant increase in noise level. Thus, the signal-to-noise ratio may be improved by techniques such as impedance matching, which increase the circulating power in the oscillator. In general, efforts to improve noise performance with regard to additive noise often become a matter of literally overpowering the inherent noise of the circuit. Additive noise is usually dominant in the noise floor area.

In contrast, multiplicative, or modulation, noise cannot be improved by increasing the signal power as the noise level is directly dependent on the signal level. A simplified case involving additive and multiplicative noise sources is shown in Figure 5, where the time varying capacitance can modulate both the phase and amplitude of the carrier signal. At resonance, phase fluctuations will be much larger than amplitude fluctuations. Clearly, increasing the signal power will improve the noise performance with respect to the additive noise source, but will offer no improvements with respect to modulation noise.

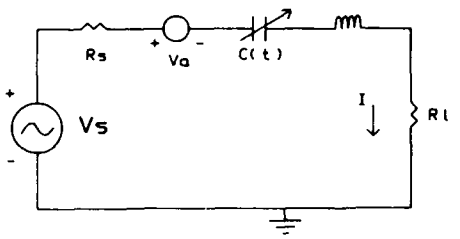


Figure 5. Ideal signal source with additive and multiplicative noise sources.

Summing the voltages around the loop for this circuit gives

$$V_s = L\frac{di}{dt} + \frac{1}{C(t)}\int i\,dt + i(R_s + R_L) + V_a \tag{6}$$

Although this is a linear differential equation, the presence of time varying coefficients significantly alters (and complicates) the equation. The capacitor will effectively serve to modulate the phase of the instantaneous voltage across its terminals such that the resonant frequency becomes a function of time. Clearly, the capacitor will act as a modulator and will induce noise with multiplicative characteristics into the circuit. Since any of the circuit elements could be time varying, one could generalize on this to say that multiplicative noise is not necessarily a result of circuit nonlinearities. As noted by Bell [7], a weakness of Kurokawa's noise analysis [8] is the neglect of fluctuations in reactive circuit elements. It would seem that at least two possible mechanisms for low frequency flicker noise to be imposed onto the carrier are 1) parametric modulation of circuit elements by low frequency noise sources (ie. modulation of semiconductor junction capacitances) and 2) mixing of low frequency noise as a result of non-linear operation of the oscillator. In an analysis of FET oscillators, Siweris and Schiek [9] found that phase fluctuations were entirely attributable to gate-source capacitance fluctuations, while nonlinear device transconductance was responsible for amplitude fluctuations [8].

Low frequency flicker noise (actual frequency, not offset frequency) is often considered to be a result of modulation processes and therefore exhibits multiplicative properties. For example, the flicker noise in resistors is a result of resistance fluctuations [14] and is not necessarily a consequence of current flow. Current flow only provides a means of detecting the resistance fluctuations.

Resonator flicker noise is certainly not additive, since higher power does not improve noise performance. Clearly, the resonator also cannot supply the oscillator with power, but can only modulate the frequency of circulating power.

It is interesting to note that several expressions for oscillator phase noise appear in the literature. For example, Leeson proposed an oscillator noise model where

$$S_{\phi\,out} = S_{\phi\,in}\left[1+\left(\frac{f}{2\cdot Q_L\cdot f}\right)^2\right] \tag{7}$$

One commonly cited [10],[11],[12] extension of Leeson's model which takes into account upconverted flicker noise is

$$\mathcal{L}(f) = \frac{FkT}{2P_{in}}\left[1 + \frac{f_\alpha}{f} + \frac{1}{f^2}\left(\frac{f_o}{2Q_L}\right)^2 + \frac{f_\alpha}{f^3}\left(\frac{f_o}{2Q_L}\right)^2\right] \tag{8}$$

while other sources [13] give a form similar to

$$\mathcal{L}(f) = \frac{FkT}{Pin}\left[1 + \frac{1}{f^2}\left(\frac{f_o}{2Q_L}\right)\right] + \alpha\left[\frac{1}{f} + \frac{1}{f^3}\left(\frac{f_o}{2Q_L}\right)^2\right] \tag{9}$$

where α is the flicker noise constant and f_α is the residual flicker noise corner frequency.

In the first case, it might seem that all noise sources are considered to be additive, while in the second case thermal noise is additive and flicker noise is multiplicative. If, however, the flicker noise corner frequency becomes a function of power [13], so that

$$f_\alpha = \alpha P_{in}/FkT \tag{10}$$

then the equations reconcile, with the exception of the factor of two (3dB), and all flicker noises are multiplicative. Later refinements of eq. (9) which reconcile the 3 dB factor are found in ref. [6].

The residual noise corner frequency (offset frequency) is generally a function of the signal power level, and the one Hertz intercept point is a device constant. One should then not expect the low frequency corner frequency, f_c, and the residual noise corner frequency, f_α, to necessarily be equivalent.

As the physical mechanisms of flicker noise are not entirely clear, additional research into the physics of flicker noise and the mechanisms by which low frequency noise is imposed onto the carrier needs to be performed.

IV. The Oscillator at Steady State with Additive Noise

As the amplitude of oscillation grows, circuit limiting begins to occur and the closed-loop Q asymptotically approaches infinity from the negative direction. The impedance values at start-up change as the oscillator enters the limit cycle such that the positive and negative resistive portions approach balance. In addition, it must be expected that reactive elements may also change under large signal conditions so that some frequency change will occur. Element values used in the steady state model must

change to reflect the steady state values.

As noted by Robins [1], an oscillator can be considered to be a very high-gain tuned amplifier. The positive feedback produces a nearly infinite Q which amplifies any noise present in the circuit. Robins used this principle to show that a prediction of phase noise which is effectively equivalent to what is obtained from Leeson's model [15] can be derived by standard circuit analysis. Parzen later extended this principle by introducing the concept of differential resistance [16], and showed that any noises present in the circuit cause a slight imbalance between the positive and negative circuit resistances. Differential resistance was also inferred by Kurokawa as a consequence of any changes in oscillation amplitude [8]. Note that, unlike Leeson's model, these models did not impose the restriction that phase shift around the loop be zero for very short periods of time. As will be seen, this has significant implications for multiplicative noise sources as well.

Hence, in physical oscillators some small differential resistance must be present, and the steady state Q must be positive. A net positive circuit Q also implies that any circuit behavior due to the negative Q at start-up will eventually decay (ie. the homogeneous solution of the differential equation will tend to zero for long time periods).

One significant implication of these analyses is that simple circuit analysis can be used to model the steady state oscillator with additive noise if provisions for nonlinearities are allowed.

A simple series resonant oscillator with a noise generater can be used to illustrate this effect. Noise sources at the input of the amplifier are referred to the output as shown in Figure 6.

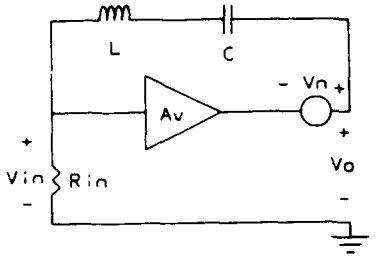


Figure 6. Steady State Oscillator model

Summing voltages at the output yields

$$V_o = V_n + A_v V_{in} , \tag{11}$$

where

$$\frac{V_{in}}{V_o} = \frac{R_{in}}{R_{in} + j(\omega L - 1/\omega C)} = \frac{1}{1 + jQ\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)} , \tag{12}$$

but $\omega/\omega_o \approx \omega_o/\omega \approx 2f_m/f_o$, where $f_m = f - f_o$, so that

$$\frac{V_{in}}{V_o} = \frac{1}{1 + jQ2f_m/f_o} = A_v . \tag{13}$$

Substituting this into eq. (11) gives

$$V_o = V_n + \frac{A_v V_o}{1 + jQ2f_m/f_o} \tag{14}$$

As $f_m \to 0$, we find

$$A_v = 1 - V_n'/V_o' \approx 1, \tag{15}$$

where V_n' is the residual noise at the carrier and V_o' is the output voltage signal. Note that as noise is introduced into the circuit, the loop gain decreases to a value below unity. Finally we obtain

$$\frac{V_o(f)}{V_n(f)} = \frac{1 + jQ2f_m/f_o}{1 - A_v + jQ2f_m/f_o} \tag{16}$$

As $f_m \to 0$, the output becomes quite sensitive to the voltage gain, or the residual voltage noise. This expression effectively gives equivalent results to the restatement of Leeson's model by Parzen [17]. In a practical sense though, the noise very close to the carrier is determined by a number of factors such as resonator flicker noise, thermal fluctuations, and other effects. Note that if A_v is assumed to be unity and the voltage gain is converted to power gain, eq. (16) reverts to the standard form of Leeson's expression given by eq. (7).

The above limiting condition on voltage gain can also be viewed as a limiting between the positive and negative circuit resistance. Consider the circuit of Figure 7.

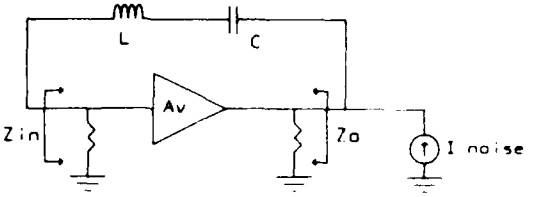


Figure 7. Feedback oscillator with noise current.

Of course, the noise current could be transformed into a noise voltage at the frequency of oscillation, but the current is simply more convenient in this case. In a more practical oscillator, both voltage and current noise sources will be present so that noise power will vary with offset frequency. As derived in Appendix I, the impedance looking into the output port of the amplifier with the load removed will be given by

$$Z = \left\{ \frac{1}{Z_o} + \frac{1}{Z_f + Z_{in}} - \frac{A_v Z_{in}}{Z_o(Z_f + Z_{in})} \right\}^{-1} \tag{17}$$

An alternative realization of this is shown below for the case when Zin and Zo are resitive and equal, along with the noise current source.

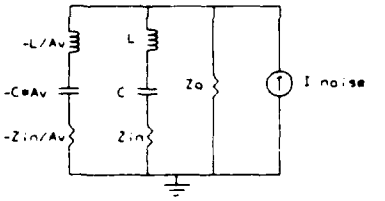


Figure 8. Negative Resistance Model of Feedback Oscillator from Figure 7.

Clearly, this circuit will exhibit a high impedance state at resonance. Also notice that the Q of the positive impedance branches of the circuit is nearly equal to the magnitude of the negative impedance branch. The positive Q is the same Q as one would obtain from the more conventional approach of using the open-loop phase slope (ie. the loaded Q in Leeson's model). It is simple to show that the Q of either a series or parallel resonant circuit having two resistive elements is obtained from the reciprocals of the Q for either element, so that it would be expected that the magnitudes of the positive and negative Q branches be nearly equal. That is to say, the closed-loop Q, Q_{cl} is obtained from

$$\frac{1}{Q_{cl}} = \frac{1}{Q_p} + \frac{1}{Q_n} \tag{18}$$

and it is known that the closed-loop Q is nearly infinite. Thus, it is seen that $Q_p \approx -Q_n$. Since dissipation factor is the reciprocal of Q, it is seen from eq. (18) that the closed-loop dissipation factor is nearly zero. In essence, noise signals within the 3 dB bandwidth of the negative Q branch are amplified by the negative resistance. For this simple circuit, the transition from white phase (flat) to white frequency (20 dB/decade) noise will be determined by $f_o/2Q_p$, which is the same result as obtained from Leeson's model. By including flicker noise sources, the 10 and 30 dB/decade slope regions can also be obtained.

Of course, if the negative resistance is exactly equal to the positive resistance, then the total circuit exhibits an infinite Q. Since this is clearly not the case, the finite signal power implies a limiting relationship on loop gain, or equivalently negative resistance. Parzen has used this concept to show that phase noise can be estimated without opening the loop by using the difference between positive and negative circuit resistance, termed dR, and the resulting high circuit Q.

It can now be seen that a negative resistance circuit can be used to estimate phase noise through use of Leeson's model, and that either the positive or negative Q may be used to determine the transition frequency of $f_o/2Q_L$. Noise at frequencies within this bandwidth are simply amplified by the negative circuit resistance. Although this approach does not explictly require that the closed-loop gain be determined, the loaded or positive Q must be determined. This may not be a trivial task in oscillators having multiple feedback paths, or in oscillators where circuit impedances are greatly disturbed by opening the loop.

Alternatively, given an intimite knowledge of noise sources in the circuit and the exact loop gain, one can obtain the noise performance by simple circuit analysis. It is not necessary to explicitly determine the loaded Q with this approach.

V. The Oscillator at Steady State with Multiplicative Noise

Consider again the simple series resonant circuit of Figure 5. This circuit will have a 3 dB bandwidth given by f_o/Q. An example of a circuit with a time varying resonance such as this would be a crystal which is being mechanically vibrated. As demonstrated by Driscoll, a resonator in vibration which is driven by a sinusoidal generator will induce phase modulation sidebands within the resonator bandwidth, and the sidebands will be attenuated by the resonator outside of the loaded bandwidth [18]. For the case of a single resonator, the sidebands will roll-off at approximately 20 dB/decade for frequencies well outside of the passband, as confirmed in the data of Driscoll.

It is important to see that if the frequency of the resonator is being modulated, the ensuing phase modulation becomes a function of the loaded Q, as described by Parker [6]. An alternative way of viewing this can be found in the equation

$$\Delta\phi = \frac{d\phi}{d\omega}\,\Delta\omega = 2\pi\tau_g\Delta F = 2Q_L\,\frac{\Delta f}{f_o} \qquad (19)$$

Increasing the loaded Q causes a proportional increase in group delay, so that the phase fluctuations increase. This must be so, since an increase in loaded Q cannot improve the stability of a circuit where the resonator frequency is fluctuating. Using eq. (19), the spectral density of phase fluctuations for a resonator is related to the spectral density of fractional frequency fluctuations by

$$S_{\phi R}(f) = (2Q_L)^2 S_{yR}(f) \qquad (20)$$

If the total open-loop phase fluctuations, S_ϕ', are caused by some combination of phase fluctuations, $S_{\phi E}'$, and frequency fluctuations, S_y', then

$$S_\phi' = S_{\phi E}' + 4Q_L^2 S_y' \qquad (21)$$

This leaves the dilemma of separating S_ϕ and S_y from measurements. By experimentally evaluating $\Delta S_\phi'/\Delta Q_L$, S_y' can be found from

$$\frac{\Delta S_\phi'}{\Delta Q_L \cdot 8Q_L} = S_y' \qquad (22)$$

so that the frequency fluctuations can be determined. Whether a particular device is predominantly a phase or a frequency modulator is essentially determined by the change in phase noise with respect to loaded Q.

Returning to the circuit of Figure 5, if the source and load resistances are reduced while maintaining constant phase fluctuations in the passband, then the circuit bandwidth will decrease. As shown in Figure 9, a family of curves can be generated as the circuit Q is varied.

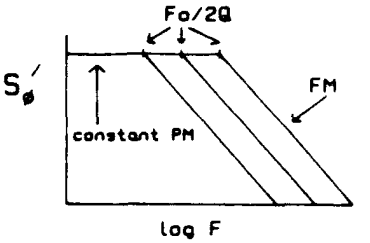


Figure 9. Frequency response with varying Q and constant phase modulation in the passband.

Note that while the S_ϕ is constant in the passband, the f^{-2} attenuation for frequencies outside of $f_o/2Q$ in effect yields fm outside of the passband. By including an amplifier in the circuit, an open-loop oscillator could also be characterized in the preceeding manner, with the amplifier serving as the phase modulator.

From eq. (20), it becomes apparent that the case of resonator frequency fluctuations (or any other frequency modulator) produces a different family than the case of constant phase fluctuations, as shown in Figure 10. While phase fluctuations inside the passband increase by Q^2, the reduction in bandwidth yields a constant fractional frequency fluctuation outside of the passband.

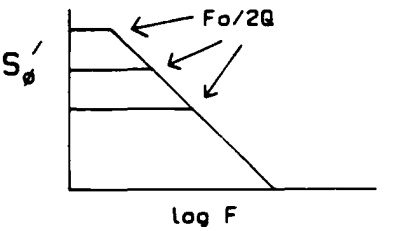


Figure 10. Frequency response with varying Q and constant frequency modulation of the resonator.

If the source and load resistors could be made arbitrarily small, the circuit Q would approach infinity and any phase modulation sidebands would roll off at 20 dB/decade for a single pole feedback network. Thus, in the limit as $Q \to \infty$, any phase modulation sidebands will be attenuated at 20 dB/decade, and will essentially appear as f_m sidebands outside of the passband. Of course, there is a physical limit as to how small the circuit resistances can become. In addition, to obtain any useful output power in the load, the circulating current will be impractically large. However, if the source resistance is allowed to take on negative values while the load remains positive, then the circuit Q can approach infinity and useful output power can be obtained. In addition, either phase or frequency modulation by loop components will show up as frequency modulation across the load due to the filtering. This is exactly the behavior which would describe an oscillator.

Thus, an oscillator can be viewed as the limiting case of a filter where circuit resistors can take on negative values such that the width of the passband approaches zero Hertz and the circuit is driven by noise generators. All equations for open-loop can now be applied to the closed-loop case if the large-signal gain is adjusted to produce the proper Q. In addition, white noise sources inside the loop generate signals which are filtered to produce a f^{-2} response, provided that the circuit has a single resonance, so that phase modulation is converted into frequency modulation.

This realization provides additional insight into other areas such as factors which limit the rate of modulation in oscillators and other oscillator phenomenon. For example, it becomes apparent that with a feedback network providing 20 dB per decade of attentuation, there is essentially no upper limit on the maximum frequency of modulation by an external source, as noted by Barnes [5]. Of course, this assumes that the modulation index is large enough that sidebands do not become masked by the intrinsic oscillator noise. However, if the feedback network has some other response, such as a spurious mode in the crystal, then some deviation from ideal frequency modulation will occur. Deviations from an ideal single pole feedback network in the oscillator can then be seen as a possible source of distortion in frequency modulation systems.

A vibrating crystal oscillator can be used to illustrate the effects of resonator frequency noise. For purposes of illustration, the vibration is assumed to be random with a uniform spectral density as shown in the upper plot of Figure 11. It is well known that the frequency of a vibrating crystal can be modulated by external mechanical vibrations [19]. The typical open-loop noise of a crystal resonator and amplifier under static conditions and in vibration would then be as shown in Figure 11. If the circuit Q is increased to a nearly infinite value, then the vibration will be seen to induce frequency modulation.

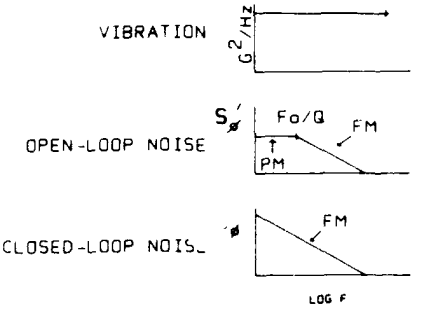


Figure 11. Vibration profile with corresponding open- and closed-loop noise.

Inside the open loop bandwidth, phase modulation by the resonator will be upconverted to frequency modulation, and outside of the loop, the phase modulation is attenuated by 20 dB per decade under closed-loop conditions. If the vibration has a f^{-1} characteristic such that the resonator exhibits flicker noise, then the corresponding closed-loop noise will roll off by f^{-3}, both inside and outside of the loop until the noise floor is reached. It would be expected that the same results would apply under static conditions, as well, for the case of resonator flicker noise.

It must be noted, however, that there appears to be some discrepency on whether flicker noise of a resonator outside of the loop bandwidth will drop off by f^{-2} or f^{-3}. Experimental data by Walls and Wainwright in 1974 shows a f^{-2} behavior [20] and more recently, Kroupa also suggests f^{-2} characteristics [21]. As was noted by Kroupa though, work done by Gagnepain and others would infer a f^{-3} response outside of the resonator bandwidth [22].

It is well known that a f^{-1} random vibration spectral density induces a -30 dB/decade response in the spectrum of an oscillator, both inside and outside of the open loop bandwidth [19]. However, unless the residual f^{-1} noise of a resonator exhibits different properties than resonator noise induced by f^{-1} mechanical vibration, one would expect a f^{-3} oscillator response for residual resonator flicker noise, both inside and outside of the open-loop bandwidth.

A simple experiment was performed to support the contention that an oscillator is merely a tuned noise amplifier, and that phase modulation sidebands are filtered by the extraordinarily high circuit Q to produce frequency modulation sidebands. A voltage controlled crystal oscillator (VCXO) was constructed with a second crystal and series resistor placed in parallel with the first crystal as shown in Figure 12. The series resistor was chosen to ensure that the circuit was oscillating only on the first resonator. Sidebands were measured as a function of the modulation frequency.

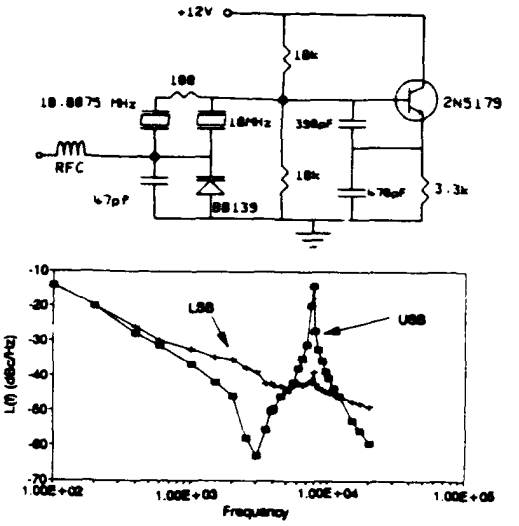


Figure 12.VCXO with Two Crystals and Measured Sideband Levels.

Although some departure from ideal fm occured in the lower sideband, the upper sideband clearly shows that the second resonator allows high amplification of the induced sideband by the circuit negative resistance. The frequency of high sideband amplification corresponds to the difference frequency between the parallel resonant frequencies of the two resonators. It should be noted that the second order sideband was also very pronounced when the modulation frequency was half the difference frequency. In fact, at a modulation frequency of 3800 Hz, the first upper sideband was -50.2 dBc, and the second order sideband was -38.2 dBc, located 7600 Hz above the carrier.

One would conclude from this that the mechanism of frequency modulation is more involved than simply maintaining zero phase shift around the loop at all times, and that the concurrent amplification and filtering of phase modulation sidebands may provide a realistic model of oscillator performance.

It also becomes apparent that the resonators for wideband VCXO's must be free of strong spurious modes within the modulation frequency bandwidth.

VI. Nonlinear Considerations

The preceeding discussion is based on linear models of the oscillator. As the oscillator enters into a limit cycle, nonlinear operation will result. Using the principle of harmonic balance, a small signal model of the oscillator can be contructed. The equivalent impedances can be determined from

$$Z_n = \frac{V_n}{I_n}, \qquad (23)$$

where n represents the harmonic number. The voltages and currents can be determined by performing a transient analysis followed by a Fourier transform on the non-linear circuit, or by other appropriate nonlinear techniques such as harmonic balance. The impedance values can then be used to form the small signal model for the fundamental signal used in subsequent analyses. Accuracy limitations may mandate some trimming on the amplifier gain to maintain the appropriate oscillation level. It is interesting to note that the circuit impedances at the various harmonics can be transformed to the time domain to obtain what is effectively sampled impedance in the time domain.

Finally, some limitations of this approach are noted. As shown by Grozkowski [23], the principle of reactive power balance implies that harmonic content may influence the frequency of an oscillator. Groszkowski's result is a consequence of conservation of energy, and does not consider that impedance values are affected by circuit limiting. This implies that even if the fundamental impedance values for nonlinear elements are used in a small signal analysis,

the total circuit impedances will not precisely balance. However, the effect is proportional to Q_L^{-2}, so that errors will be quite small for high Q circuits. In addition, other nonlinear effects such as squegging due to low frequency time constants may be present. These types of behaviors usually become immediately apparent when the real oscillator is constructed, and bias values or time contants are altered to eliminate the behavior.

VII. Summary and Conclusions

It has been shown that the total circuit Q of an oscillator is negative during the starting phase, and eventually transitions into an extremely large positive Q circuit during the limit cycle. Additive and multiplicative noises have been examined in feedback and negative resistance oscillators, and the transformation between feedback models and negative resistance models has been examined. By examining a simple RLC circuit, it is also seen that a continuous transition can be made between open- and closed-loop circuits by allowing circuit resistances to take on negative values.

References

[1] Robins, W.P., Phase Noise in Signal Sources, London: Peter Peregrinus Press Ltd, 1984.

[2] Hayt, W.H., Kemmerly, J.E., Engineering Circuit Analysis, McGraw-Hill, New York, 1978.

[3] ECA-2 Electronic Circuit Analysis, Version 2.60, Tatum Labs, 3917 Research Park Dr. B-1, Ann Arbor, MI, 48108

[4] Warner, A.W., "Design and Performance of Ultraprecise 2.5-Mc Quartz Crystal Units," Bell System Tech. J., 39, (Sept. 1960), pp. 1193-1217.

[5] Barnes, J., "Noise and Time and Frequency - A Potpourri," Proceedings of the 42nd Annual Symposium on Frequency Control", 1988.

[6] Parker, T.E., "Characteristics and Sources of Phase Noise in Stable Oscillators," Proceedings of the 41st Annual Symposium on Freqeuncy Control", 1987.

[7] Bell, D.A., Noise and the Solid State, New York: Wiley, 1985, pp. 119-122.

[8] Kurokawa, K. "Some Basic Characteristics of Broadband Negative Resistance Oscillator Circuits," Bell Syst. Tech. J., (48), 1969.

[9] Siweris, H.J., Schiek, B., "Analysis of Noise Upconversion in Microwave FET Oscillators," IEEE Trans. on Microwave Theory and Tech., vol. MTT-33, No. 3, March, 1985, pp. 233-241.

[10] Vendelin, G.D., Design of Amplifiers and Oscillators by the S-Parameter Method, New York: Wiley, 1982, pp. 145-161.

[11] Rohde, U.L., Digital Frequency Synthesizers, Prentice Hall, 1983, pp. 76-83.

[12] Rhea, R.W., Oscillator Design and Computer Simulation, Prentice Hall, 1990, pp. 82-85.

[13] Gerber E.A., Ballato, A., et. al., Precision Frequency Control, New York, Academic Press, 1985, pp. 72-75.

[14] van der Ziel, A., Noise in Solid State Devices and Circuits, New York: Wiley, 1986, pp. 145-175.

[15] Leeson, D.B., "A Simple Model of Feedback Oscillator Noise Spectrum," Proceedings of the IEEE, Vol. 54, No.2,pp.329-330, February, 1966.

[16] Parzen, B., "Universal, Computer Facilitated, Steady State Oscillator, Closed Loop, Analysis Theory," Proceedings of the 44th Annual Symposium on Frequency Control", 1990.

[17] Parzen, B., "Clarification and a Generalized Restatement of Leeson's Oscillator Noise Model," Proc. of 42nd Ann. Freq. Contr. Symp., 1988, pp. 348-351.

[18] Driscoll, M.M, "Quartz Crystal Resonator G Sensitivity Measurement Methods and Recent Results," IEEE Trans.Ultrason. Ferro. Freq. Contr., vol. 37, no. 5, pp. 386-392.

[19] Filler, R.L., "The Acceleration Sensitivity of Quartz Oscillators: A Review," IEEE Trans. Ultrason. Ferro. Freq. Cont., UFFC-35, No. 3, pp. 297-305, May, 1988.

[20] Walls, F.L., Wainwright, A.E., "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications," IEEE Trans. Instrum. Meas., vol. IM-24, pp. 12-20, Mar. 1975.

[21] Kroupa, V.F., The State of the Art of Flicker Frequency Noise in BAW and SAW Quartz Resonators," IEEE Trans. on Ultrason. Ferro. Freq. Contr., vol. 35, no. 3, May, 1988, pp. 406-420.

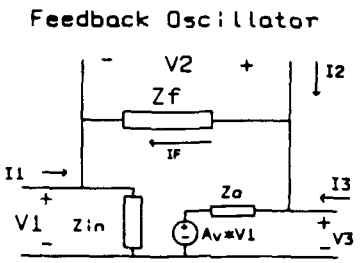
[22] Gagnepain, J.J., "Fundamental Noise Studies of Quartz Crystal Resonators," Proc. 30th Ann. Freq. Contr. Symp., 1975.

[23] Grozkowski, J., Frequency of Self Oscillations, MacMillan, NY, 1964, pp. 207-231..

Appendix I.

Negative Impedances in a Simple Feedback Oscillator.

If reverse transmission effects are included in the feedback network, then a simple oscillator can be described as shown below. All elements, including the voltage gain, can be complex. It is desired to determine the impedance looking into all three ports - input, output, and feedback. Test currents are independently applied to each port to determine the impedance (ie. i2 and i3 are zero when i1 is applied).



Summing voltages around the loop yields

$$V_1 = (i_1 + i_f)z_{in} = A_v V_1 - i_f(Z_o + Z_f)$$

and

$$i_f = \frac{-V_1 + A_v V_1}{Z_o + Z_f} ,$$

so that

$$\frac{V_1}{i_1} = Z_1 = \frac{Z_{in}(Z_o+Z_f)}{Z_o+Z_f+Z_{in}(1-A_v)} .$$

Clearly, the impedance can be made negative with sufficient gain. Looking into the feedback port gives

$$V_1 + V_2 = A_v V_1 - Z_o\left[\frac{V_2}{Z_f} - i_2\right].$$

Using $V_1 = V_2 Z_{in}/Z_f$ yields

$$\frac{V_2}{i_2} = Z_2 = \frac{Z_o}{1 + \frac{Z_o}{Z_f} + \frac{Z_{in}}{Z_f}\left[1-A_v\right]} ,$$

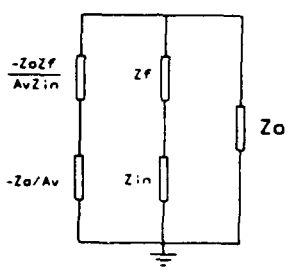
while summing currents at the output port gives

$$i_3 - \frac{V_3}{Z_f+Z_{in}} + \left[\frac{A_v Z_{in} V_3}{(Z_{in}+Z_f)Z_o} - \frac{V_3}{Z_o}\right] = 0$$

Rearranging this to obtain the impedance

$$\frac{V_3}{I_3} = Z_3 = \left\{\frac{1}{Z_o} + \frac{1}{Z_f+Z_{in}} - \frac{A_v Z_{in}}{Z_o(Z_f+Z_{in})}\right\}^{-1}$$

An alternative realization of Z3 is shown below.



It is interesting that the typical oscillator configurations can be obtained with proper placement of circuit elements.

Oscillator	Z_{in}	Z_f	Z_o
Pierce	capacitor	crystal	capacitor
Colpitts	crystal	capacitor	capacitor
Butler	resistive	crystal	resistive

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ANALYSIS OF START-UP CHARACTERISTICS OF CRYSTAL OSCILLATORS

Jin-Qin Lu, LSI Design Division,
SONY LSI Design Corporation, Tokyo 141, Japan
Yasuo Tsuzuki, Division of Electrical & Computer Engineering,
Yokohama National University, Yokohama 240, Japan

Abstract

In this paper, an analysis method of the start-up characteristics of crystal oscillators has been presented. First, the origin causing oscillation to start just after *d.c.* power supply is switched on has been discussed by means of a computer-aided analysis and a small signal equivalent circuit analysis. Then, this analysis method, suitable for practical uses, has been outlined which can calculate both the overall start-up characteristics of the amplitude and frequency of oscillation. Finally, experimental results have been given to verify the usefulness of this analysis method.

Introduction

Many investigations on the start-up behavior of crystal oscillators from the initial oscillation to the steady-state oscillation have been made from the age of vacuum tubes. It has been shown by experiments that the amplitude of oscillation increases in an exponential manner. In one of the publications of IEC (International Electrotechnical Commission), an approximation formula has been given to calculate the start-up time [1]. The formula also indicates that the amplitude of oscillation increases exponentially, however, the method of determining the initial and the steady-state amplitude of oscillation has not been shown in it. Consequently, this formula can not be applied to the calculation of the start-up time. Since the origin causing oscillation to start has not yet been clarified, the calculation of the overall start-up characteristics, including both the amplitudes and frequencies of oscillation, can not be done in principle.

It would be possible to calculate the overall start-up characteristics by using general-purpose circuit simulators such as *SPICE*. However, the total computation time is vast because the quality factor Q of a crystal resonator is so high that the number of period of oscillation up to the steady-state is even over the order of Q. Accordingly, this calculation method is not suitable for practical uses. In this paper, we attempt to develop a simple analysis method of the overall start-up characteristics.

In the first section, we discuss the start-up behavior, especially the origin causing oscillation to start, i.e. oscillation triggering, by using the simulator *SPICE* and a small signal equivalent circuit analysis method. In the next section our analysis method of the overall start-up characteristics is outlined. In the last section, experimental results to verify the usefulness of present analysis method are described.

Start-up Behavior of Oscillator

The oscillator circuit configuration considered here is the very common *Colpitts* crystal oscillator using a bipolar transistor, and is shown in Fig. 1. The left side of the terminals

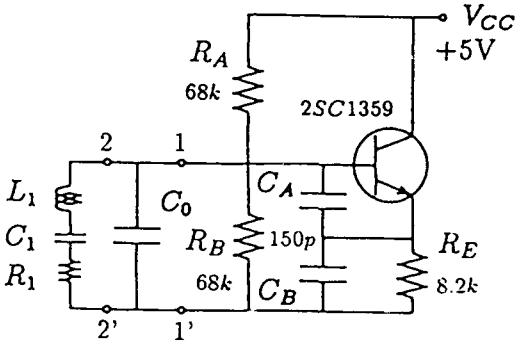


Figure 1: *Colpitts* crystal oscillator circuit.

1-1' in Fig. 1 is a usual equivalent circuit of a crystal res-

CH2965-2/91/0000-360 $1.00 © 1991 IEEE

onator with parameters given in Table 1. The *Gummel-Poon* model for bipolar transistors is used in circuit simulation.

Table 1: Equivalent circuit parameters of crystal resonator.

$f_s[kHz]$	$L_1[H]$	$C_1[pF]$	$R_1[\Omega]$	$C_0[pF]$
10,017	1.87×10^{-2}	1.35×10^{-2}	11.9	2.75

The overall start-up behavior of oscillation in this circuit is evaluated by means of a computer-aided analysis method. The start-up behavior of the oscillation current I through the terminals 2-2' is analyzed by using *SPICE* with a usual transient analysis option. In this analysis, *d.c.* supply is treated as an ideal step voltage without any random noise sources which are usually considered as oscillation triggers. The simulation result of the oscillation current I is shown in Fig. 2. As shown in this figure, the oscillation amplitude

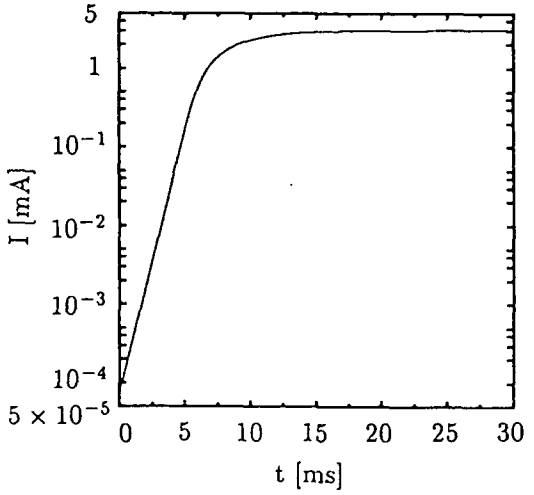


Figure 2: Start-up characteristic of the oscillation current of motional arm of crystal resonator.

suddenly builds up to a certain value when *d.c.* supply is turned on. Then, the amplitude of oscillation I increases in an exponential manner, in other words, it increases linearly in logarithmic scale. Finally, the oscillation amplitude gradually approaches the steady-state level. It took very long run-time to obtain this simulation result. For instance, it costed about 7 hours of CPU time for a 20 *MIPS* workstation to get only one point data at 16 [ms]. Therefore, usual transient analysis methods for the calculation of the overall start-up characteristic is obviously not suitable for practical uses.

In order to clarify the origin causing oscillation to start, the oscillation current I and the voltage across the resonator were evaluated in the first several oscillation periods with relatively small steps in transient analysis of *SPICE*. According to this simulation results, it has been shown that there appeared a step-like voltage across the resonator when the initial oscillation current I begun to oscillate, just after *d.c.* supply is turened on as seen in Fig. 2. The rise time of the step-like voltage is even as short as the order of 10^{-10} second.

In the well-known *Gummel-Poon* model for bipolar transistors, there are only two junction capacitances on base-to-collector and base-to-emitter and a base-spreading resistance before an operating point of a transistor builds up. Therefore, the circuit shown in Fig. 2 can be equivalently converted to the circuit shown in Fig. 3 immediately after the application of *d.c.* supply. Usually, the junction capaci-

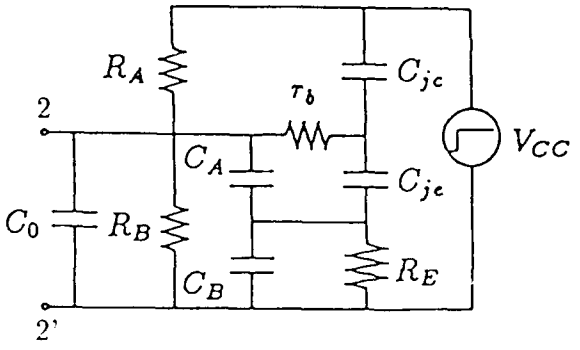


Figure 3: Equivalent circuit immediately after the application of *d.c.* supply.

tances of a transistor for high-frequency use are so small that the junction capacitances are charged much faster than other capacitances in the oscillator circuit, and hence, the step-like voltage appears across the resonator. Then the motional arm of the equivalent circuit of the resonator is triggered simultaneously by this step-like voltage and the current I starts to oscillate and remains constant due to the high Q value of the resonator untill the operating point of a transistor is established. Both of the rise time and the value of the step-like voltage can be estimated using this small signal equivalent

circuit shown in Fig. 3, and therefore, the initial amplitude of oscillation current I can be calculated using the motional arm parameters, i.e. the LCR equivalent circuit of the resonator with the applied step-like voltage across it.

Based on these discussions, it has been shown clearly that *d.c.* supply switching is the most important oscillation trigger in this circuit configuration and a circuit simulation can be performed without any random noise sources in the oscillator circuits.

Outline of Analysis Method

As shown in Fig. 4, a crystal oscillator can be equivalently converted to a series LCR circuit where L is the motional inductor L_1 of the resonator, C and R are determined by the parameters of resonator C_1, R_1, and the active circuit's negative resistance $-R_L$ and equivalent capacitance C_L containing the shunt capacitance C_0 of the resonator. From the

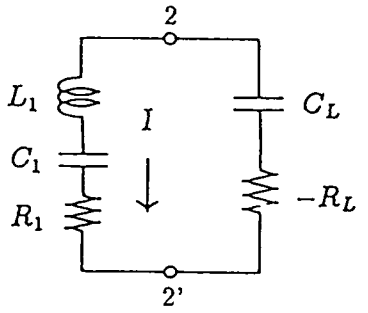


Figure 4: Equivalent circuit of crystal oscillator.

well-known property of the series LCR circuit, the start-up time, denoted as t, is given as follows;

$$t = \frac{2L_1}{R_L - R_1} \ln \frac{I_s}{I_0}, \tag{1}$$

where I_0 is the initial amplitude of the oscillation current I and I_s is its amplitude at the steady-state. The IEC publication [1] gives the same formula for calculation of the start-up time. However, due to the nonlinearity of a transistor, both magnitude of the equivalent capacitance C_L and that of the negative resistance $-R_L$ change with oscillation signal level, i.e., the level of the oscillation current I. Therefore, the start-up time calculated by Eq. (1) is valid only in the linearly increasing region as seen in Fig. 2. In a relatively short time interval Δt, however, this LCR circuit can be treated as a linear circuit since Q is so high. In other words, both the equivalent capacitance C_L and the negative resistance $-R_L$ can be considered approximately as linear elements in such a short time interval. Therefore, the increment ΔI of the oscillation current I in this circuit cab be calculated by using Eq. (1) in this short time interval Δt as shown below;

$$\Delta t = \frac{2L_1}{R_L(I) - R_1} \ln \frac{I + \Delta I}{I}, \tag{2}$$

where the negative risistance $-R_L(I)$ is a function of the current level I only and not of the increment ΔI. Furthermore, by employing the relation between the resonator current level and the negative resistance, the overall amplitude characteristic from the start-up to the steady-state can be calculated by integrating the contributions in every short time interval. The initial amplitude of the resonator current I can be calculated by using a conventional simulator, and the level of the resonator current I_s at the steady-state can be calculated from the condition that the negative resistance $-R_L$ is equal to the equivalent resistance R_1 of the resonator [2].

On the other hand, the start-up characteristic of oscillation frequency can easily be obtained by the above-mentioned idea. Here, the calculation formula for frequency is given as follows;

$$\omega_f = \sqrt{\frac{1}{L_1 C'(I)} - \sigma^2} \tag{3}$$

where

$$C'(I) = \frac{C_1 C_L(I)}{C_1 + C_L(I)}, \qquad \sigma = \frac{R_L(I) - R_1}{2L_1},$$

and the equivalent capacitance $C_L(I)$ is a function of the current level I.

Experimental Results

The analysis method presented in the above section was applied to two circuit examples designed by different policies. Circuit no. 1 was designed to have a small margin of the nagative resistance $(-r_L)$ at small signal levels, while circuit no. 2 is opposite. The two circuits have the same oscillation current I_s at the steady-state to see the differences of the start-up characteristics between them. Their circuit parameters are listed in Table 2 where R_A, R_B and C_A, C_B are set

Table 2: Two circuit design examples.

No.	R_A [$k\Omega$]	R_E [$k\Omega$]	C_A [pF]	I_c [mA]	$-r_L$ [Ω]
1	68	8.2	150	0.21	−69
2	4.7	1.8	75	0.35	−325

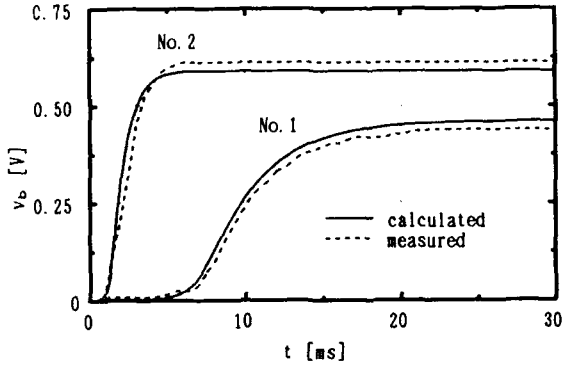


Figure 5: Start-up characteristics of voltage across the terminal of resonator of the two circuits.

to be equal for design convenience and I_c is a d.c. operating point of the transistor.

By using the present analysis method, the overall start-up characteristics of the oscillation current I in these two circuits were easily obtained only in about 10 minuts of CPU time including the calculation time of the charactristics of the current level I versus the equivalent capacitance $C_L(I)$ and the nagative resistance $-R_L(I)$. The relative errors of time were under 1% compared with the results by usual transient analysis methods at the same current level shown in Fig. 2.

Comparision between experiment and calculation was made, and the results are shown in Fig. 5. The curves show the amplitude of the voltage v_b across the terminal of the resonator by multiplying the amplitude of the current I by the input impedence of the terminals 2-2' shown in Fig. 4. Good agreement between experiment and calculation demonstrated that the present analysis method is quite effective for practical uses.

Evaluation of the overall start-up characteristics of oscillation frequency of the two circuits was also achieved by using Eq. (3). The results are shown in Fig. 6 where the ordinate describes the relative variation of start-up frequency to the frequency at the steady-state.

It can been seen that the oscillation frequency of circuit

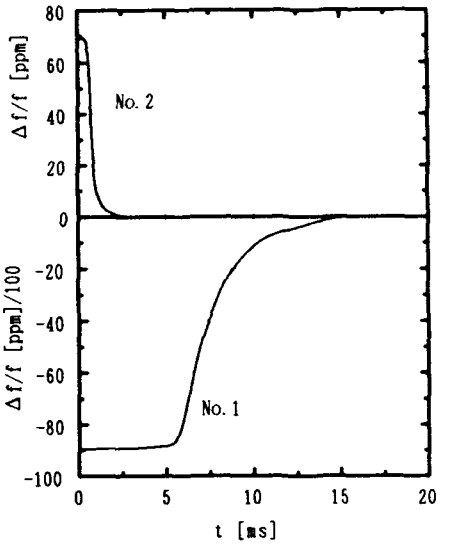


Figure 6: Start-up characteristics of oscillation frequency of the two circuits.

no. 2 approaches the steady-state much faster than that of circuit no. 1. In contrast to these frequency changes, the maximum frequency variation of circuit no. 1 is only about 1 [ppm], 70 times lower than circuit no. 2's variation.

Conclusions

In this paper, a simple analysis method of the overall start-up characteristics of crystal oscillators has been desribed. After clarifying the origin causing the oscillation current to start by means of circuit simulations and a small signal circuit analysis, the analysis method, empolying the relation between the oscillation current and the nagative resistance and the equivalent capacitance, has been developed. It has been demonstrated that the present analysis method can easily calculate the overall start-up characteristics of the amplitude and frequency of oscillation, and can reduce the total computation time significantly, while maintaining good precision in calculation for practical uses.

References

[1] International Electrotechnical Commission, "Quartz crystal units for frequency control and selection," IEC Standard Publication 122-2, Second Edition, 1983.

[2] T. Adachi, M. Hirose and Y.Tsuzuki;"computer analysis of Colpitts Crystal Oscillator" , Proc. of 39th AFCS, pp.178-182, May, 1985.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

OSCILLATOR DESIGN USING S PARAMETERS AND A PREDETERMINED SOURCE OR LOAD

R. J. Weber

Department of Electrical Engineering and Computer Engineering
Iowa State University
Ames, Iowa 50011

Abstract

There are an infinite number of terminal impedances which satisfy the general oscillation condition $|S\ S' - I| = 0$ for a three terminal device which is operated below it maximum frequency of oscillation. S is the linearized device three port scattering matrix characterized or measured at some specified terminal condition (e.g. power level) (1) and S' is the scattering matrix of the embedding network for the three terminal device. The design engineer almost always has some specification that is to met with the oscillator. For example, maximum power output, minimum noise shelf, or maximum device gain for the amplifying device in the oscillator may be specified. This paper gives a closed form solution to the oscillation condition under the conditions of a specified impedance presented to either the source or the load port of the amplifying device.

Previous design approaches to S parameter design have generally not specified a port operating impedance (e.g. 2,3) or have required iteration or root finding to determine the impedances (4).

Introduction

Consider the two port device shown in Figure 1. The device designated by S could be a bipolar, FET, or other two port device with gain. If the unsaturated device is to develop the most power (i.e. maximum DC to RF conversion), the equivalent parallel resistance R of the load conductance in the linear range is determined by the mutual satisfaction of the following two conditions (5).

```
sqrt(R) = (V - Vsat)/sqrt(2.* P)
```

and

```
sqrt(R) = sqrt(2. * P)/(I - Imin)
```

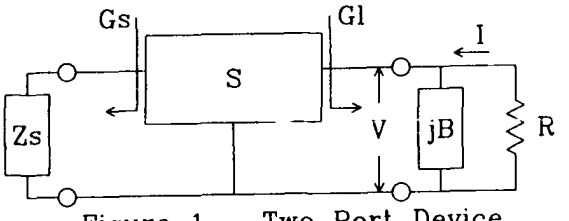


Figure 1 – Two Port Device

where V is the device DC bias voltage, I is the DC bias current, Vsat is the saturation voltage of the device, Imin is the cutoff current of the device, and R is the value of the shunt load resistance on the device when the shunt susceptance of the device has been canceled out. It should be pointed out in passing that the amplifier gain is not maximized since the optimum load admittance is not the admittance for conjugate matching. We desire maximum power conversion, not maximum power gain.

If the input impedance of the device is negative real for the desired load (R and shunt susceptance), then an oscillator can be made by putting the input termination on the device given by:

$$\Gamma s = (1. - S22 * \Gamma l)/(S11 - \Delta s * \Gamma l)$$

where Δs is the determinant of the S matrix, the S matrix is characterized or measured at the required power output P, and Γs and Γl are the source and load reflection coefficients respectively. However, the source impedance resulting from the reflection coefficient Γs will not in general be purely imaginary. This results in power being delivered to the input. What is really desired is that all the power be delivered to the load.

If two circuits are cascaded with the load R as shown in Figure 2, and external shunt feedback is put around the device as shown, the Barkhausen criteria can easily be calculated to be (1):

$$\Delta s + S12 + S21 = 1$$

However, circuits CKT1 and CKT2 are not in general easy to determine and have unique values only under some set of constraints e.g. simultaneous conjugate match.

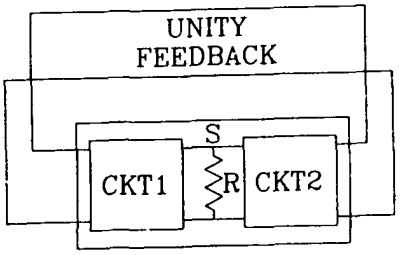


Figure 2 – Two Port Matched Devices. Load Included in S

Consider now converting the two port S parameters to three port S parameters with the conversion:

CH2965-2/91/0000-364 $1.00 © 1991 IEEE

$$M1 = \begin{pmatrix} 2 & 0 & (S11 + S12 - 1) \\ 0 & 2 & (S21 + S22 - 1) \\ 1 & 1 & 1 \end{pmatrix}^{-1}$$

$$M2 = \begin{pmatrix} 2*S11 & 2*S12 & (1 - S11 - S12) \\ 2*S21 & 2*S22 & (1 - S21 - S22) \\ 1 & 1 & 1 \end{pmatrix}$$

```
S3 = M1*M2
```

where S3 is the three port indefinite matrix derived from the two port scattering matrix S. This transformation assumes that several feedback and shunt elements on the common terminal are negligible. In general they need to be considered and can be included if the full three port S parameters are measured or calculated.

Whether the above transformation is used whenever the given assumption is valid, or if the full three port S parameters from measurements or characterizations are used, the following equations will determine the values for Γ1, Γ2, and Γ3 using a specified load or source impedance for the device.

Consider the three port circuit shown in Figure 3. Let V23 be the voltage from the output port terminal to the common terminal of the device. Then to maintain a given load on the embedded device, the impedance:

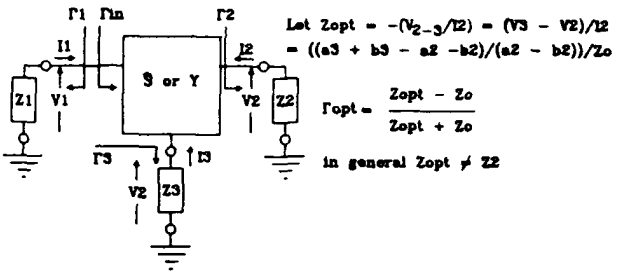
$$Z1 = -V23/I2$$

must be maintained for any terminating impedances on the three terminals.

Let Γ1 and Γ3 be the reflection coefficients of the impedances put on terminal one and terminal three respectively. For these impedances to be reactive, the magnitudes of Γ1 and Γ3 must be equal to 1. Applying these constraints, a closed form solution for Γ3, Γ1, and a new load reflection coefficient Γ2 can be found. The new load Γ2 maintains the relationship on the device such that the terminal voltage across the device, V23, and the device terminal current, I2, are related by the optimum load Zopt which is in general different from the load specified by Γ2. Since no power flows into Γ1 and Γ3 if they are of magnitude of one, all the power will appear in Γ2 and thus the maximum power available from the device will be delivered to Γ2 as an oscillator if Zopt is the load for maximum power delivered from the device when the device is used as an amplifier. By interchanging the terminals 1 and 2 of the device, power can be taken of the "input" of the device in a manner which will present the optimum noise source impedance to the device. Therefore the oscillator can be designed for either maximum power generation or minimum device noise.

Solving for the magnitude of Γ1 = 1 with the constraint that the magnitude of Γ3 = 1 gives a closed form solution for Γ3. Γ1 is determined from the above equations, and Γ2 is determined from the general three port equations. These solutions are in terms of familiar bilinear transformations.

OPTIMUM LOAD FOR THE OSCILLATOR



Given a Three Port Device:

$$\Gamma in = \frac{a\Gamma opt\Gamma 3 + b\Gamma opt + c\Gamma 3 + d}{e\Gamma opt\Gamma 3 + f\Gamma opt + g\Gamma 3 + h}$$

For a given Γopt,

$$\Gamma in = \frac{A\Gamma 3 + B}{C\Gamma 3 + D}$$

Figure 3: Three Port Oscillator Configuration

It can be shown that Γin is equal to NUM(Γin)/DEN(Γin) where:

```
NUM(Γin) = Γopt(2 S12 S21 S31 -   S12 S31 S31
               + S11 S32 S31 - 2 S11 S22 S31)
  + Γopt Γ3(2 S12 S31 S23 S31 - 2 S12 S21 S33 S31
          - 2 S13 S31 S31 S22 + 2 S11 S22 S33 S31
              - S12 S31 S31 - 2 S11 S23 S31 S32
              + S11 S32 S31 + 2 S13 S31 S21 S32)
          + Γ3(  S12 S31 S31 -   S11 S32 S31
               + 2 S13 S31 S31 - 2 S11 S33 S31)
  + (2 S11 S31 + S12 S31 S31 -   S11 S32 S31)

        = AΓ3 + B ;  for a given Γopt.
```

```
DEN(Γin) = Γopt(S31 S32 - 2 S22 S31)
+ Γopt Γ3(2 S22 S31 S33 - 2 S23 S32 S31 + S31 S32)
          - Γ3(S32 S31 + 2 S31 S33)
             + (2 S31 -   S32 S31)

        = CΓ3 + D ;  for a given Γopt.
```

$|\Gamma in| = 1$ AND $|\Gamma 3| = 1$ imply no power dissipation at ports 1 or 3.

Let $\Gamma in = \Gamma o + \Gamma p\, e^{-j\Phi i}$ where $\Gamma o = 0$.
and $\Gamma 3 = \Gamma' + \Gamma 3r\, e^{-j\Phi}$ where $\Gamma' = 0$.

$$\text{Then, } |\Gamma in|^2 = \left(\frac{A\Gamma 3 + B}{C\Gamma 3 + D}\right)\left(\frac{A\Gamma 3 + B}{C\Gamma 3 + D}\right)^*$$

$$|\Gamma_p|^2 = \left(\frac{A\Gamma 3r\, e^{-j\Phi} + B}{C\Gamma 3r\, e^{-j\Phi} + D}\right)\left(\frac{A\Gamma 3r\, e^{-j\Phi} + B}{C\Gamma 3r\, e^{-j\Phi} + D}\right)^*$$

$$|\Gamma_p|^2(CC^*\Gamma 3r^2 + C^*D\Gamma 3r\, e^{-j\Phi} + CD^*\Gamma 3r\, e^{+j\Phi} + DD^*) = AA^*\Gamma 3r^2 + A^*B\Gamma 3r\, e^{-j\Phi} + AB^*\Gamma 3r\, e^{+j\Phi} + BB^*$$

If Γ3r is specified, then the equation can be solved for Φ and vice versa. Rearranging:

$$(A^*B\Gamma 3r\, e^{-j\Phi} - |\Gamma_p|^2 C^*D\Gamma 3r\, e^{-j\Phi} + AB^*\Gamma 3r\, e^{+j\Phi} - |\Gamma_p|^2 CD^*\Gamma 3r\, e^{+j\Phi}) = |\Gamma_p|^2(CC^*\Gamma 3r^2 + DD^*) - AA^*\Gamma 3r^2 - BB^*$$

Let:
$w\, e^{j\Phi} = AB^*\Gamma 3r\, e^{+j\Phi} - \Gamma_p{}^2 CD^*\Gamma 3r\, e^{+j\Phi}$
and:
$K = |\Gamma_p|^2(CC^*\Gamma 3r^2 + DD^*) - AA^*\Gamma 3r^2 - BB^*$

Solving the equation for Φ:

$w\, e^{j\Phi} + w^*\, e^{-j\Phi} = K$
$(u + j\,v)(\cos\Phi + j\,\sin\Phi) + (u - j\,v)(\cos\Phi - j\,\sin\Phi) = K$
$2u\cos\Phi - 2v\sin\Phi = K$
$u\cos\Phi - v\sin\Phi = K/2$
$\sqrt{u^2 + v^2}\cos(\Phi + \theta) = K/2$
$u = \sqrt{u^2 + v^2}\cos(\theta)$
$v = \sqrt{u^2 + v^2}\sin(\theta)$

$\theta = \arctan(v/u)$
$\phi = -\theta + \arccos(K/(2\sqrt{u^2 + v^2}))$
ϕ has two possible solutions.
If $\Gamma 3r = 1$, then one solution has a $\Gamma p > 1$ and one solution has $\Gamma p < 1$.
If $K/(2\sqrt{u^2 + v^2}) > 1$, then no oscillator exists for the specified conditions.

The solution can be redone for:

$$\Gamma in = \Gamma o + \Gamma p\, e^{-j\phi i}$$
$$\Gamma 3 = \Gamma' + \Gamma 3r\, e^{-j\phi}$$

in general. Two solutions will still exist but both solutions could have negative real impedances for terminations. If the termination conditions for the input and the common terminal are the conditions for reflection coefficients representing Q's - loaded, unloaded, or external, then two sets of circles exist and four solutions are possible. All four, several, or none or the solutions may be realizable for any particular set of Q's. The solutions for this set of conditions has been given (6).

Consider an example of a MRF901 transistor at 1 GHz operating at 10 V for Vce and 20 mA Ic. The S matrix of the device given by the manufacture is:

S11 = .51 at 162 degrees S12 = .08 at 69 degrees
S21 = 3.51 at 71 degrees S22 = .39 at -41 degrees

The output capacity of this part is nominally 0.4 pF. This part should require a shunt load of:

$$R = (10 - 2)/.02 = 400 \text{ ohms}$$

in parallel with an inductive reactance of j398 ohms. A small amount of negative resistance is needed at the input to overcome the loss in elements used to synthesize $\Gamma 1$ and $\Gamma 3$. This is included in the equations by solving for a $\Gamma 1$ which is slightly smaller than one in magnitude (6). The solution for the magnitude of $\Gamma 1$ not equal to one is included in the solution given above.

Substituting the above values into the equations for $\Gamma 1 = 0.9$ gives:

$Z1 = 0.8424 - j6.4983$
$Z3 = 0. - j14.8578$
$Z2 = 189.6 + j200.$

Notice that the input load has a small series resistance in it to account for component losses. An oscillator was built using these values and generated 50 mW of RF power at 1 GHz. The procedure has been used for many such oscillators which have been built in the RF design classes at Iowa State University with similar results.

An example run for a 500 MHz oscillator using a MRF901 device follows. The output listing is from a computer program that solves the above equations. A circuit analysis listing is also included for EEsof's Touchstone program given although this particular data set was not calculated using Touchstone.

```
ENTER ZO
50
ENTER     2    FOR 2 PORT DATA
ENTER     3    FOR 3 PORT DATA
2
ENTER     S    FOR S PARAMETERS
ENTER     Z    FOR Z PARAMETERS
ENTER     Y    FOR Y PARAMETERS
S
ENTER THE S PARAMETERS IN  VOLT-MAGNITUDE
                           PHASE-DEGREES
ENTER S11,S12,S21,S22
0.5 -166 .05 57 6.81 92 .41 -35
ENTER THE OPTIMUM IMPEDANCE
288.409 265.78

ENTER DATA FOR GAMMA SOURCE
ENTER THE TYPE OF INPUT YOU WISH

ENTER  QE  FOR EXTERNAL Q
ENTER  QU  FOR UNLOADED Q
ENTER  QL  FOR LOADED Q
ENTER  MA  FOR MAGNITUDE OF GAMMA
ENTER  CR  FOR CENTER AND RADIUS OF CIRCLE
MA

ENTER RADIUS OF REFLECTION COEFFICIENT
0.251

ENTER THE DATA FOR GAMMA COMMON TERMINAL
ENTER THE TYPE OF INPUT YOU WISH

ENTER  QE  FOR EXTERNAL Q
ENTER  QU  FOR UNLOADED Q
ENTER  QL  FOR LOADED Q
ENTER  MA  FOR MAGNITUDE OF GAMMA
ENTER  CR  FOR CENTER AND RADIUS OF CIRCLE
MA

ENTER RADIUS OF REFLECTION COEFFICIENT
0.95

SOLUTION ONE

REFLECTION COEFFICIENTS IN MAG AND ANGLE-DEG

  INPUT REFLECTION COEF     3.984062  -164.3204
2PORT LOAD REFLECT COEF     .8297441   9.961959
GROUND INVERSE REF COEF     1.052632   101.7578
 SOURCE REFLECTION COEF     .2510001   164.3204
3PORT LOAD REFLECT COEF     .8386903   10.87746
 GROUND REFLECTION COEF     .9500000  -101.7578

IMPEDANCES IN REAL AND IMAGINARY FORM

              INPUT IMP  -30.29769  -4.386847
        TWO PORT LOAD IMP  288.4090   265.7800
GAMMA INVERSE GROUND IMP  -2.129126   40.62005
             SOURCE IMP   30.29769   4.386847
     OSCILLATOR LOAD IMP   264.0737   281.8258
             GROUND IMP   2.129124  -40.62005

SOLUTION TWO

REFLECTION COEFFICIENTS IN MAG AND ANGLE-DEG

  INPUT REFLECTION COEF     3.984064  -21.33271
2PORT LOAD REFLECT COEF     .8297441   9.961959
GROUND INVERSE REF COEF     1.052632   54.46371
 SOURCE REFLECTION COEF     .2510000   21.33271
3PORT LOAD REFLECT COEF     .8545094   12.02498
 GROUND REFLECTION COEF     .9500000  -54.46370

IMPS IN REAL AND IMAGINARY FORM

              INPUT IMP  -78.68701  -15.33593
        TWO PORT LOAD IMP  288.4090   265.7800
GAMMA INVERSE GROUND IMP  -6.107611   96.85227
             SOURCE IMP   78.68701   15.33593
     OSCILLATOR LOAD IMP   229.9467   303.4441
             GROUND IMP   6.107609  -96.85227
```

```
ENTER   G   FOR A DIFFERENT VALUE OF GAMMA
ENTER   Z   FOR A DIFFERENT VALUE OF LOAD
ENTER   N   FOR STOPPING
Stop - Program terminated.
```

Solution one was chosen since this is a common emitter device and the resistance of Z3 is minimized to maintain maximum output power.

A Touchstone file to calculate the output impedance follows:

```
! Osc circuit design for a MRF901 at 500 MHz
!
DIM
  FREQ GHZ
  RES OH
  IND NH
  CAP PF
  LNG MIL
  TIME PS
  COND /OH
  ANG DEG

VAR

EQN

CKT

! oscillator S-parameter definitions
   S2PA_Q1 50 60 8  MRF901H.s2p
! from Motorola data disk at 10V 15mA
!
! put in .1 inch
! of microstrip line on Er=2.45 substrate
! on collector and base
     tlin 50 51 z=50 f=.5 e=2.191
!    tlin 60 61 z=50 f=.5 e=2.191
! comment above tl out to see output
! impedance at the device
     tlin 60 61 z=50 f=.5 e=0.000001
!  RESONATOR ON BASE
   RES 1 0 R=20
   CAP 1 2 C=.15915494
   IND 2 3 L=636.6198
   CAP 3 0 C=0.75
   RES 3 4 R=10
!
!  MATCHING SECTION - CRYSTAL TO DEVICE
!  PHASE SHIFTER
     tlin 4 51 z=50 f=.5 e=7.32948
!
!  COMMON TERMINAL IMPEDANCE
   RES 8 9 R=2.129124
   CAP 9 0 C=7.836275
!
!  OUTPUT MATCH
   IND 61 7 L=89.707938
!  89.707938nH is j281.8258 ohms at 500 MHz
!  RES 7 0 R=264.0737
!  ABOVE COMMENTED OUT TO LOOK
!  AT OUTPUT IMPEDANCE

  def1p 7 out

FREQ
  SWEEP 0.498 .502 .0004

OUT

    out re[z1] gr1
    out im[z1] gr1
```

The output impedance of the oscillator follows. Notice that the output resistance is -264 ohms and the output reactance is approximately zero. The external Q is approximately 78.5.

OUTPUT IMPEDANCE OF 500 MHz OSCILLATOR

FREQ-GHz	RE(ZOUT)	IM(ZOUT)
.498000	-.159683E+03	-.386601E+02
.498400	-.178036E+03	-.362511E+02
.498800	-.197883E+03	-.317480E+02
.499200	-.219082E+03	-.246056E+02
.499600	-.241316E+03	-.142219E+02
.500000	-.264065E+03	.265083E-01
.500400	-.286536E+03	.187036E+02
.500800	-.307654E+03	.422068E+02
.501200	-.326080E+03	.706074E+02
.501600	-.340337E+03	.103521E+03
.502000	-.349001E+03	.139988E+03

Conclusions

An approach to microwave oscillator design has been given which will solve in closed form the impedances necessary to be presented to the device for some optimum condition, e.g. maximum power or minimum noise sheif. This approach is applicable whenever S parameters are available including millimeter wavelengths and even at optical frequencies with the appropriate S matrix. The discussed example used series terminations. With the appropriate delta to wye transformations, the analysis is applicable to shunt oscillator designs as well (6).

Acknowledgments

I would like to acknowledge the efforts of Mr. Andrew Oliver, a student in electrical engineering at Iowa State University, for measuring several MRF901 devices under large signal conditions using the linearized S Parameter technique

References

1. R. J. Weber, Referenced in a MS thesis by Philip Howard Thompson, "Design and Characterization of Voltage Controlled Oscillators Using Thin Film Resonator Based Filters," Iowa State University, Ames, Iowa, 1989.

2. George D. Vendelin, Anthony M. Pavio, and Ulrich L. Rohde, Microwave Circuit Design Using Linear and Nonlinear Techniques, New York, Wiley-Interscience, 1990.

3. David M. Pozar, Microwave Engineering, Massachusetts, Addison-Wesley, 1990

4. Rowan J. Gilmore and Fred J. Rosenbaum, "An Analytic Approach to Optimum Oscillator Design Using S-Parameters," IEEE Transactions on Microwave Theory and Techniques, VOL. MTT-31, NO. 6, pp 633, Aug 1983.

5. R. J. Weber, "Oscillator Design Techniques Using Calculated and Measured S-Parameters," Tutorial presented at the 45th Annual Symposium on Frequency Control, May 1991.

6. R. J. Weber, "A Microwave Oscillator S-Parameter Design Technique and Computer Algorithm," 34th Midwest Symposium on Circuits and Systems, May 1991.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

UNIVERSAL, COMPUTER FACILITATED, STEADY STATE OSCILLATOR, CLOSED LOOP ANALYSIS THEORY AND SOME UHF AND MICROWAVE APPLICATIONS

Benjamin Parzen
consulting engineer
3634 Seventh Avenue
San Diego, CA 92103

ABSTRACT

This paper is a continuation of Ref 4. It includes most of the theory of that reference plus additional theory to facilitate analysis of UHF and microwave oscillators.
The theory of oscillator analysis in the immittance domain is presented. This theory enables the computer simulation of the steady state oscillator. The simulation makes practical the calculation of the oscillator total steady state performance, including noise at all oscillator locations. Some oscillator applications of PC program, BPT, created for the simulation, are listed and some specific microwave oscillators analyzed.

PART 1 THEORY

1. INTRODUCTION

1.1 During the past 16 years, the writer has been developing theory for oscillator analysis in the immittance domain. "IMMITTANCE DOMAIN" means that the basic relationships are expressed in terms of immittance,

$Z = R + j X$ or $Y = G + j B$

Effective and facile application of the theory to real problems also necessitated the creation of a PC computer circuit analysis program.

The results of this effort are
A remarkably simple oscillator theory which fully describes the oscillator operation and which is readily translatable into the real world.
A very user friendly multipurpose and powerful circuit analysis program, BPT.

The theory and program are universal in that they apply to all oscillators, past, present, and future.

The theory and program are symbiotic, in that
To apply the program to oscillators, the theory must be used.

The program is very helpful and almost indispensable in applying the theory to real oscillator problems including that of improving the theory.

It should be noted that, as BPT does not incorporate any basic information peculiar to oscillators, it follows that the theory may be used, in conjuction with other circuit analysis programs, to analyze oscillators, but with much greater, and perhaps prohibitive, difficulty.

1.2 The theory is basesd upon the combination of 2, compatible and equally important, oscillator models.
The noise source, amplifier, filter, model [3] in which, positive feedback produces the necessary extremely high effective amplifier gain, as the primary model.

The negative resistance model [2] as the secondary model.

2.THE NOISELESS OSCILLATOR
(See Ref 2)

2.1 Fundamental Relationships

At steady state, in any and every mesh of a hypothetically noiseless oscillator,

$$Zt = \sum Z = 0 \qquad (1)$$

From which,

$$\sum R = 0 \qquad (1a)$$

$$\sum X = 0 \qquad (1b)$$

Defining,

$$RN = \sum Rnegative \qquad (2a)$$

$$RT = \sum Rpositive \qquad (2b)$$

Then,

$$RT = -RN \qquad (3)$$

The dual statement of Eq 1 is,
At steady state, across any and every 2 nodes

CH2965-2/91/0000-368 $1.00 © 1991 IEEE

of a hypothetically noiseless oscillator,

$$Yt = \sum Y = 0 \qquad (4)$$

It is noteworthy that Eqs 1 and 4 are completely analogous to Kirchoff's first and second network laws. It may, therefore, be in order to name Eqs 1 and 4 as the oscillator first and second laws. It is interesting to speculate that similar sets of laws may exist in other fields and that efforts should be made to discover and apply them.

For convenience, Eq 1 is called the Negative Resistance Oscillator Model, NROM. Similarly Eq 4 is called the Negative Conductance Oscillator Model, NCOM.

Theoretically, both models are completely equal. Practically, for a particular application, one model may have advantages and disadvantages over the other. The choice of use depends upon the application. In many applications, it is desirable to use both models. Examples are described in Ref 7.

2.2 Non-linearity considerations

If the linear and non-linear elements are grouped together, then, from Eqs 1

$$\sum Rlin + \sum Rnlin = 0 \qquad (5)$$

from which,

$$\sum Rlin = - \sum Rnlin \qquad (5a)$$

Similarly,

$$\sum Xlin = - \sum Xnlin \qquad (5b)$$

where the non-linear values are the effective values.

Eqs 5 are very important since they state that it is only necessary to study the linear R and X elements to compute the totals of the non-linear R and X elements. This fact drastically reduces the labor involved in the analysis and, very often, it eliminates the need for considering the detailed behavior of non-linear elements.

2.3 Noiseless oscillator model

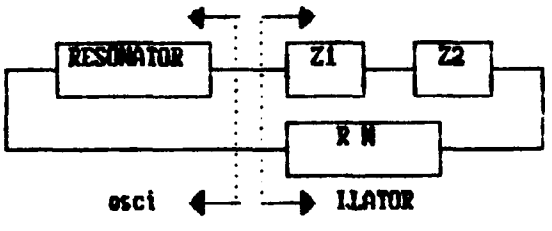


Fig. 1 Noiseless Oscillator Model

Fig 1 is the simplified diagram of the oscillator resonator mesh. (The mesh containing the resonator is usually chosen for analysis as it has the most information.)
As shown in Fig 1, that part of the oscillator containing the resonator circuitry is called the osci, symbol os. The remainder is called the llator, symbol LL.

From Eq 1,

$$Z_{LL} = - Z_{os} \qquad (6)$$

Also,

$$Delta\ F = - Delta\ X_{LL} / (\delta X_{os}/\delta f) \qquad (7)$$

Eq 7 enables the easy determination of oscillator frequency shifts caused by llator changes.

3. THE REAL OSCILLATOR

3.1 Introduction

The real oscillator always includes one or more llator noise sources, in addition to the resonator frequency noise. The calculation of the contribution of the resonator noise to the oscillator noise is very simple (See Section 3.3). However, the calculation of the contributions of the llator noises is extremely difficult and constitutes a major topic of this paper.

3.2 Frequency relationships

The total frequency, F, is described by

$$F = fo + f \qquad (8)$$

where fo is the carrier frequency
f is the offset or Fourier frequency

3.3 Contribution of the resonator frequency noise to the oscillator noise.

Let the resonator frequency noise be described as

$$| S_f (f) |_{os} \qquad (9)$$

Then the resonator frequency noise contribution to the oscillator frequency noise at all oscillator locations is identical to that in Eq 9.
The resonator frequency noise contribution to the oscillator phase noise at all oscillator locations is

$$| S_f (f) |_{os} / f^2 \qquad (10)$$

The total phase noise at any oscillator location, m, is given by the sum of that of Eq 10 and the contributions of the llator noise sources to the phase noise at that location, $PStm$.

3.4 Contributions of the llator noise sources to the oscillator noise at location m.

Let the noise of noise source, n, be given by power spectrum,

$$PSn\ (f) \qquad (11)$$

then the contribution of this source to the oscillator noise power spectrum at location m, is

$$PSmn\ (f) \qquad (12)$$

and

$$PSmn\ (f) = PSn\ (f) * CFmn\ (f) \qquad (12a)$$

where $CFmn\ (f)$ is defined as the contribution function of noise source, n, to the oscillator noise at location m.

The total oscillator noise, PSt , at location m, obviously is

$$PStm\ (f) = \sum_{1}^{n} PSmn\ (f) \qquad (13)$$

3.5 Real oscillator, special case, ZN configuration

Fig 2 is the complete diagram for the ZN configuration (N meaning noise, Z meaning set up by means of the Z configuration) of a special case real oscillator. The fact that makes this oscillator a special case is that the noise source, Vn , represents the total equivalent noise contributions of all the noise sources in the llator.

All real oscillators always contain a non-physical very small value resistance, dR. The value of dR is determined by the oscillator limiting, or ALC, system which sets the level of the oscillator output.

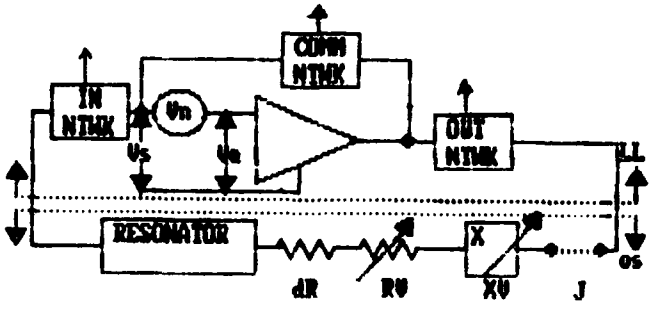


Fig 2. Real Oscillator, Special Case, ZN Config.

Fig 2 includes the input, output, and common networks and the multiple potential output points, available in every real oscillator. The noise in the output depends upon the output point location.
Fig 2 also shows RV and XV which do not exist in the real oscillator. Their function is described in Sect 3.7.

The oscillator, at steady state, at $f = 0$, (the carrier frequency) satisfies Eqs 14 to 16, similar to and derived from Eqs 1 to 3,

$$Zt = \sum Z = dR \qquad (14)$$

From which,

$$\sum R = dR\ , \quad \sum X = 0 \qquad (14a)$$

and

$$RT = -RN + dR \qquad (15)$$

The equivalent relations for the NCOM are

$$Yt = \sum Y = dG \qquad (16)$$

From which,

$$\sum G = dG\ , \quad \sum B = 0 \qquad (16a)$$

Also, it has been proven that

$$|RN / dR|^2 = |Vs\ (0) / Vn\ (0)|^2 \qquad (17)$$

This equation is used in calculating the oscillator operating Q.

Let

$$Ar^2 = (RN / dR)^2 \qquad (18)$$

3.6 Real oscillator operation

a. The oscillator is constructed and power is applied.
b. The operator sets the output level by adjusting the limiting circuitry, and thus setting dR.
c. The oscillator determines fo, Vn, RT , and all other operating conditions.
d. The oscillator determines the phase noise as a function of f.

3.7 The real oscillator, special case, Z configuration

Fig 3 is the complete diagram for the Z configuration (Z meaning impedance) of the special case real oscillator.
The purpose of this configuration is to enable the precise setup of the necessary and sufficient oscillatory condition of Eq 14. Once this is done, we are certain that this configuration, when converted to the ZN configuration, represents a true oscillator at the desired frequency.

As will be noted, the Z configuration differs from the ZN configuration in the following:

a. The value, Vn , of the noise is 0.

b. The circuit has been broken and a 1 A current source inserted. The value of the voltage, *Vz*, across this source is the value of the impedance *Zt* of Eq 14. When *Zt* has been precisely adjusted to *dR* at the desired frequency, *fo*, then this Z configuration is ready for conversion to a ZN configuration. The adjustment procedure is called "zeroing".

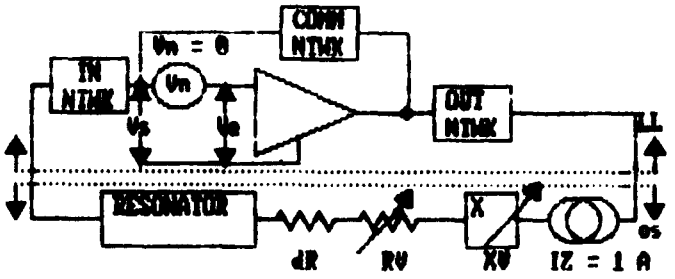


Fig 3. Real Oscillator, Special Case, Z config.

RV and *XV* are very small value trimmable resistor and reactor, respectively, provided to facilitate zeroing. Their values are identical in the Z and ZN configurations.

The zeroing procedure has been automatized in the computer simulated laboratory and thus simplifies the computer oscillator simulation.

3.8 Computer simulated oscillator operation

In the program oscillator, Z config. (See Fig 3)

a. The oscillator is constructed by entering the circuit into the computer.
b. The operator sets *fo* , and *Vn* = 0
c. The program sets the effective gain of the active circuitry, so that Eq 14 is almost satisfied.
d. The program adjusts *RV* and *XV* so that the mesh impedance, *Zt* = *Vz* in Fig.3, satisfies Eq 14 to a very high degree of precision. The Z configuration is then converted into the ZN configuration.

In the program oscillator, ZN config. (see Fig 2)

e. The operator sets *Vn* to *Vn* real.
f. The operator sets *dR* to the value where the desired oscillator output is obtained as measured by the ac current in any component or ac voltage between any 2 points.
g. The program determines the voltages and currents at all other points.
h. The program determines the phase noise at and between all points.
i. The program determines *Zt* (*f*), (see Fig 3) when *dR* is at the value of Step f, and *Vn* is set to 0.

3.9 Computation of the contribution functions, *CF*mn in the real oscillator

This is best done by using the program. The procedure is the following

a. Start with the oscillator configured as in Sect 3.8, Step d. Include all the noise sources, which are assumed to be uncorrelated.
b. Create the equivalent ZN configuration.
c. To compute *CF*mn, set the magnitudes of all noise sources to 0 excepting the source at location n.
d. Make the source, *Vn*, a unit white noise source.
e. Determine the noise response at location m, *PS*mn (*f*)

f. Then from Eq 12a,

$$CF\text{mn}\ (f) = PS\text{mn}\ (f) \qquad (19)$$

g. Repeat steps c to f for all noise sources.

3.10 Oscillator noise in all real oscillators, due to *Vn*
(See Fig 2)

3.10.1 The resonator current, *Ix* , phase noise, due to *Vn* is given by

$$\mathcal{L}_{Ix}(f) = \mathcal{L}_{Vn}(f) * |RT / Zt(f)|^2 \qquad (20)$$

Zt (*f*) is obtained with the Z configuration and

$$\mathcal{L}_{Vn}(f) = PS\text{Vn}\ (f) / (Vs\ (0))^2 \qquad (21)$$

3.10.2 Noise at other locations

The phase noise in *Vs* is almost the same as in *Ix* ,except at very high values of *f*.

Now that the phase noise in *Ix* is known, the determination of the phase noise, at all other locations, is straightforward but, depending upon the circuit complexity, can be very difficult and tedious, and is best done with the program.

However those, desiring to perform any of the calculations, should keep in mind the following rules when combining noises at any location:
a. Noise voltages and currents, due to the same noise source, should be combined as phasors.
b. Noise powers, due to different uncorrelated noise sources, should be combined as scalars.

3.11 The general relationship between $PS(f)$ and the phase noise, at the same location,

$$\mathcal{L}(f) = PS(f)/\text{Carrier Power} \quad (22)$$

$$= PS(f) / PS(0) \quad (23)$$

for all values of f when $PS(0)$ does not approach infinity.

Eq 22 is also useful for the case where $Vn(f)$ has a flicker or other noise component which theoretically approaches infinity as f approaches 0 and thus theoretically also makes $PS(f)$ approach infinity as f approaches 0. In this case, the oscillator noise power spectrum, at f when $Xt(f) >> dR$, is independent of dR, which is set by the oscillator limiting system at the desired carrier. Therefore, at these values of f,

$$\mathcal{L}(f) = PS(f) / PC \quad (24)$$

where PC is the arbitrary desired carrier reference power.

3.12 The circuit noise transformation function $CTRm(f)$, applied in the ZN configuration of Fig 2.

Eq 20, repeated here for convenience,

$$\mathcal{L}_{Ix}(f) = \mathcal{L}_{Vn}(f) * |RT / Zt(f)|^2 \quad (20)$$

(term 0) (term 1) (term 2)

may be considered a special case of general Eq 25

$$\mathcal{L}_{m}(f) = \mathcal{L}_{R}(f) * CTRm(f) \quad (25)$$

(term 0) (term 1) (term 2)

Term 0 is the oscillator phase noise at location m, where, in Eq 20, m = Ix. Term 1 is called the residual phase noise [6] of the active or other device. Term 2 is called the circuit transformation of residual noise function at location m.

The importance of the residual noise lies in the fact that it can be measured independently of the oscillator. Then, one needs only to compute the applicable $CTRm(f)$ and then, using Eq 25, determine the oscillator noise, at any and every location.

Since it is stipulated that the noise is due to Vn, then

$$\mathcal{L}_{R}(f) = PS_{Vn}(f) / [Vin(0)]^2 \quad (26)$$

$$= \mathcal{L}_{Vn}(f) * [Vs(0)/Vin(0)]^2 \quad (27)$$

where Vin is the carrier input voltage at which the residual phase noise is measured and Vs is defined in Fig 2.and used in Eq 21.

3.12.1 The computation of $CTRm(f)$

From Eq 22,

$$\mathcal{L}_{m}(f) = PSm(f) / PSm(0) \quad (28)$$

From Eq 12,

$$PSm(f) = CFmVn(f) * PSVn(f) \quad (29)$$

Let

$$ROm = PSm(0) / PSVin(0) \quad 30)$$

Then

$$PSm(0) = PSVin(0) * ROm \quad (31)$$

Combining Eqs 28 and 29 we obtain,

$$\mathcal{L}_{m}(f) = PSVn(f) * CFmVn(f)/PSm(0) \quad (32)$$

Combining Eqs 21, 26, 31, and 32 we obtain,

$$\mathcal{L}_{m}(f) = \mathcal{L}_{R}(f) * CFmVn(f)/ROm \quad (33)$$

Comparing Eq 33 with Eq 25, we see that

$$CTRm(f) = CFmVn(f) / ROm \quad (34)$$

$CFmVn(f)$ is calculated as in Sect 3.8.

3.12.2 The calculation of ROm with the BPT program. (Note that ROm is independent of Vn and dR)

a. Enter the applicable ZN configuration of the oscillator.
b. Set $F = fo$ and dR = 1E-9 ohm.
c. Make Vn a VW component (white noise voltage source) of convenient magnitude.
d. Execute Option C and note the magnitudes Vm, at locations m and Vin.

Then

$$ROm = (Vm / Vin)^2 \quad (35)$$

or in dB using the DB option,

ROm (dB) = Vm referred to Vin

3.12.3 Notes for Sect 3.12

a. Validity of this section -- The reader is reminded that, for flicker noise, this section is valid only for Fourier frequencies, f, at which $Xt(f) >> Rt(f)$ as stipulated in Sect 3.11
b. Figure of merit -- $CTRm(f)$ is a very useful figure of merit of the transformation of device residual noise into oscillator noise at all locations.
c. Residual noise measurements -- In order

to achieve good agreement between the measured oscillator noise performance and the predicted noise performance based upon the residual noise measurement, it is essential that the measurement source and load impedances, at every applicable Fourier frequency, be identical to the equivalent impedances in the closed loop oscillator, as the contributions of the noise current sources to the effective Vn are strong functions of the impedances seen by these sources. Program BPT can compute these inpedances in the closed loop oscillator.

d. Multiple noise devices -- This section applies only to oscillators with a single or single major noisy device. The program must be used for more complicated circuits.

e. If, for practical reasons, the residual noise cannot be measured, the oscillator phase noise is determined by including all the significant noise sources in the ZN configuration and then computing the total noise at each location, as described in this paper.

f. If it is desired to ascertain the magnitudes of the voltages and currents at all locations at f = 0, then

1. Make *dR* a value such as that of Step 3.12.2.b.
2. Set the Nauqahyde of *Vn* in step 3.12.2.c so that *Vin*(osc) becomes equal to *Vin* (residual measurement input *V*) by means of Option E.
3. Execute Option C and record the data.

3.13 The oscillator *Qopm (f)*

Qop, the oscillator operating *Q*, is generally defined by

$$Qop = (dx / df)_{\varepsilon \rightarrow 0} * fo / 2RT \quad (36)$$

It is seen that *Qop* applies only to low *f*.

It is proposed that Eq 36 be extended to be

$$Qopm\ (f) = (dx / df) * fo / 2RT \quad (37)$$

(which includes *Qop*)

as it will yield more information.

It can be shown that

$$Qopm\ (f) \approx fo / \{2 * SQR\ [CTRm\ (f)] * f\} * Vin/Vs \quad (38)$$

Both *Qopm* and *CTRm* will become more important with the expanded use of multi-section resonators, for which *dx/df* is not constant with f and Leeson's noise model [1][5] does not apply.

3.14 Oscillator circuit configurations

Thus far the Z and ZN configurations have been described. There are additional useful configurations which are considered in detail in Ref 7. Some of these are

3.14.1 N config -- This is the raw complete oscillator circuit It is assigned node numbers and then entered into the computer.

3.14.2 YN config --This is the N config. and the noise sources which has also been provided with *Y* tuning elements *BV* and *GV*, *dG*, and a jumper to enable *Z* measurements.

3.14.3 Y config --This is the YN configuration with the amplitudes of the noise sources set to 0.

3.14.4 ZYN config --This is the ZN configuration which has been provided with *BV*, *GV*, and *dG*.

YN , Y, and N configurations have the important advantage of having fewer nodes. There are also many more possible YN and ZYN configurations since tuning elements can be connected between any 2 nodes. Choosing the optimum node pair is very difficult.

4. PROGRAM APPLICATIONS IN OSCILLATORS

Some general applications to oscillators are cited below to demonstrate the utility and power of program BPT when used in conjunction with the theory. It should be remembered that BPT's high resolution and accuracy and virtual ground node facility are of prime importance as they enable the precise zeroing of the configuration required to simulate the oscillator.

4.1 Automatic zeroing of the Z oscillator configuration (coarse, active device gain setting and fine, impedance trimming) for the derivation of the Zt (f) relationship.

4.2 Automatic calculation of the oscillator operating frequency.

Both 4.1 and 4.2 functions are high speed and high resolution executions of their respective functions.

4.3 Calculation of the DC operating point of oscillators.

4,4 Includes a procedure for determining the AC operating point of a self-limiting oscillator.

4.5 Includes a procedure for setting the AC and DC operating points in ALC type oscillators.

4.6 The program is a linear one but is designed to be capable of being interfaced with non-linear operating conditions. Items 4.3 to 4.5 are examples.

4.7 Investigation of the Zt (f) or Zt (?) function of an oscillator or circuit such as a llator.

The following 5 examples illustrate the very useful information obtained from noiseless llator studies.

4.7.1 The effect of component changes upon frequency.
4.7.2 Overtone and mode selector gain margins and the effect of the resonator overtone and mode selector circuitry upon the oscillator stability..
4.7.3 Starting gain margin (loop gain, ALo)
4.7.4 Effect of power dissipating components upon the loop gain and the operating Q.
4.7.5 Effect of component tolerance and environment upon the above 4 items

4.8 Determination of bypass and coupling capacitor adequacy.

4.9 Examination of the performance of subcircuits such as tuning networks.

4.10 The determination of circuit isolation properties.

4.11 Calculation of operating *Q*, with the aid of the *CTRm* and *Qopm* functions.

4.12 Studies of impedance properties of resonators and other devices requiring high resolution.

4.13 Investigation of the effect of component and subcircuit noise upon oscillator noise.

4.14 Determination of the circuit configuration for optimum noise performance.

4.15 Calculating and plotting oscillator phase noise for white noise and white plus flicker noise sources. It is not necessary to assume symmetrical noise sidebands.

4.16 Calculating and plotting the oscillator noise for white noise sources from the *Zt* (*f*) relationship.

4.17 In setting up an oscillator, the oscillator Z configuration is first created and then zeroed. (See Item 4.1). A facility is provided to automatically convert the zeroed Z configuration into an ZN configuration and vice versa.

4.18 Calculation of equivalent Y parameters for S parameters.
4.19 Automatic calculation of the *CTRm* (*f*) function.

4.20 While BPT is basically intended for analysis, CAA, it, in common with other analysis programs, is also very useful in computer aided design, CAD, activities.

4.21 Quasi-instantaneous determination of the oscillator operating frequency, with a resolution better than 1 part 10^{15} ,the loop gain, and the oscillator line width assuming white noise sources.

5. THE PROGRAM AS A RESEARCH TOOL

The following suggested applications illustrate the power and usefulness of the program in research activities:

5.1 The formulation, checking, and confirmation of new theory. and new oscillator designs.

5.2 Determination of component aging from experimentally derived llator aging.

5.3 Determination of component temperature characteristics from experimentally derived llator temperature performance.

5.4 Determination of llator noise by experimentally deriving the oscillator phase noise and then program calculating the residual noise required to produce that oscillator noise.

5.5 Exploring and confirming new oscillator theory and designs prior to attempting experimental verification.

6. CONCLUSIONS FOR THE THEORY

Extremely simple and powerful oscillator theory has been presented. This theory has been used in the creation of a computer program for the universal analysis of oscillator steady state performance.

The importance of *Zt* (*f*), *RT*, and *dR* and the tremendous power of analysis in the immittance domain complemented by computer aided analysis have been demonstrated. An important additional advantage is that this analysis method provides considerably greater understanding of the operation of the real oscillator.

PART 2 SOME TYPICAL UHF AND MICROWAVE APPLICATIONS

7. INTRODUCTION

This part describes some applications at UHF and microwave frequencies. It is important to note that the treatment is exactly the same as that for oscillators at lower frequencies. Indeed, the basic treatment does not vary with frequency. The only difference is the frequent use of transmission line and other distributed elements and the description of components by means of S and Y parameters.

Such customary procedures as open loop analysis, Bode plots, Nyquist diagrams, Smith charts, and k factors are conspicuous by their absence. It should be pointed out that these procedures were invented to avoid the extensive, tedious, labor and time consuming, and, at that time, practically prohibitive computations required by the use of Kirchoff's network laws in closed loop analysis. Of course, these calculations are now markedly facilitated by the ubiquitous and tireless computer directed in accordance with the theory of Part 1. Additional benefits obtained are greater understanding, power to generate information previously unattainable, extraordinary simplicity, and very large savings in time and labor.

The data for these applications are presented in the form of simplified schematics, netlists, a typical parts list, and plots of the more important, and infrequently or not previously published, operating characteristics. Comments on the data are also included.

The data was obtained with program BPT as directed by the user guided by the above theory. The circuit is entered into the computer as a NETLIST via a file or the keyboard. The computer translates the netlist into a PARTS LIST which is readily understood by any user. The user then interactively directs the computer to generate the desired data.

It will be noted from the netlists and parts list that the microwave active devices are characterized by fixed frequency or frequency dependent S or Y parameters. These parameters may be small signal for starting conditions or large signal at steady state conditions. The large signal parameters must be measured at the equivalent input level in the oscillator. (See Sects 2.2 and 3.12).

Such steady state data as power output and the ac voltages and currents at the various locations are not included as they are very easily calculated as described in Sect 3.12.3f.

The applications are:

Llator studies of a 100 MHz Driscoll oscillator and a brief discussion of some types of information available from llator studies.

Analysis of

A low Q, 2 GHz transmission line oscillator.
A high Q, 1.1 GHz transmission line oscillator.
A 4 GHz dielectric resonator oscillator.

The oscillator plots are for 2 quantities versus the Fourier frequency, f.

The circuit transformation of residual noise at location m, CTR_m (f)
The magnitude of the closed loop impedance, Zin (f), at the input terminals of the active device.

The CTR_m function is described in Sect 3.12

If the noise performance of the oscillator, $\mathcal{L}_m(f)$, is known, then as pointed out in Sect 5.4, CTR_m (f) is calculated and the residual noise can then be calculated from Eq 25.

The Zin quantity determines the contribution of the device input noise current, In, to the oscillator noise as it produces a noise voltage, $V_{rn} = In * Zin$, to be added to the device noise voltage, Vn. It is therefore very important, when measuring the device residual noise, that the device be terminated by the impedances present in the closed loop oscillator.
In this connection, the noise currents may be determined by measuring residual noise at various known terminations and then calculating the corresponding noise currents (see Sect 3.12.3c).

8. 100 MHz LLATOR STUDIES

Sect 2.3 defines the osci and llator. THe llator may be treated as a subassembly of the oscillator and is capable of being studied and pretested independently of the oscillator and osci. This is desirable because llator study, measurement, test, manufacture, and trouble shooting are often much simpler than those of the oscillator. (See Ref 2)

Eq 7 relates the llator instability to the oscillator frequency. It is seen, that while the major contributor to the oscillator frequency instability is always the osci, the llator too can deteriorate the performance somewhat. The effect of the llator tuning on the oscillator stability is investigated in this study. This effect is quite important as most oscillators, wherein the emitter or equivalent current is also the osci current, are usually, in practice, not properly tuned for maximum stability since it is much easier to tune for maximum instability.

Fig 4 is the schematic of the llator which is tuned by *Cv*. *IZ* enables the calculation of Z_{LL}.

Table 1 is the corresponding netlist which is entered into the computer.

(see next page for Fig 4 and Table 1)

Z_{LL} vesrus *Cv* is plotted in Fig 5 in terms of R_{LL}(R) and X_{LL}(X). $-R_{LL}$ roughly corresponds to the oscillation amplitude and $-X_{LL}$ corresponds to the llator's contribution to the oscillator frequency. The substance of Eq 7 for the applicable resonator is given in line 3 of that figure.

At *Cv* = 18 pF, $-R_{LL}$ is maximum, 70 ohms. Unfortunately, the X plot shows *dX/dCv* to be maximum or at the point of maximum frequency instability due to instability in *Cv*. The same plot shows the greatest stability (0 slope) at *Cv* = 14 and 22 pF. The 22 pF point is unusable because of low $-R_{LL}$, 15 ohms. Thus the preferred point is at 14 pF where $-R_{LL}$ is 37 ohms. These facts translate into the practical tuning rule: "Start at mimimum *Cv*. Increase *Cv* until the output is maximum. Back off until the output is about 3 dB below maximum.".The plot shows the theoretical difference to be about 5.5 dB. The smaller number is used to allow for limiting effects.

The 100 MHz resonator is a 5th overtune *SC* cut quartz crystal unit. After the llator is properly set for magnitude of *-RLL* and frequency stability, we turn our attention to the study of the possibily of oscillation at one of the other crystal overtone or mode frequencies. These are listed in line 2 of FIg 5. Associated with each frequency, n, is a measured crystal resistance *RIn*. Oscillation at any frequency except 100 MHz would constitute a catastrophic failure. Therefore this should be checked.

A frequency sweep of R_{LL} is made. THe results are shown in Fig. 6.

If $(R_{LL} / RI)_{100} < 2 * (R_{LL} / RI)_n$, for any value of n

that would be cause for concern and a redesign of the llator should be made. 2 is the factor of safety and opinions may vary as to what this factor should be. This writer's opinion is that it should be at least 2. Often it is unwittingly considerably less.

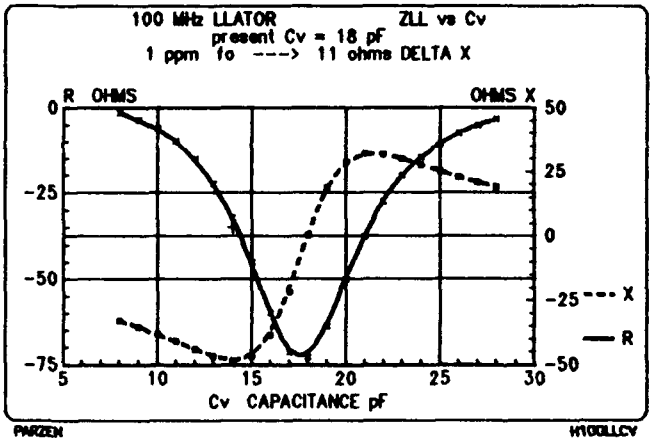


Fig 5 Llator R_{LL} snd X_{LL} vs Cv

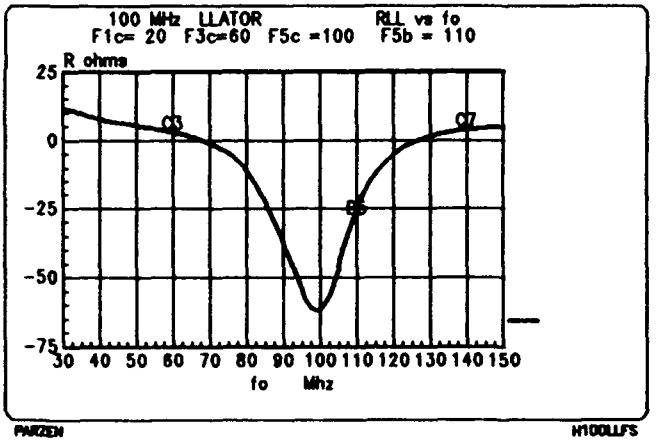


Fig 6 Llator R_{LL} vs fo

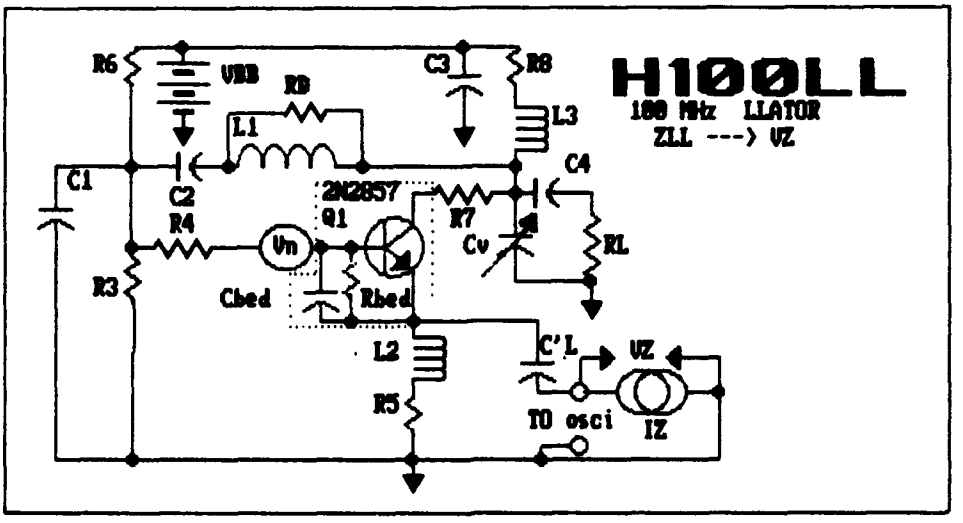


Fig 4 100 MHz Llator
SCHEMATIC

```
100 MHZ NO MS DAMPED LL, 0
C1,C, 0 , 2 , 9E-011 , 0 , 0 , 0
C2,C, 2 , 3 , 1E-009 , 0 , 0 , 0
R3,R, 3 , 0 , 1000 , 0 , 0 , 0
R4,R, 3 , 4 , 9 , 0 , 0 , 0
NN1,N, 5 , 6 , 7 , .2077780189025116 , 100 , 1200
(Rbe),R, 5 , 6 , 481.2828639343197 , 0 , 7 , 0
(Cbed),C, 5 , 6 , 2.7557415645146666E-011 , 0 , 7 , 0
(Vn),V, 4 , 5 , 0 , 0 , 1 , 0
L2,L, 6 , 8 , 8.1999999999999999E-006 , 0 , 0 , 0
R5,R, 8 , 0 , 300 , 0 , 0 , 0
R6,R, 3 , 11 , 3300 , 0 , 0 , 0
VBB,BV, 0 , 11 , 0 , 0 , 0 , 0
R7,R, 7 , 9 , 10 , 0 , 0 , 0
L3,L, 9 , 10 , 2.7E-006 , 0 , 0 , 0
R8,R, 10 , 11 , 10 , 0 , 0 , 0
C3,C, 11 , 0 , 1000 , 0 , 0 , 0
L1,L, 9 , 2 , 1.8E-007 , 0 , 0 , 0
CV1,C, 0 , 9 , 1.8E-011 , 0 , 0 , 0
RD,R, 9 , 2 , 1000 , 0 , 0 , 0
C4,C, 9 , 13 , 1E-009 , 0 , 0 , 0
RL,R, 0 , 13 , 1200 , 0 , 0 , 0
C'L,C, 6 , 12 , 2.7660579263622215E-010 , 0 , 0 , 0
(IZ),I, 0 , 12 , 1 , 0 , 2 , 0
```

Table 1 100 MHz Llator
NETLIST

9. OSCILLATOR ANALYSES

9.1 Introduction

This section starts the oscillator analysis and it begins with stating several observations, cautions, and caveats applicable to all the oscillator analyses herein.

10.1.1. All the oscillator configurations are ZN as that is the easiest to understand, zero, manipulate, and modify. This is in spite of the 3 fewer nodes required by the YN configuration. (See Sect 3.14)

10.1.2. Many writers would classify these oscillators as "negative resistance analysis type" as distinct from open loop analysis type. This writer feels the classification superfluous as negative resistance is a useful concept in all oscillators and any rigorous method of analysis useful for any oscillator should be, and is, useful for all oscillators. In the real world, the method chosen reflects the user's past experience, history, likes and dislikes, prejudices, and training, and the availability of, and capability and willingness to use, tools such as theory, calculators, computers, and software, rather than any requirements posed by the oscillator. The choice should depend only on the completeness of the pertinent information and knowledge gained from the analysis and the ease of use.

9.1.3. The low values of the *CTR* data at the higher values of f, shown in the plots to illustrate some important points, are meaningless for 3 reasons

a. The noise corresponding to the *CTR* is so small that in practice it would be swamped by the following amplifier or other device noise.

b. At the higher frequencies, the values of *CTR* is strongly influenced by the stray and parasitic elements, present in the real oscillator, which have not been entered into the computer.

c. THe S parameter values are not valid at the higher frequencies.

9.1.4. The low values of *Zin* at the higher frequen-cies are probably not too far off because they are in the asymptotic region.

9.1.5. A very important property of any oscillator design is its *Qop*. It is seen from Eq 36 that *Qop* depends upon *fo*, *RT* and *dx/df*. For a fixed *dx/df*, *Qop* increases as *RT* decreases. These parameters are decribed for each oscillator.

9.2 2 GHz lo-Q oscillator

Fig 7 is the schematic diagram and Table 2 is the netlist, both on the next page.

Fig 8 shows *CTR*(RL) and *Zin* plotted versus *f*.

At *f* = 1e4 Hz, *Zin* has the surprisingly high value of 1e6 ohms which can be disasterous for transistors with substantial noise currents. This oscillator has a very low *Qop* of about 1. caused by the high *RT*, 143 ohms, in combination with the low value of the effective *dx/df*, 1.46 e-7 ohms/Hz. Since this value is typical for all oscillators with a transmission line type resonator, *RT* must be made very small to acheive high *Qop*.

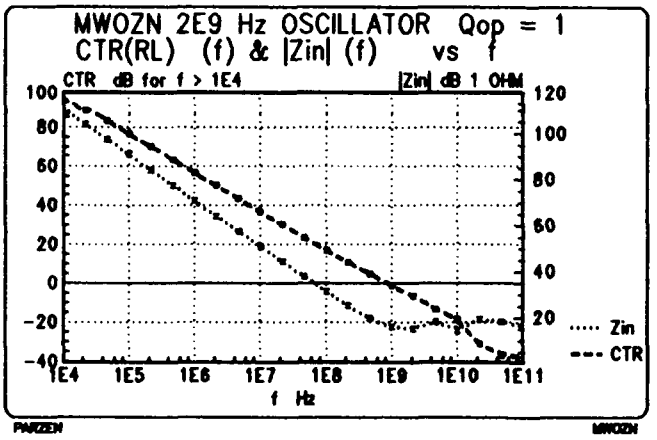


Fig 8 2 GHz Osc.
CTR and Vin vs f

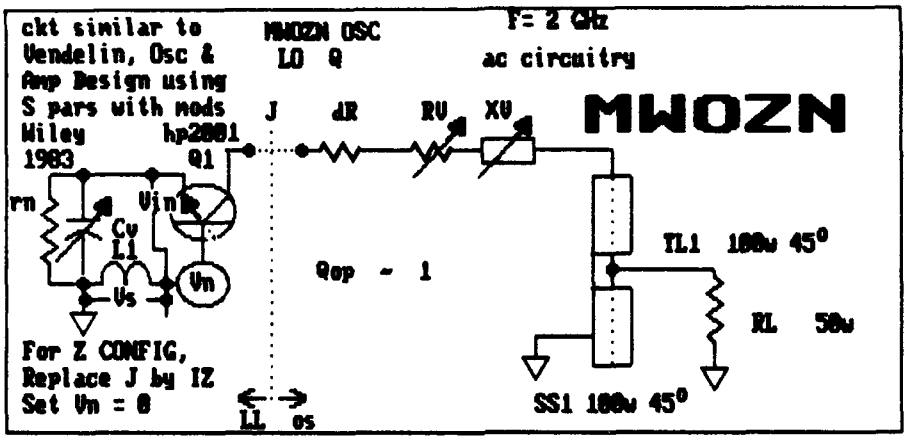


Fig 7 2 GHz OSC SCHEMATIC

```
2 GHZ OSC N CONF VGN = 3, 3
rn,R, 0 , 1 , 6.8884 , 0 , 0 , 0
CV,C, 0 , 1 , 9.1175966207E-012 , 0 , 0 , 0
Vn,VW, 6 , 2 , 3E-009 , 0 , 0 , 0
hp2001 CB Spar,SP, 2 , 1 , 2 , 50 , 3 , 0
(,, 0 , 0 , .9409999999999999 , 174 , 1.9 ,-28
(,, 0 , 0 , 1.01 ,-17 , 1.3E-002 , 98
LI,L, 6 , 0 , 5E-010 , 0 , 0 , 0
{J},J, 3 , 5 , 9.9999999999999E-021 , 0 , 0 , 0
{dR},R, 5 , 7 , 1E-005 , 0 , 0 , 0
{RV},R, 7 , 8 , 1.206733626424983E-003 , 0 , 0 , 0
{XV},X, 8 , 9 , 4.669654306585953E-004 , 0 , 0 , 0
TL1,TL, 0 , 9 , 0 , 100 , 4 , 4000
TL2,TL, 0 , 4 , 0 , 100 , 0 , 4000
RL,R, 4 , 0 , 50 , 0 , 0 , 0
Vs,TP, 0 , 6 , 1E+020 , 0 , 0 , 0
Vin,TP, 1 , 6 , 1E+020 , 0 , 0 , 0
```

Table 2 2 GHz OSC NETLIST

9.3 1.1 GHz hi-Q oscillator

Fig 9 is the schematic diagram and Table 3 is the netlist.

Table 4 is the Parts List for this oscillator and is typical. The parts list is customized for the frequency, current at the time when the list is viewed or hard copied.

Note that the netlist and the parts list include a type SPF component, which is a transistor characterized as a frequency dependent S parameter network. The values of the parameters are those at the current frequency specified in the parts list.

Note that there 2 output points, RL and RLopt. RL is at the cavity output. Rlopt is at the transistor output.
LW1 and LW2 are lead iinductances.

Fig 10 shows *CRTm* and *Zin* plotted versus *f*.
The curves are sumilar to those of the 2 GHz oscillator except for the effects of the higher Q.which shifts the curves nearer the CTR Y axis.

The *CTR*(RL) and *CTR*(RLopt) curves are identical for f < 1e6. Then the cavity filtering action for CTR(RL) evidences itself

dx/df is 1.48e-7 ohms/Hz, by coincidence almost identical to that of the 2000 GHz oscillator. However *RT* (≈ *-RN*) is .237 ohm which accounts for the much higher Qop,

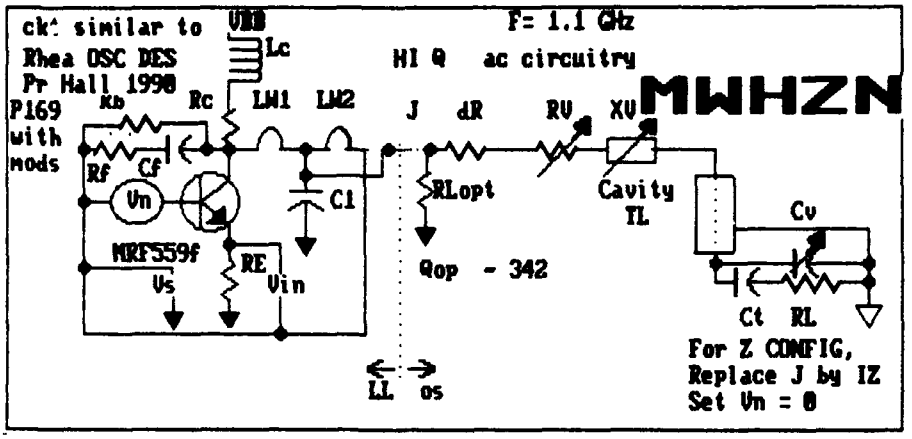


Fig 9 1.1 GHz osc
SCHEMATIC

```
1100 MHz HI Q OSC. CONF = last letter(s) of file name, 8
Vn,VW, 7 , 1 , 3E-009 , 0 , 0 , 0
{mrf559f,SPF, 14 , 1 , 14 , 50 , 2 , 0
(,, 0 , 0 , .754000014172684 , 156.3999963555955 , 1.711999870421175 , 52.1999971654632
(,, 0 , 0 , .3160000263206988 ,-99.40000546660669 , .1048000109332134 , 68.80000182220223
RE,R, 0 , 14 , 1E-003 , 0 , 0 , 0
Rf,R, 7 , 3 , 680 , 0 , 0 , 0
Cf,C, 3 , 2 , 1E-009 , 0 , 0 , 0
Rb,R, 7 , 2 , 18000 , 0 , 0 , 0
Lc,L, 4 , 5 , 1E-007 , 0 , 0 , 0
Rc,R, 5 , 2 , 100 , 0 , 0 , 0
LW1,L, 2 , 6 , 2E-009 , 0 , 0 , 0
LW2,L, 6 , 7 , 2E-009 , 0 , 0 , 0
C1,C, 6 , 0 , 6.5E-011 , 0 , 0 , 0
TL1,TL, 0 , 11 , 0 , 100 , 12 , 1135
Cv,C, 12 , 0 , 4E-015 , 0 , 0 , 0
VCC,BV, 4 , 0 , 15 , 0 , 0 , 0
RLopt,R, 0 , 2 , 1000 , 0 , 0 , 0
Ct,C, 12 , 13 , 1E-013 , 0 , 0 , 0
{J},J, 6 , 8 , 9.999999999999999E-021 , 0 , 0 , 0
{dR},R, 8 , 9 , 1E-008 , 0 , 0 , 0
{RV},R, 9 , 10 , 7.668542056988303E-002 , 0 , 0 , 0
{XV},X, 10 , 11 ,-1.280249757079139E-002 , 0 , 0 , 0
tpct,TP, 0 , 12 , 1E+020 , 0 , 0 , 0
RL,R, 0 , 13 , 50 , 0 , 0 , 0
Vs,TP, 0 , 7 , 1E+020 , 0 , 0 , 0
Vin,TP, 7 , 14 , 1E+020 , 0 , 0 , 0
```

Table 3 1.1 GHz OSC
NETLIST

```
05-18-1991    19:55:44
# OF COMPONENTS =  26   HIGHEST NODE # =  14    # OF VOLTAGE SOURCES =  2
MWHQZN                              FREQUENCY= 1100000101.233457
CIRCUIT NOTE :  1100 MHz HI Q OSC. CONF = last letter(s) of file name
  COMPNT. CONNECTED TO NODE
 #   SYB   N1-   +N2      TYPE          VALUE          PHASE ANGLE
 1   Vn      7 - 1       VW, WH NS V   1E-009
 2   (mrf559f
                IN LO= 14 ; HI= 1 -- OUT LO= 14 ; HI= 2 ; Zo= 50
  F DEPENDENT  SPP, S PARAMETER NTWK INCLUDING THE NEXT TWO CPTS
( MRF559F 10V 10MA 4F 250-1500 MHZ
 3   MAG s11 =0.7540  ANG s11 =+156.4  MAG s21 =   +1.712  ANG s21 = +52.2
 4   MAG s12 =0.1048  ANG s12 = +68.8  MAG s22 =0.3160  ANG y22 = -99.4
 5   RE      0 - 14      RESISTOR      1E-003
 6   Rf      7 - 3       RESISTOR      680
 7   Cf      3 - 2       CAPACITOR     1E-009
 8   Rb      7 - 2       RESISTOR      18000
 9   Lc      4 - 5       INDUCTOR      1E-007
 10  Rc      5 - 2       RESISTOR      100
 11  LW1     2 - 6       INDUCTOR      2E-009
 12  LW2     6 - 7       INDUCTOR      2E-009
 13  C1      6 - 0       CAPACITOR     6.5E-011
 14  TL1    IN LO= 0 ; HI= 11 : OUT LO= 0 ; HI= 12  TL, TRANSMISSION LINE
                          Zo =  100   ,LAMBDA/4 FREQ (MHz)=  1135
 15  Cv      12 - 0      CAPACITOR     4E-015
 16  VCC     4 - 0       BV, DC VLTG   15
 17  RLopt   0 - 2       RESISTOR      1000
 18  Ct      12 - 13     CAPACITOR     1E-013
 19  {J}     6 - 8       JUMPER,     R =  9.9999999999999999E-021
 20  {dR}    8 - 9       RESISTOR      1E-008
 21  {RV}    9 - 10      RESISTOR      7.668542056988303E-002
 22  {XV}    10 - 11     X, REACTANCE      -1.280249757079139E-002
 23  tpct    0 - 12      TESTPOINT SET,    R =  1E+020
 24  RL      0 - 13      RESISTOR      50
 25  Vs      0 - 7       TESTPOINT SET,    R =  1E+020
 26  Vin     7 - 14      TESTPOINT SET,    R =  1E+020
```

Table 4 1.1 GHz OSC PARTS LIST

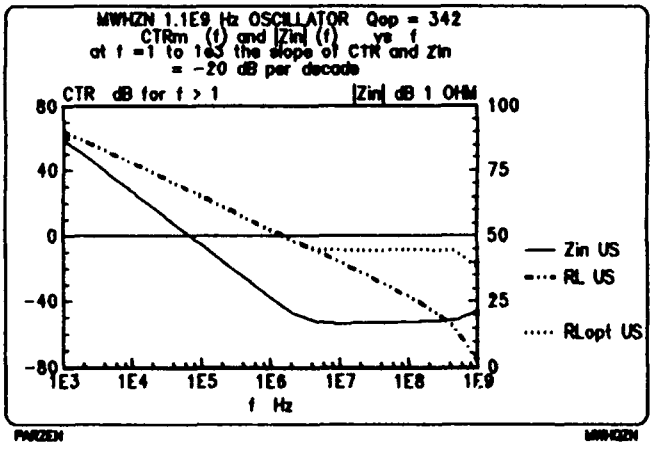


fig 10 1.1 GHz OSC
CTR and Zin vs f

9.3 4 GHz dielectric resonator oscillator

Fig 11 is the schematic diagram and Table 5 is the netlist, both on the next page.

Ls and Lx are lead inductance.

Fig 12 shows *CRTm* and *Zin* plotted versus f.

These curves include data for both the upper and lower sidebands, US and LS, of the spectrum since they may not be symmetrical. The asymmetry is caused by the fact that the signal, at the location being observed, is the sum of at least 2 signals arriving via different paths. If there is only 1 major llator noise source then the signals are correlated and must be combined as phasors.

The relative phase varies with the frequency *f*, and at *f* = *fa* the signals will be in phase in one sideband and out of phase in the other sideband. The out of phase signals causes dips in the *CTR* function in the region of *fa*. The value of *fa* has a strong dependence upon *Qop*, being closer to *fo* the greater the *Qop* because the phase shifts more rapidly.

The magnitude of the dip is a function of the equality of the magnitudes of the signals. For example, a 10 MHz crystal oscillator, having a Qop of 1E6, may have a noise dip greater than 20 dB in its lower sideband output at *fa* = 20 Hz. THere is no sign of a dip in the resonator current noise since only a single path is then involved.

This effect may be of great importance in systems which require a low noise signal in a relatively narrow *f* region close to the carrier.

CTRm curves are provided for 2 locations, RL and Ldc.(Ld). Ldc is at a location suitable for extracting power. Also at this location the asymmetry is much greater than at RL. Curve E for CLd US (*CLD* ≡ *CTRLd*) shows a 10 dB dip at fa ≈ 2E5 Hz. Curve C, shows a negligable dip in CRL US.

A strong dip also exists in curve A, the curve for *Zin* US. at a somewhat higher *fa*.

RT = 5.5 ohms and *dx/df* = 1.216E-5 ohm/Hz. The *Qop* is therefore 4420.

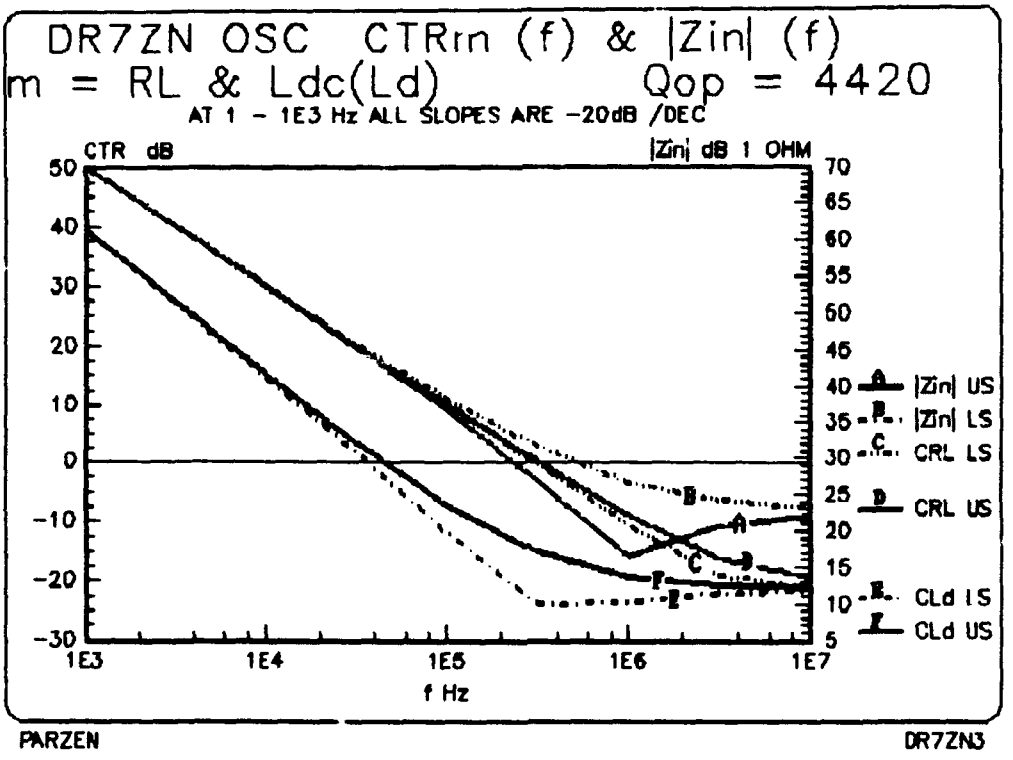


Fig 12 4 GHz DRO CTRm and Zin vs f

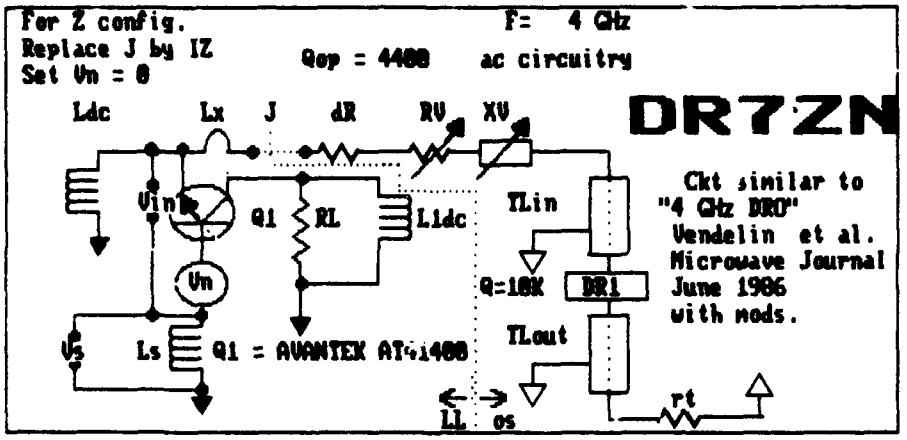


Fig 11 4 GHz DRO SCHEMATIC

```
DIELECTRIC RSONATOR 4 GHZ OSC N CONFIG, 7
{J},J, 7 , 13 , 9.999999999999999E-021 , 0 , 0 , 0
{dR},R, 13 , 1 , 1E-007 , 0 , 0 , 0
{XV},X, 1 , 2 ,-3.917136581082582E-005 , 0 , 0 , 0
{RV},R, 2 , 3 ,-2.781566539213837E-002 , 0 , 0 , 0
TLin,TL, 0 , 3 , 0 , 50 , 4 , 4592
TLout,TL, 0 , 5 , 0 , 50 , 6 , 5000
rt,R, 0 , 6 , 50 , 0 , 0 , 0
Ldc,L, 0 , 7 , 1E-005 , 0 , 0 , 0
Lx,L, 7 , 8 , 2.015205757595715E-010 , 0 , 5.064764482827642 , 4000000000
Q1-CESP,SP, 8 , 9 , 8 , 50 , 10 , 0
(,, 0 , 0 , .8057 , -176.74 , 2.599 , 74.77
(,, 0 , 0 , .4306 , -22.94 , 3.16E-002 , 56.54
Vn,VW, 9 , 11 , 3E 09 , 0 , 0 , 0
Ls,L, 0 , 11 , 3.1E-010 , 0 , 0 , 0
DR1,XD, 4 , 5 , 4.1E-010 , 10000 , 0 , 4000
L2dc,L, 0 , 10 , 1E-005 , 0 , 0 , 0
RL,R, 0 , 10 , 50 , 0 , 0 , 0
Vs,TP, 0 , 11 , 1E+020 , 0 , 0 , 0
Vin,TP, 8 , 11 , 1E+020 , 0 , 0 , 0
```

Table 5 4 GHz DRO NETLIST

10 REFERENCES

[1] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," Proc. I.E.E.E., vol 54, pp. 329-330, Feb. 1966.

[2] B. Parzen, Design of Crystal and Other Harmonic Oscillators. New York:Wiley, 1983.

[3] W.P. Robins, Phase Noise in Signal Sources. London:Peter Peregrinus Ltd, 1982.

[4] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory," Proc. 44th Annual Frequency Control Symposium, pp. 536-541. May 1990.

[5] B. Parzen, "Clarification and a Generalized Restatement of Leeson's Oscillator Noise Model," Proc. 42nd Annual Frequency Control Symposium, pp. 348-351, June 1988.

[6] G. Montrose et al., "Residual Noise Measurements of VHF, UHF, and Microwave Components", Proc. 43rd Annual Frequency Control Symposium, pp. 349-359, May 1989.

[7] B. Parzen, Oscillator and Stability Analysis in the Immittance Domain. In preparation.

PROCEEDINGS OF 1991 ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

QUARTZ CRYSTAL OSCILLATORS WITH DIRECT RESONATOR HEATING

BRUCE LONG, GREGORY WEAVER
Piezo Crystal Company
100 K Street POB 619
Carlisle, PA 17013

(717) 249-2151

Abstract

This paper describes a newly patented, low power, ovenized oscillator with the crystal heated directly by means of a thin film resistive heater deposited on its surface. Mounting the resonator inside an evacuated enclosure forms a miniature "thermos" bottle. The only significant path for heat loss is through the resonator mounting ribbons and posts and these, being long and thin, have quite high thermal resistance. Pulse duration modulation controls resonator heating current. Phase noise, Allan variance, and aging all appear unaffected by direct resonator heating. Practical results for an oscillator employing these concepts are encouraging. Frequency-temperature stability better than $\pm 2.5 \times 10^{-8}$ from 0 to 70°C. has been achieved using a third overtone 10.230 MHz SC-cut resonator with room temperature power consumption of about 300 mW. Theoretical analysis shows substantial heating power reduction is possible by introducing additional thermal resistance in the form of a glass ring as part of the resonator mounting structure. Room temperature power consumption under 100 mW appears feasible.

Introduction

Quartz crystal resonators are the backbone of modern communications electronics. One observer remarked "Quartz is such an unexpected blessing its very existence might well be recalled in any debate over the benevolence of the creator". [1] But quartz crystal resonators are not perfect. In many applications the basic frequency temperature stability of the resonator is insufficient. Assuming the crystal cut has been wisely selected, and the angle precisely maintained during manufacturing, only two general techniques exist to improve oscillator frequency temperature stability. The first is compensation. In Temperature Compensated Crystal Oscillators (TCXO) temperature sensitivity is reduced by adjusting the oscillator circuit to compensate for temperature induced frequency shifts. The prevailing technique employs a thermistor network developing a temperature dependent correction voltage applied to a voltage variable capacitance thereby restoring frequency stability.

TCXO Limitations

Every resonator has a unique frequency temperature characteristic requiring custom compensation --a major liability in mass production. Low cost TCXOs have thermistor compensation networks designed solely upon the basis of the crystal frequency temperature curve. Better TCXOs are subjected to a requirement run revealing the frequency temperature behavior of the oscillator as a whole. A second frequency temperature test, the confirmation run, confirms correct compensation. Additional temperature testing and adjustment of the thermistor network is often necessary.

Temperature compensated oscillators have not improved greatly in recent years. The factors limiting TCXO performance are exhaustively researched and documented. One important factor is hysteresis. Repeated frequency temperature curves, or curves made in opposite directions fail to overlap. While speculation to the cause of hysteresis is plentiful, practical measures for its elimination or reduction are scarce. Process control and cleanliness have some effect but hysteresis is also in part an unavoidable intrinsic property of quartz. As a practical matter, hysteresis guarantees temperature compensated oscillators rarely perform quite as well in the field as they do during the factory confirmation run.

Quartz crystal resonators age. Accordingly most precision oscillators have a electrical or mechanical frequency adjustment. Unfortunately aging adjustment upsets temperature compensation. This inconvenient phenomenon is the trim effect. Aging adjustment changes the slope or linearity of the varactor tuning network. Correction voltage from the thermistor network is no longer accurate. Attention to tuning network linearity and using the same tuning network for both temperature compensation and aging adjustment reduces the trim effect. Some digitally compensated oscillators avoid the varactor tuning network altogether. Pulse dropping corrects oscillator frequency, circumventing the trim effect. However hysteresis still limits compensation accuracy and the missing pulses play havoc with the oscillator short term stability.

CH2965-2/91/0000-384 $1.00 © 1991 IEEE

TCXO stability also suffers if temperature changes too quickly. It is not possible to place the thermistor network directly on the resonator as it is mounted inside an evacuated enclosure for high Q and low aging. Thermistors mounted on the crystal case respond to sudden temperature changes before the resonator. The compensation network compensates for the wrong temperature causing large transient frequency shifts.

Finally high performance doubly rotated resonators cannot be used in TCXOs. TCXO crystals often have frequency temperature inflection points just outside the specified temperature range. This sets up a largely linear frequency temperature characteristic across the specified temperature range simplifying the compensation network but also demands a wide frequency adjustment range from the oscillator. TCXO resonators cannot be "stiff". Doubly rotated resonators especially the SC-cut have many desirable properties including lower phase noise, better aging and reduced acceleration sensitivity but are too stiff to temperature compensate. Consequently TCXOs are limited to AT-cut resonators and the improved performance possible with doubly rotated resonators remains out of reach.

Other compensation schemes develop correction voltages by means of a digital look-up table and a digital to analog converter, or by analog computation of the required cubic correction factor from a linear temperature sensor but these techniques also share the fundamental shortcomings mentioned before. In short TCXO technology is inadequate in many applications with further improvement unlikely.

Ovenized Oscillators

Of course undesirable frequency temperature excursions can be avoided by simply maintaining the resonator at a constant elevated temperature by means of electrical heating controlled by a thermostat or proportional controller. This is the basis of oven controlled crystal oscillators (OCXOs). Frequency temperature stability can be as much as one thousand times better than a TCXO but improved performance comes at a steep price. OCXO power consumption exceeds that of TCXOs by a factor of at least several hundred. Furthermore OCXOs, burdened by the need to surround the resonator and associated oscillator circuity first by a heated metallic enclosure and then by thermal insulation, are difficult to miniaturize.

Ovenized oscillators, like TCXOs, represent mature technology. Recent introduction of doubly rotated resonators has enabled improvement in the areas of aging, phase noise and frequency temperature stability but on the whole the pace of development has slowed. Power consumption remains a critical issue.

Oven Power Reduction

There are two ways to reduce oven power. First, make the heated volume as small as possible. This reduces the area from which heat is lost and, for fixed outside dimensions, maximizes the volume available for insulation. Only the oscillator components that are unavoidably temperature sensitive should be heated. Ideally the resonator should be the only temperature sensitive component. Most oscillators have some temperature sensitivity which is minimized by stable bias, avoidance of active device saturation and high permittivity ferrites, and by using high quality temperature stable components. Second, the thermal insulation must be as effective as possible. While urethane foam and fiber glass type material are commonly used, vacuum is the ideal thermal insulator.

By this line of reasoning the ultimate ovenized oscillator is a temperature insensitive amplifier and associated feedback loop components connected to a heated resonator, the latter suspended in an evacuated enclosure. This immediately suggests depositing a resistive thin-film heating element on an otherwise ordinary resonator inside a standard evacuated crystal enclosure forming a directly heated resonator. Figure 1 shows a directly heated resonator mounted inside a conventional TO-5 enclosure. The heating element, split into two semi-circular sections to avoid shorting the resonator electrodes, runs along the edge. The heating elements are applied with the same vapor deposition process used to deposit the resonator electrodes but with reduced thickness in order to increase electrical resistance.

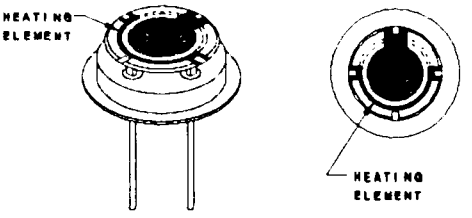


FIGURE 1
CRYSTAL RESONATOR
WITH THIN FILM RESISTIVE HEATER

Quartz Crystal Thermal Model

A quartz crystal resonator is a complex thermal system. Because the resonator enclosure is evacuated, heat is not conducted efficiently into the resonator. Convection is absent, and heat transfer by radiation is negliable for all but very large temperature differences. The only significant heat conduction path into the resonator is through the resonator support posts. The long thin posts present a substantial thermal resistance. Thermal resistance refers to a resistance to heat flow in an analogy with Ohms Law. Heat flow (in Watts) corresponds to electrical current. A heat source models as a current source. Thermal insulation offers a resistance to heat flow with units degrees Centigrade per Watt. In this analogy, a voltage drop corresponds to a temperature difference. Thermal resistance between the resonator and its enclosure is responsible in part for the poor performance of TCXOs in response to temperature slew. Ovenized oscillators are also affected. Because it takes time for heat to travel through the crystal leads into the resonator, ovenized oscillators continue to drift after the oven temperature has stabilized. The thermal time delay in both cases is analogous to a RC low pass filter with the specific heat of the quartz resonator functioning as the shunt capacitor. Quartz crystal resonators, especially AT-cuts, are further affected by mechanical stress accompanying the thermal gradients induced as heat flows from the mounting posts at the edge towards the active region underneath the electrodes in the center.

Directly Heated Resonators

The ideal of depositing a heating element directly upon the resonator surface has been around for some time - a patent was issued in 1969 [2] - but directly heated resonators have never been widely used. Two problems stand in the way of commercialization.

The first centers on the difficulty of controlling resonator temperature. Conventional ovenized oscillators sense the temperature of the heated enclosure surrounding the oscillator. This approach fails when the resonator alone is heated. In principal a sensor could be placed on the resonator surface but several difficulties arise on this score. The sensor mass loads and damps the resonator, and might compromise aging.

The second problem is more fundamental. Directly heated resonators experience significant mechanical stress coming from large thermal gradients underneath the thin-film heater. Quartz crystal resonators react to stress by changing frequency. Greenhouse, et. al., [3] encountered stress induced frequency shifts in an AT-Cut ovenized crystal oscillator using a directly heated resonator to reduce warm-up time, reverting to conventional heating during operation. Thermal frequency shift is a central issue in the development of practical directly heated resonators.

Resonator Temperature Control

Maintaining a constant resonator temperature despite ambient temperature variations is the function of the heater control circuitry. The control circuit transfer function, having units of Watts per degree, may be thought of as an equivalent thermal conductance ideally the reciprocal of resonator thermal resistance. It increases heating current by a set amount for each degree drop of the enclosure temperature.

Three parameters are needed to accurately estimate resonator temperature without direct sensor contact; case temperature, resonator-to-case thermal resistance, and resonator heater power. Case temperature is easily measured by a variety of temperature sensors. Because thermal resistance is determined primarily by the resonator mounting structure, it can be measured once for a given mechanical configuration and considered constant thereafter. Measured variation between resonators is small. Resonator heating power is monitored by observing heating current.

Efficient control of heating current is necessary. Conventional ovenized oscillators control heating element current with a variable resistance pass element, typically a transistor. Voltage division insures a sizable portion of the heating power dissipates in the control device instead of the heating element. This heat is wasted unless the control device is mounted on the temperature controlled structure, an impractical solution in the case of directly heated resonators. Control device dissipation also disrupts the otherwise linear relationship between heating current and element power.

Pulse width modulation of heating current overcomes both objections. Because the control device is either on or off and ideally dissipation free, heating power is applied directly and solely to the resonator. A single switching power supply integrated circuit provides all necessary control functions.

The control system described earlier is open loop having no feedback and it is not able to correct for errors in the controller gain setting or variation of resonator thermal resistance. Still it provides acceptable performance. Adding thermal insulation external to the resonator case is

a worthwhile improvement as it introduces a useful measure of thermal negative feedback. Power consumption is also slightly reduced.

A fully closed loop control system, capable of compensating for all gain mismatch errors can be built using a dual mode oscillator and resonator self-temperature sensing as described by Schodowski [4]. Two resonator modes are excited into oscillation simultaneously. One mode, being largely temperature insensitive, generates the oscillator output signal. The frequency of the second mode indicates resonator temperature.

Dual mode oscillators have generated a lot of interest lately. In the dominant concept, output pulses are dropped one at a time to maintain a constant average output frequency as the resonator frequency changes with temperature. This technique has potentially very low power consumption - no heating power is used - but the phase perturbations accompanying the missing pulses are severe. Phase locking and direct digital synthesis are proposed to overcome this limitation but these complicate an otherwise elegant scheme and increase power consumption.

Combining resonator self-temperature sensing and direct heating gives a stable, spectrally pure oscillator. A simple phase lock maintains a constant beat note frequency between the oscillation modes and insures constant resonator temperature. Crystal hysteresis is avoided and power consumption, while not as low as in the pulse dropping scheme, is still quite respectable when one considers additional phase noise clean up circuitry is not unnecessary.

Thermal Stress

Thermal stress induced frequency shift is a key issue affecting development of directly heated resonators. It has two components, a transient shift occurring whenever heating power is applied or removed, and a static shift present continuously while the resonator is heated. Figures 2 and 3 show the frequency shift of an oscillator using a directly heated 10 MHz, fundamental mode, FC-cut resonator. Transient shift stands out in Figures 2.1 and 2.2. With 300 mW applied to the heater, the resonator temperature rises about 60°C. The transient shift dies out quickly without noticeable effect provided the heater PWM frequency is sufficiently high.

FIGURE 2
THERMAL STRESS INDUCED
FREQUENCY SHIFT

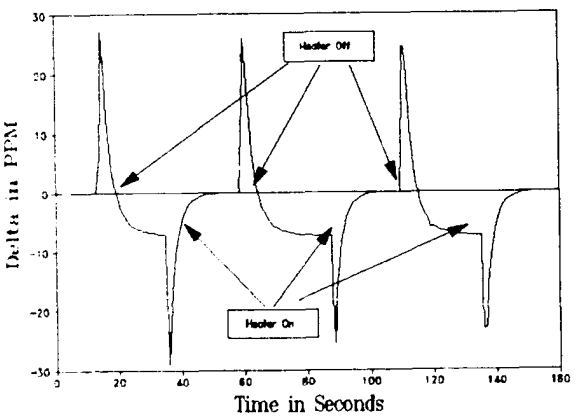


FIGURE 2.1

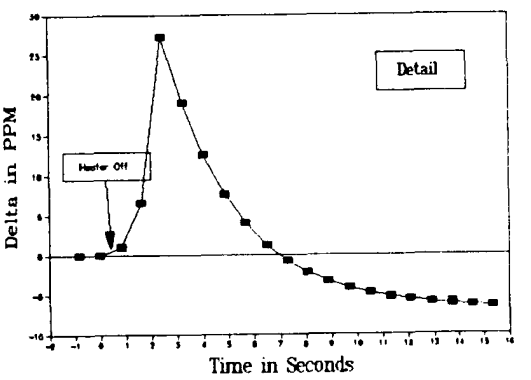


FIGURE 2.2

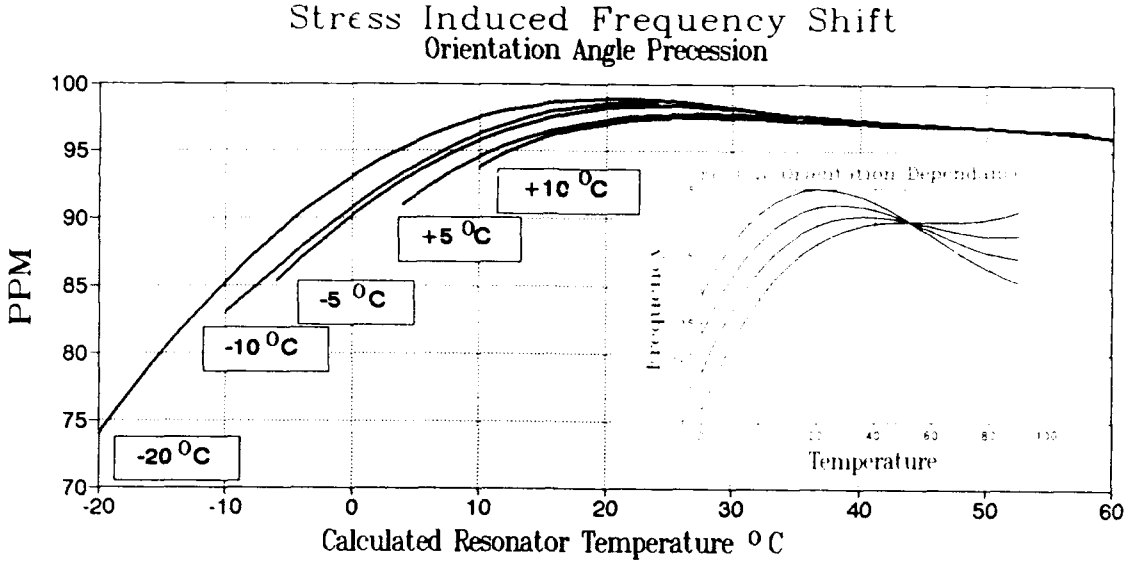


FIGURE 3

Figure 3 shows static frequency shift for five resonator heating powers. The horizontal axis represents estimated resonator temperature calculated from the product of the resonator to case thermal resistance, and the resonator heating power. Without thermal stress, all curves should overlap being simply the same frequency-temperature curve repeated five times.

Measurements of this type were our first attempt to measure resonator to case thermal resistance. We hoped to shift the curves until they overlapped. Then, for a given heater power, the horizontal shift equals the resonator temperature rise and division yields the resonator thermal resistance. [illegible] the curves don't overlap. A hump or bulge begins to form below 25°C, growing larger with heating power. The family of curves in Figure 3 bears a striking resemblance to FC-cut frequency-temperature curves as a function of orientation angle. Direct resonator heating appears to shift the effective resonator angle.

At first we thought stress induced frequency shift would obstruct practical exploitation of directly heated resonators but later observations suggest frequency shift is largely a linear function of applied heating power and can be dealt with by compensation, or by adjusting the heater control circuitry to slightly underheat the resonator. In the latter case, resonator temperature drops as the ambient temperature falls, but the thermal stress induced frequency shift compensates the frequency drop that would otherwise occur retaining frequency stability and slightly reducing power consumption. FC-cut resonators, which are only slightly stress compensated give acceptable results used in this fashion and SC-cut resonators, having greater stress immunity, work better. The exact relationship between thermal stress and frequency shift is unknown but the under-adjustment compensation scheme based upon the assumption of linearity works well. Dauwalter [5] reported a linear frequency shift accompanying compressive stress in 15 MHz AT-Cut resonators with less than one part in ten thousand non-linearity. This implies frequency shift compensation of DHXO oscillators can be simple and very accurate.

In the future we hope to modify the heating element shape in order to reduce resonator stress. The existing heater pattern applies heat more or less evenly around the outer edge of the resonator. Concentrating more heat near the resonator support posts which draw heat from the resonator is a step in the right direction.

Another untried approach sets the resonator operating temperature at the inflection point instead of the usual lower turn point. This relies upon the assumption

thermal stress causes a precession of apparent resonator angle without affecting the location of the inflection point. If correct, this means thermal frequency shift is absent at the inflection point.

Finally a method of increasing resonator thermal resistance is described latter in this paper. Decreased heating power consumption is the prime objective but thermal stress is also reduced. To summarize thermal stress induced frequency shift in directly heated crystal oscillators can be successfully circumvented and further prospects for improved performance via various stress reduction measures are promising.

Applications and Results

Our development effort so far concentrates on two specific embodiments. The first is a very low power oscillator intended for battery operation as encountered in rescue beacons. The most important factors in this application are power consumption and short term frequency stability under temperature slew. Phase noise, aging and frequency-temperature stability are secondary but still important factors and the expected ambient temperature ranges from -20 to + 60°C.

The second variation is a low power high performance oscillator suitable for use in portable instrumentation and navigation equipment. This oscillator consumes more power but the frequency-temperature stability is much better. The standard commercial temperature range, 0 to 70°C applies to this oscillator. The primary difference between the two versions is the choice of crystal. The first uses a 10 MHz, fundamental mode FC-cut resonator. The second employs a 10 or 10.230 MHz, third overtone SC-cut resonator.

The wide region of near-zero frequency-temperature coefficient exhibited by FC-cut resonators between approximately 30 and 70°C enables significant power savings applications needing limited temperature stability. Conventional ovenized oscillators heat the resonator well above the highest expected ambient temperature achieving excellent frequency temperature stability at the expense of power consumption. At average and low ambient temperatures power consumption is much larger than need be.

While rescue beacons can experience very high temperatures, daytime desert heat for example, less extreme temperatures are more likely especially for beacons floating in seawater. Battery size is limited and the system frequency temperature stability needs are modest. Taken together, trading temperature stability for increased battery life is attractive. Between 30 and 60°C,

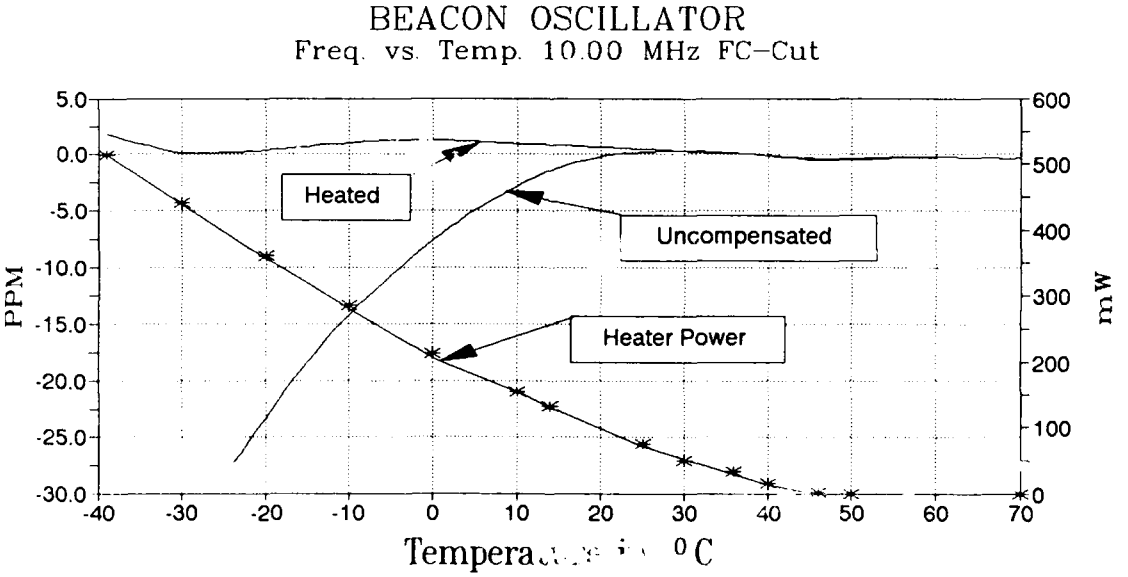


FIGURE 4

FC-cut temperature stability is sufficient without heater power. Below 30°C, applying heater current as the ambient temperature drops maintains resonator temperature and oscillator stability. Figure 4 shows the result. Frequency-temperature stability between -20 and +60°C is better than ± 1.2 ppm. The transition between heated and unheated operation begins at 30 degrees. Temperature slew at a 1.5°C per minute does not affect stability. Allen variance is quite good, superior to comparable TCXO performance.

A small production run of these devices showed quite acceptable unit to unit consistency. Initially, we thought each oscillator might require a unique controller thermal gain to compensate for unit to unit variations of the resonator thermal resistance and thermal stress compensation, however after experimentally determining the correct setting for the first oscillator, the remaining oscillators showed nearly identical performance without additional adjustment. Further work directed towards size and cost reduction is underway.

The second directly heated crystal oscillator (DHXO) product under development offers an order of magnitude improvement in frequency temperature stability compared to the best TCXOs, at on tenth the power consumption of conventional crystal oscillators. Figure 5 shows the frequency temperature performance of a laboratory prototype. Total deviation over 0 to 70°C is less than ± 2.5 x 10^{-8} with heating power consumption of 270 mW at 25°C.

The oscillator uses the Piezo standard low phase noise circuit with no modification or component sorting. Crystal frequency is 10.230 MHz. Close in phase noise is quite good, -123 dBc at 10 Hz offset in a one Hertz bandwidth and is unaffected by direct resonator heating. Figure 6 shows oscillator phase noise with and without resonator heating. The close in phase noise slope of 30 dB per decade extends inward to within 10 mHz of the carrier. The phase noise floor beyond 100 Hz does increase by about three dB and spurious at multiples of the heating PWM frequency are present. Spurious had been much higher before some obvious ground loops were eliminated. Even then the laboratory breadboard construction and layout was less than ideal. Insufficient isolation remains a likely cause of switching frequency spurious. On the other hand, spurious might be thermal stress modulation of the crystal resonant frequency but it must be noted heating frequency sidebands were no higher on the FC-cut DHXO which should, in principal, be more susceptible to stress modulation. Stress modulation can be reduced by converting the heater current pulse train into direct current by means of an LC filter and a commutating diode. Increased switching rate also helps. Allan variance is unaffected by resonator heating, Figure 7.

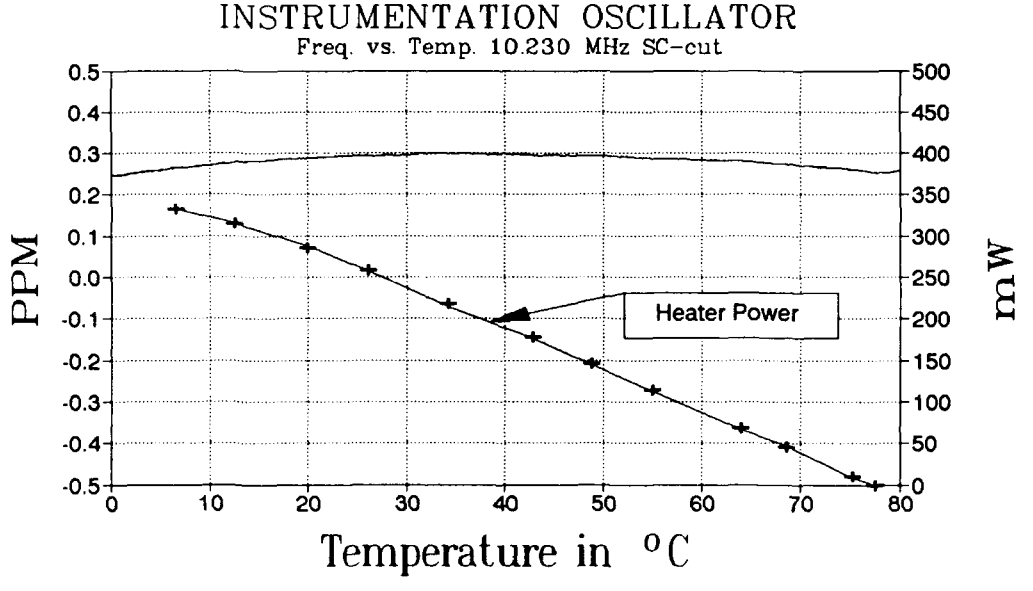


FIGURE 5

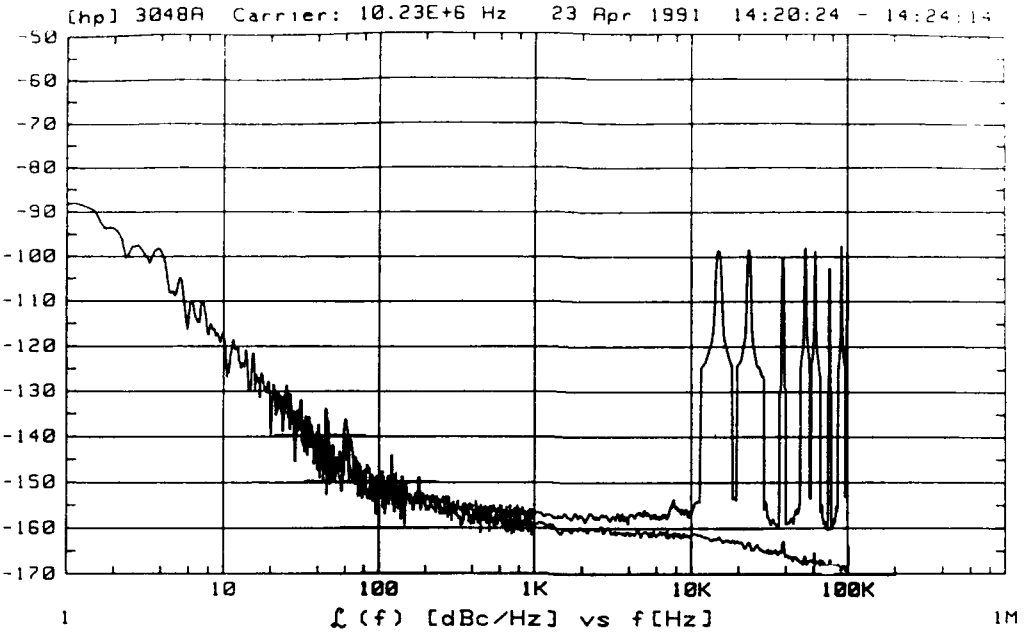


FIGURE 6
INSTRUMENTATION OSCILLATOR PHASE NOISE

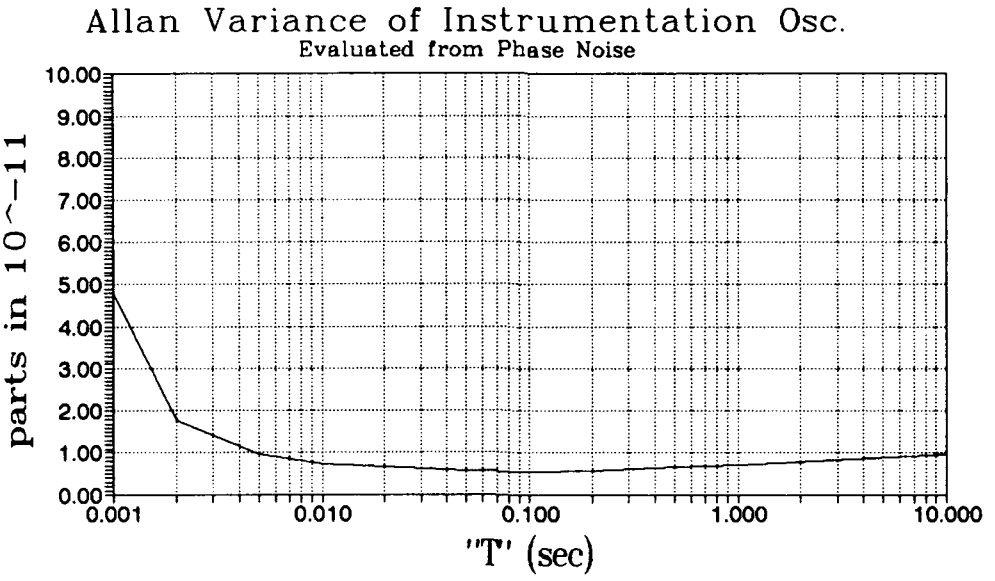


FIGURE 7

DHXO Power Reduction

A simple modification of the crystal mounting structure reduces DHXO heating power consumption. Thermal resistance of the mounting posts is increased by adding an insulative ring as illustrated in Figure 8. Resonator mounting ribbons attach midway between adjacent mounting posts providing maximum thermal resistance. Thin film conductors deposited upon the ring surface maintain electrical continuity. The insulative ring can be glass or quartz or any other thermally insulative material compatible with metallic surface thin films. A hand built prototype nearly halved heating power. Computer thermal modeling predicts much larger potential power savings suggesting the prototype thin film conductors might have been too thick. Reduced heating power brings about decreased stress induced frequency shift, easing concerns on that account. The insulative ring is also an ideal location for the temperature sensor. Reduced thermal time constant betrween the resonator and the sensor improves response to thermal transients including warm-up. Thermal resistance between the sensor and the ambient environment adds negative feedback to the control circuit reducing the sensitivity of the controller thermal gain adjustment.

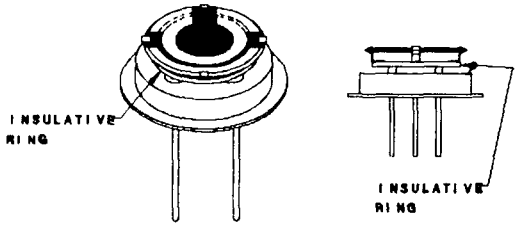


FIGURE 8
DIRECTLY HEATED RESONATOR
WITH INSULATIVE RING

Acknowledgments

The authors thank Charles Jensik and the management of Piezo Crystal Company for supporting the development of the directly heated quartz crystal. Additionally, we acknowledge the contributions of William Hanson. We also appreciate the encouragement and support of Dr. Joseph Suter and Mr. Joseph Wall of Johns Hopkins Applied Physics Laboratory.

References

[1] Sherman, John H., "Quartz Analogues," Proceedings 40th Annual Frequency Control Symposium, 1986.

[2] Garland, et. al., United States Patent 3,431,392, "Internally Heated Crystal Devices," Hughes Aircraft Company, Culver City, CA., March 4, 1969.

[3] Greenhouse, McGill, Clark, "A Fast Warmup Quartz Crystal Oscillator," Proceedings 32nd Annual Frequency Control Symposium, May 1978.

[4] Schodowski, S., "Resonator Self-Temperature-Sensing Using a Dual-Harmonic Mode Crystal Oscillator, Proceedings 43rd Frequency Control Symposium, 1989.

[5] Dauwalter, Charles, "The Temperature Dependence of the Force Sensitivity of AT-Cut Quartz Crystals" Proceedings 26th Annual Frequency Control Symposium, 1972.

Additional References

Rosati, Schodowski, and Filler, "Temperature Compensated Crystal Oscillator Survey and Test Results," Proceedings 37th Annual Frequency Control Symposium, 1983.

Rosati, Thompson, "Further Results of Temperature Compensated Crystal Oscillator Testing," Proceedings 38th Annual Frequency Control Symposium, 1984.

Filler, Raymond L., "Thermal Hysteresis in Quartz Crystal Resonators and Oscillators," Proceedings 44th Annual Frequency Control Symposium, 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PERFORMANCE TESTS ON AN MCXO COMBINING ASIC AND HYBRID CONSTRUCTION

A. BENJAMINSON, General Technical Services, Inc.
105 Clermont Avenue, Alexandria, VA 22304

BRIAN ROSE, Q-Tech Corporation
10150 W. Jefferson Blvd., Culver City, CA 90232

Introduction

This new Microcomputer-Compensated Crystal Oscillator employs hybridized crystal oscillator circuits combined with an Application-Specific Integrated Circuit and a Microcontroller. It provides both frequency and time accuracies of 30 parts per billion over the temperature range of −55 to +85 degrees Celsius with negligible warm-up time and low power consumption.

The ASIC contains the signal mixers, divider chains, counters, phase comparators, digital control logic and a Direct Digital Synthesizer. During calibration the frequency/temperature (F/T) characteristics of both the fundamental and the 3rd overtone of the SC-cut resonator are measured, as sustained by a Dual Overtone Crystal Oscillator [1]. Two 7th-order polynomials are fitted to the F/T data collected for each crystal and the coefficients are stored in the non-volatile memory of the MCXO. Through the microcomputer, the polynomial controls the correction frequency generated by the DDS which, when summed with the output of the 3rd overtone frequency oscillator, (using the phase-locked loop) produces the very stable temperature-corrected output frequency.

A block diagram of the MCXO is shown in Figure 1. Detailed explanations of this system and the factors controlling its performance were described at the 43rd and 44th Annual Frequency Control Symposia [2,3] and will not be repeated here.

This paper describes the performance of a prototype 10 MHz. design using a new ASIC initially developed by General Technical Services with LABCOM support. Additional research and development are continuing under the aegis of the Q-Tech Corporation, who will be producing more prototypes for market analysis.

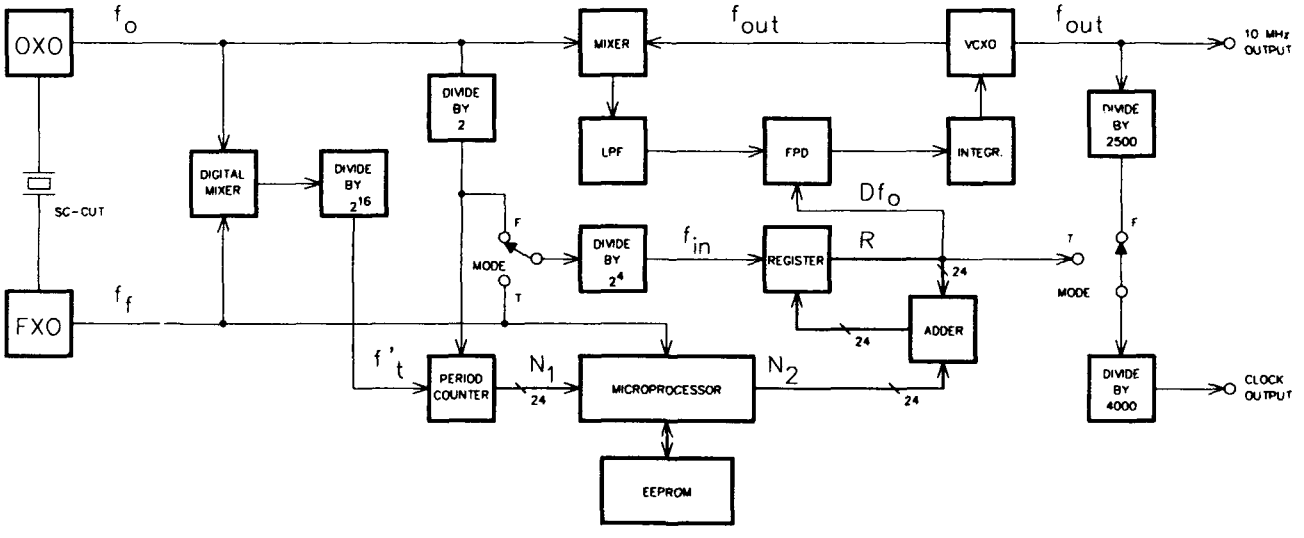


Figure 1. MCXO Block Diagram

CH2965-2/91/0000-393 $1.00 © 1991 IEEE

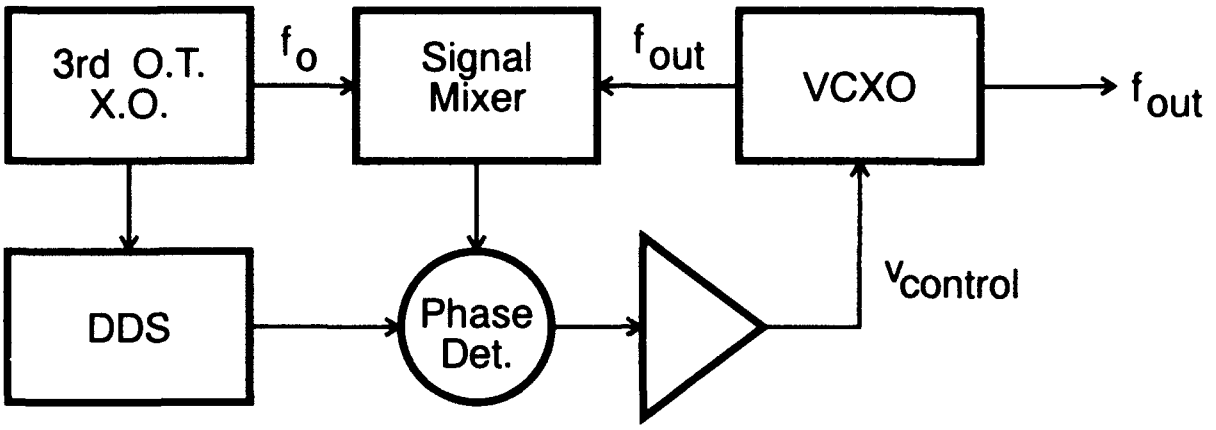


Figure 2a. Frequency Correction Method—Summing Loop

A Low Noise MCXO

The MCXO developed by G.T.S. is unique in that it uses a direct digital divider (DDS) to generate a correction frequency equal to the difference between the 3rd O.T. frequency of the SC-cut resonator and the desired 10 MHz. output (see Figure 2a). The two frequencies are summed directly in a PLL/VCXO circuit without the use of digital dividers, resulting in a stable temperature-corrected 10 MHz. output frequency.

This approach differs from designs employing the pulse-deletion technique, which is illustrated in Figure 2b. In this technique the 3rd O.T. frequency is always higher than the target frequency, requiring that pulses be deleted in an approximately periodic fashion to achieve the target frequency's repetition rate (e.g. 10 MHz.). Further processing is required and is accomplished through the use of two high-ratio digital divider chains, one following the pulse deleted signal, the other dividing the output of the VCXO. The dividers culminate in a phase detector which serves to lock the VCXO to the target frequency.

Dual-Mode Hybrid Oscillator

The dual-mode hybrid oscillator is constructed using two thick film substrates mounted in a butterfly type of package, approximately 1.1 x .97 x .17 inches in size. A four volt regulator and a "kill" circuit are included in the package. (The overtone oscillator must be turned off in certain modes.) Components are mounted with epoxy and interconnected with 1 mil. gold wire. The total current for both oscillators and associated buffers is less than 2 milliamperes. Since both oscillators are series-mode types, pretuning of the circuits is accomplished by using resistors in place of the crystal and then adjusting the L-C oscillator frequencies to within a few percent of the crystal resonant frequencies.

MCXO System Configuration

The majority of the digital circuitry is now contained in the ASIC, a custom Gate Array, manufactured by LSI Logic according to design specifications by Practical Micro Design. All signal routing, measurement and digital generation occur in the gate array.

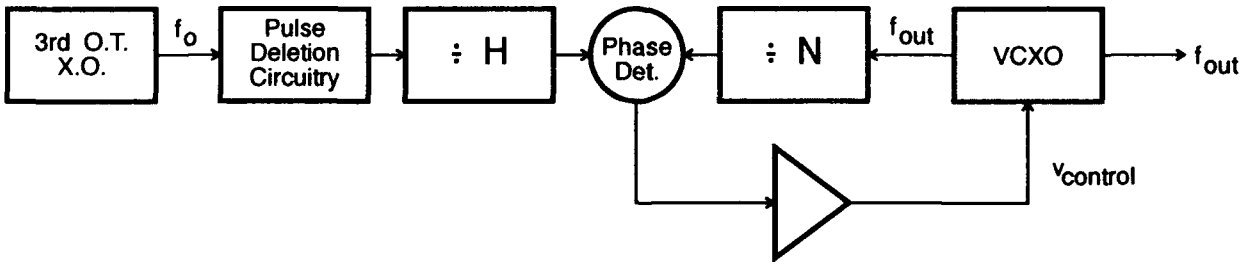


Figure 2b. Pulse Deletion Method—Cleanup Loop

Both the gate array and the microcomputer utilize CMOS technology to provide low power operation in compact 44 pin surface mount packages. The microcomputer is an Intel 87C51FA with 8K of EPROM, 256 bytes of RAM and four 8-bit parallel ports. The 87C51 interfaces with the gate array and performs all numeric computations, control functions, mode switching and similar tasks. The 87C51 is connected to a 93C46, which is a 64x16 serial NVRAM, which can retain the polynomial coefficients, even when the MCXO is without power.

In order to save power, the gate array and microcomputer together provide a variety of modes. These are selectable through a pair of pins on the MCXO connector. The primary mode is the Frequency Mode, which offers continuous corrections to the 10 MHz. output frequency, while simultaneously generating 1 PPS. time ticks.

The next mode is the same but with power conservation via the "sleep" function in which the microcomputer and portions of the gate array are shut down for up to 32 seconds at a time, as long as the MCXO is at a stable or slowly changing temperature. If the temperature starts to change while the unit is sleeping, the frequency error may increase temporarily.

In the Time-keeping Mode, only the 1 PPS output is available. The MCXO may sleep for as long as 128 seconds if the temperature is stable. In this mode, the VCXO is shut down, while the 3rd OT oscillator is powered up at the end of each sleep period only long enough to measure the crystal temperature.

The fourth mode is the Calibrate Mode, used for measuring the two SC-cut resonator frequencies as a function of temperature, and for serial communication with the microcomputer to download calibration information into the NVRAM. This mode requires special calibration equipment and is not intended for field operation.

Gate Array Elements

The gate array can "power down" the various crystal oscillators during modes when they are not required. Many multiplexers are employed in the gate array both to control mode switching and also to permit changes in the various divider ratios. This permits power vs. speed trade-offs to be arranged to meet a variety of applications.

The Mixer Circuits

The gate array contains two different signal mixer sections: the first mixes Fo, the overtone frequency with Fout, the output of the VCXO. This is an exclusive-or gate (XOR), followed by a low-pass L-C filter which extracts the difference frequency Fp. This signal is later compared with the output of the DDS in the gate array's frequency/phase comparator.

The second mixer section beats Ff, the fundamental mode frequency, against Fo after the latter is divided by three. This section offers a choice of mixers: the first is a single digital mixer, in which both the mixing and low-pass filtering are performed in the digital domain. The second is a dual analog mixer, comprising two XOR mixers in cascade. The first analog mixer heterodynes Ff and Fo/3. After low-pass filtering the difference frequency is mixed again with $Fo/3/2^6$ and presented to a multiplexer for selection by the reciprocal (N1)counter.

The N1 Counter

The N1 counter uses the selected mixer output as its gate input and Fo as the clock. This is a 24-bit counter which is read by the microcomputer eight bits at a time.

In order to permit changes in the gate period or to adjust to the selected mixer, the counter is preceded by a prescaler followed by an 8-to-1 multiplexer. The counter runs during the first half-cycle after each reset. This allows the microcomputer to sample the resulting N1 value as soon as possible after reset.

The Direct Digital Synthesizer

The DDS is also preceded by a prescaler with a 4-to-1 multiplexer, so that it can operate on a variety of source clock frequencies. Although the sideband level is lowest at the highest clock frequency, the power consumption is also proportional to clock frequency, so that power conservation dictates the sideband levels. The clock source for the prescaler is Fo/2 in the full frequency mode, or Ff in the timekeeping mode, in which mode the overtone and VCXO oscillators are shut down and the only output from the MCXO is 1 PPS. The DDS generates a 4 KHz square wave in this mode, using the correction polynomial for the fundamental frequency oscillator. This is subsequently divided by 4000 to produce the clock pulses.

The DDS itself comprises a 26-bit register and a 26-bit adder. The N2 number is loaded from the microcomputer into the adder, and the results stored in the reg-

ister. The high bit from the register provides the output signal (DFo) from the DDS. It is connected either to the frequency phase detector in the Frequency mode, or to the 1 PPS divider chain in the Time mode.

The resolution of the correction system is set by the DDS resolution which is better than 0.5 PPB in the Frequency mode and better than 1 PPM in the Time mode. In the latter case, these steps integrate toward zero (or the aging rate) as small changes in temperature occur.

The Phase-locked Frequency-correction Loop

The PLL provides the link between the output of the first mixer (Fp) and the output of the DDS (Dfo). It functions through the VCXO to equalize the two frequencies by forcing Fp to match DFo. Since DFo is the calculated error between Fo and the calibration standard (Fstd), and Fp is the difference between Fo and Fout, it results in Fout being equal to Fo - DFo, and therefore equal to Fstd.

Because the VCXO and the 3rd O.T. frequencies differ by as little 200 Hz.it is important that the initial search pattern begin with the VCXO frequency (Fout) at the high end of its range (Fout > Fo) and be forced to sweep down until the phaselock occurs with Fout= Fo - DFo. This is accomplished by the microcomputer, which forces Fout high and enables the Frequency/Phase comparator to drive the frequency to the lock point.

Divide by 10,000,000

The MCXO provides a 1 PPS output in both the Frequency mode and the Time mode. This is accomplished in the gate array with a segmented divide by 10 divider chain. When operating in the Frequency mode, the first segment divides the Fout signal by 2500 to create a 4 KHz. signal. This passes through a multiplexer to the second segment which provides the final division to 1 PPS. When the Time mode is in operation, the VCXO is disabled to save power, and the fundamental mode oscillator, operating through the DDS, provides the 4 KHz input to the second divider section.

Aging Correction

The MCXO is capable of automatic aging correction in the field. This adjustment is made on the 10 MHz. output frequency by modifying values in the polynomials that describe the frequency/temperature characteristics of the overtone and fundamental frequencies of the SC-cut resonator. It does not modify the oscillator circuits in any way and does not degrade their inherent stability. Aging correction occurs automatically whenever an external 10 MHz. reference signal (Fref, with TTL levels) is applied to the external reference terminal. The MCXO must be in the frequency mode during the correction interval and its temperature must be maintained at a constant value, preferably between 0° and 30° Celsius since the normal temperature correction cycle is interrupted during this aging correction process. When the external reference signal is applied, the microcomputer measures the beat frequency between the Fout and Fref frequencies to a resolution of 1 PPB. The correction period may be in excess of 200 seconds. A status output is available and can directly drive a low current LED if desired. The LED is normally off, but will begin flashing when an aging correction cycle is in progress. If the correction is successful the light will go off at the end of the cycle, else the light will remain lit if the correction can not be achieved. No changes are made to the polynomials if the cycle is aborted.

MCXO Performance Data

Frequency Tolerance

Figure 3a shows the data from a calibration run on a typical MCXO assembly. Frequency readings were recorded for both fundamental and overtone modes of the dual-mode crystal oscillator after passing through the ASIC and the microcomputer. The test chamber temperature is usually stepped in increments of 2°C beginning and ending at 25°C. The cycle reverses at -57°C and +87°C.

A least-squares 7th order polynomial is normally fitted to the data. The figure shows the residuals resulting from plotting the differences between the 3rd O.T. fre-

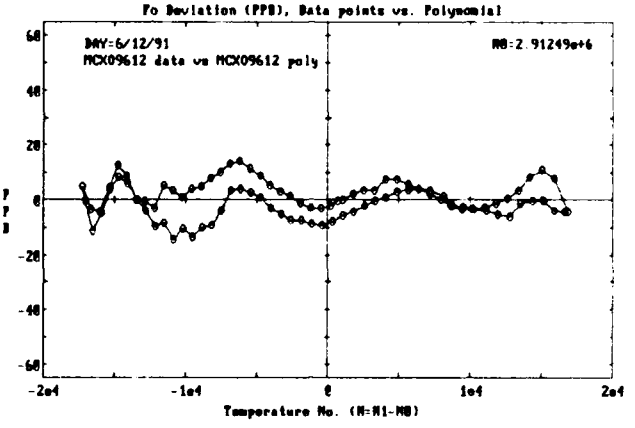


Figure 3a. MCXO Calibration Data

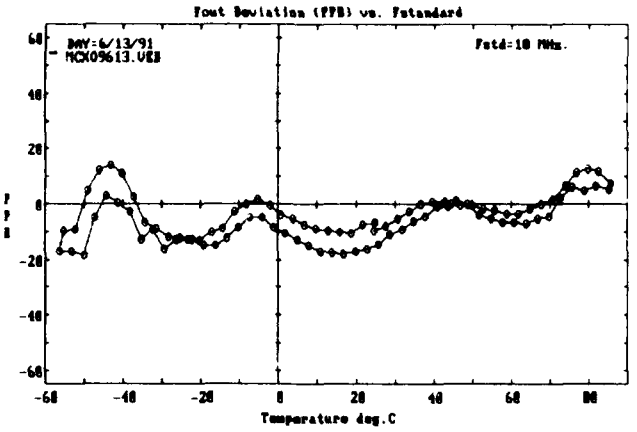


Figure 3b. MCXO Verification Run

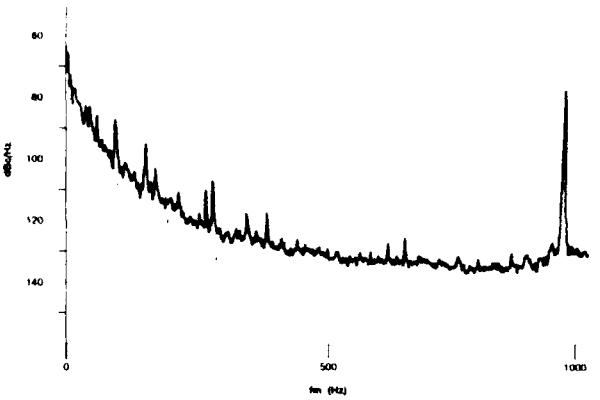


Figure 4. MCXO Phase Noise (dBc/Hz) at Fout = 10

quency and the polynomial. A similar curve would be obtained from a plot of the fundamental frequency. After the polynomial coefficients were stored in the NVRAM aboard the MCXO, a verification run was made over the same temperature range by recording the value of Fout vs. temperature. Figure 3b shows the difference between Fout and Fstd, the calibration reference frequency. Note the very close tracking between the two curves indicating the accuracy of the correction process in this type of MCXO.

Single-sideband Phase Noise Measurements

All phase noise measurements made to date on this new MCXO have been taken using the square wave output of the CMOS inverter buffering the output of the VCXO. We know that this does not produce the low levels of phase noise that a sinusoidal buffer amplifier can produce. While we are not displeased with the noise levels shown in Figure 4, we do plan to have a sinewave output buffer available on the production models. The noise floor level, which is approaching -140 dBc., can also be further reduced by raising the power level in the VCXO circuit whenever the increased power can be justified.

The spur shown at 1 KHz. offset from the carrier is generated by the beat frequency between Fo and Fout leaking into the output circuits. It is more than 70 dB. below the carrier.

Conclusion

The construction and successful testing of the present prototype models represents a significant step in the development of this important type of new oscillator. The MCXO using direct digital synthesis has been brought from the laboratory to the point where it can now be designed into manufacturer's systems. In addition, the demonstration of the use of modern digital techniques—ASIC and microcomputer—in a low power oscillator opens up the field of "intelligent oscillators" in a direction limited only by the imagination of the designers and users.

Acknowledgment

The authors wish to acknowledge the support of the U.S. Army's Electronic Technology and Devices Laboratory under Contract Number DAAL01-88-C-0804.

References

1. S. Schodowski, "Resonator Self-Temperature Sensing Using a Dual-Harmonic-Mode Crystal Oscillator." in Proceedings of the 43rd Frequency Control Symposium, 1989.

2. A. Benjaminson, S.C. Stallings, "A Microcomputer-Compensated Crystal Oscillator using a Dual-Mode Resonator." in Proceedings of the 43rd Annual Frequency Control Symposium, 1989.

3. A. Benjaminson, "Factors influencing Stability in the Microcomputer- Compensated Crystal Oscillator." in Proceedings of the 44th Annual Frequency Control Symposium, 1990.

FREQUENCY-TEMPERATURE CONSIDERATIONS FOR DIGITAL TEMPERATURE COMPENSATION

Raymond L. Filler
US Army Electronics Technology & Devices Laboratory (LABCOM)
SLCET-EQ
Fort Monmouth, NJ 07703-5601

Abstract

A fundamental requirement for compensation of the frequency-temperature (f-T) characteristic of a resonator is that the scheme used to estimate the frequency of the resonator from the temperature measurement be able to provide the necessary precision. The difference between the measured f-T data and the estimate of the frequency is a major limiting factor in any compensation scheme. If the measured f-T characteristic were sufficiently modeled by a simple function, hysteresis would be the only residual. Unfortunately, experimental results show that, in general, a simple function, for example, a polynomial of reasonable degree such as 10th, will accomplish compensation to only about 5×10^{-8}.

In this paper a description will be given of the anomalous behavior of the f-T characteristics of quartz resonators at the $pp10^8$ level which necessitates a more sophisticated estimation scheme. A review will be presented of estimation schemes, including least squares regression, minmax, segmented polynomials, and interpolation algorithms such as spline and Akima.

Introduction

The mathematical model for the frequency-temperature characteristic (f-T) of quartz crystal resonators has historically been a cubic polynomial.[1] The cubic model has been satisfactory for temperature compensating oscillators to about 1 ppm. The Microcomputer Compensated Crystal Oscillator (MCXO) using the dual-harmonic mode self temperature sensing method [2] requires a model capable of accuracy in the ppb range. Conventional temperature compensated crystal oscillators use a temperature sensing element which is remote from the actual vibrating quartz resonator. If the temperature is changing, the different thermal time constants produce a temperature gradient between the temperature sensing element and the vibrating quartz which results in "apparent" thermal hysteresis.[3] This "apparent" hysteresis has been one of the limiting factors in temperature compensation schemes. The dual-mode temperature sensing method eliminates "apparent" hysteresis and opens the way for much better temperature compensation. The cubic model is not adequate for realizing the higher accuracies. The techniques described below are useful for any digitally temperature compensated oscillator although one employing the dual-mode temperature sensing technique will make the best use of the results.

Algorithms

In order to perform digital compensation, an algorithm is required which will approximate the actual oscillator frequency from the measured temperature. For the dual-mode temperature sensing technique the temperature is represented by f_β, where

$$f_\beta = 3 f_1 - f_3 \qquad (1)$$

and f_1 and f_3 are the fundamental and third overtone frequencies, respectively. The frequency estimated by the algorithm will be referred to as $\hat{f}(f_\beta)$. In a digitally compensated oscillator, the difference between $\hat{f}(f_\beta)$ and the desired frequency is calculated and used by the compensation electronics to operate on the oscillator output to remove the effect of temperature.

The residuals, R_i, of an algorithm are the differences between the measured frequency and the estimated frequency at each measured temperature. That is

$$R_i = f(f_{\beta_i}) - \hat{f}(f_{\beta_i}); \; i = 1,\ldots,N \qquad (2)$$

where N is the number of calibration data points. The measure of quality of a temperature compensation algorithm is the maximum residual because the

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

definition of f-T accuracy is the maximum deviation from the specified nominal frequency.[4] It is assumed that the compensation electronics can eliminate any offset so that a zero residual corresponds to an output frequency exactly at nominal.

Two general classes of algorithms are examined in this paper. The first is multiple regression with integer powers of f_β as the basis functions, i.e., a polynomial fit. The familiar Method of Least Squares is an example of a multiple regression algorithm. The second class of algorithms is interpolation. For interpolation a function is found which passes through each data point while conforming to some continuity rules. Lagrange's polynomial formula and spline functions are examples of interpolation algorithms.[5]

Regression

To perform regression, a distance function with adjustable parameters is defined and then minimized with respect to those parameters.[6] The distance function used in this paper is the sum of the N residuals, R_i, raised to the exponent p. A weight is included to increase or decrease influence of selected data points. The general distance function, L_p, is

$$L_p = \sum_{i=1}^{N} w_i \, |f(f_{\beta_i}) - \hat{f}(\mathbf{A}, f_{\beta_i})|^p \qquad (3)$$

If p is chosen to be 2, the algorithm is the commonly used Method of Least Squares. If p is chosen to be ∞, the distance function is dominated by the maximum residual and the algorithm is a minimum maximum or a "minmax" algorithm.

To see that when $p=\infty$ the only term which survives is the maximum residual, consider the normalized distance function where each term is divided by the maximum residual. All of the terms, except the one involving the maximum residual, are a fraction raised to the power of infinity, i.e., each term is vanishingly small except the one involving the maximum residual, which is equal to 1.

Method of Least Squares

The most familiar least squares algorithm is polynomial regression. It uses the integer powers of the independent variable as the basis functions. The trial frequency function is

$$\hat{f}(f_\beta) = \sum_{k=0}^{D} A_k \, (f_\beta)^k \qquad (4)$$

where the D+1 coefficients are contained in the vector **A**, and D, the highest exponent in the polynomial is the degree of the polynomial. The vector **A** containing the D+1 coefficients is determined by minimizing the distance function

$$L_2 = \sum_{i=1}^{N} w_i \, |f(f_{\beta_i}) - \hat{f}(\mathbf{A}, f_{\beta_i})|^2 \qquad (5)$$

with respect to those D+1 coefficients. The RMS (or average) error of all of the residuals is minimized in Least Squares Regression.

This minimization problem is linear in the coefficients and reduces to a set of linear simultaneous equations which has a closed form solution.[5] An interesting property of a least squares fit, (only strictly true if a constant term is included) is that the algebraic sum of the residuals is identically equal to zero.

Fig. 1 shows the residuals of the least squares fit using a 10th degree polynomial on a 10MHz, 3rd O/T, SC-cut resonator.

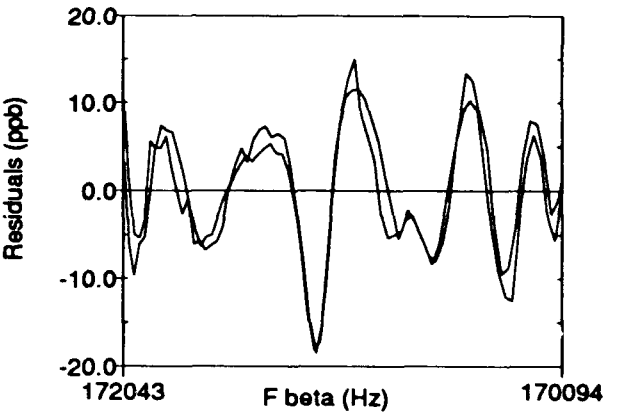


Figure 1 - Residuals from a least squares fit to a 10th degree polynomial.

The f_β range corresponds to a temperature range of -55°C to +85°C. The compensation obtained with this algorithm is about 19 ppb.

The relatively large undulations in the residuals are clearly not random. Therefore, a higher degree function is needed to properly describe the behavior of the resonator. The cause of this "microstructure" in the f-T characteristic is unknown but is repeatable for each device but not from device to device. The

task of designing an algorithm for temperature compensation would be simplified if these anomalies were smaller.

Minmax

A more appropriate distance function for the temperature compensation problem is one which conforms to the definition of stability, that is one which minimizes the maximum deviation or "minmax" criterion. The distance function to be minimized reduces to the minimization of

$$L_\infty = Max\left\{ |\hat{f}(f_{\beta_i}) - \hat{f}(\mathbf{A}, f_{\beta_i})| \right\} \qquad (6)$$

with respect to **A**. There is no closed form solution to this problem.[6]

Lawson's Algorithm: A technique known as Lawson's Algorithm uses the linear least squares technique described above with appropriate weights, w_i, in a recursive manner to converge to the minmax fit.[7] Lawson's approach is to use as the weights of the (j+1)th least square recursion the normalized residuals from the (j)th recursion. That is

$$w_i^{j+1} = \frac{w_i^j |R_i^j|}{\sum_{i=1}^{N} w_i^j |R_i^j|} \qquad (7)$$

When this method was applied to actual f-T data, the maximum residual did not always decrease monotonically, but after about 3 to 5 iterations a satisfactory minimum maximum was obtained.

Fig. 2 is a comparison of the least square fit from Fig. 1 to the fit using Lawson's Algorithm. The reduction of the maximum residual is about 30%. The penalty incurred is that the Lawson residuals are larger almost everywhere but at the maximum. This is a consequence of the fact that the Least Squares Method produces the minimum RMS residuals and any other algorithm must have the same or larger RMS error.

Direct Minmax: To test the Lawson Algorithm, and to see if there was a more convenient algorithm to find the minmax coefficients, a direct search was made of **A** space using a simplex [5] routine with the magnitude of the maximum residual as the value of the function to be minimized. This approach only reduced the maximum residual a few percent and was exceedingly time consuming.

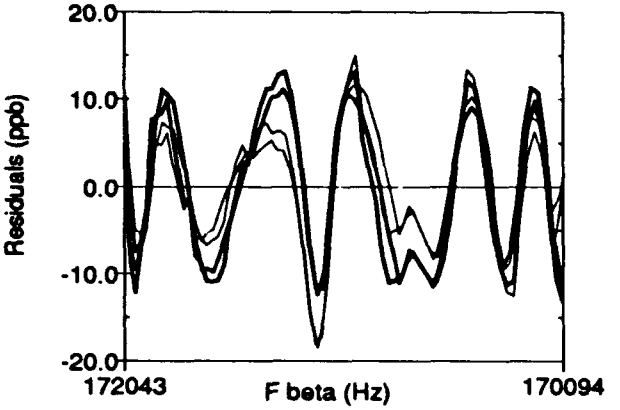


Figure 2 - Comparison of Least Squares (light) to the Lawson's Algorithm (heavy) fit for the resonator of Fig. 1.

Interpolation

In interpolation the task is to find a function which passes through every data point. There is no smoothing effect since every point has equal weight. Outliers or sharp features can produce unexpected results. The interpolation function can be a single polynomial of order N (degree N-1), where N is the number of data points, or the function can be piecewise continuous polynomials of lower degree. The continuity conditions and the degree of the polynomials define the different interpolation algorithms. The minimum continuity condition is that the function is continuous at each junction. Higher order continuity conditions consist of forcing the slope, or the slope and higher order derivatives to be continuous.

Interpolation algorithms fall into two broad classes, "local" and "global", depending on the relative influence of data far from the region where the interpolation is being applied. The more restricted the influence of remote data, the more "local" is the algorithm.

Linear Interpolation

The simplest interpolation scheme is to use straight lines between successive data points. This technique is not satisfactory because, as can be seen from Fig. 3, all of the residuals have the same sign. There will be no reduction of the accumulated error due to self-cancellation.

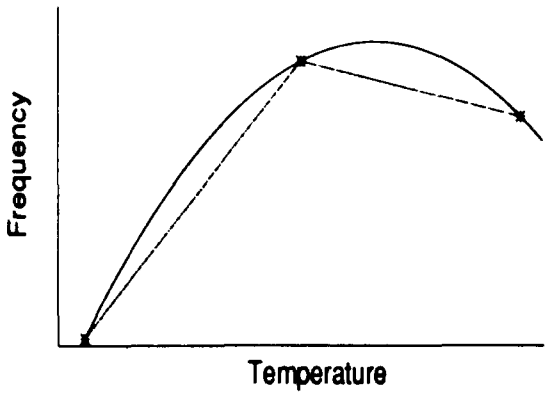


Figure 3 - Linear Interpolation.

Interpolation and Hysteresis

A problem that needs special treatment is the use of interpolation when the data includes hysteresis. Because of thermal hysteresis [8] there are two distinct f-T curves, one for the temperature increasing calibration run and one for the temperature decreasing run. To handle this, both sets of data must be kept separate and an interpolation done on each, individually. The resulting interpolant is the average of these two interpolants, as shown in Fig. 4.

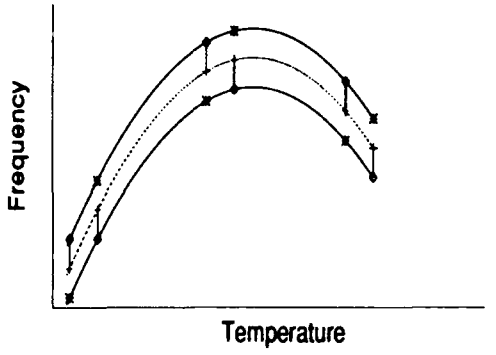


Figure 4 - Interpolation and Hysteresis - Measured Data (◊), Interpolated Value (.), Average Value (+).

Normally there are no residuals when interpolation is used since the interpolating function is a perfect fit at each data point. Hysteresis causes residuals even for the interpolation case. Of course, if more data is taken than is used in the interpolation, the quality of the algorithm can be tested independently .

Spline Interpolation

A very popular interpolation algorithm is the spline. A spline is a polynomial between neighboring pairs of points, with continuous higher order derivatives. A cubic spline is continuous through the second derivative.[5] A cubic polynomial has four coefficients. The condition of continuous second derivative allows computation of the first derivative at each data point. The two first derivatives and the values of the two data points are enough information to determine the four coefficients of each cubic interpolating function. A spline uses all of the data to determine the coefficients, so it is, therefore, a "global" algorithm.

Akima Interpolation

The problem with global interpolation algorithms is unwarranted undulations in the resulting curve. Fig. 5 shows a spline function applied to a constructed f-T anomaly.

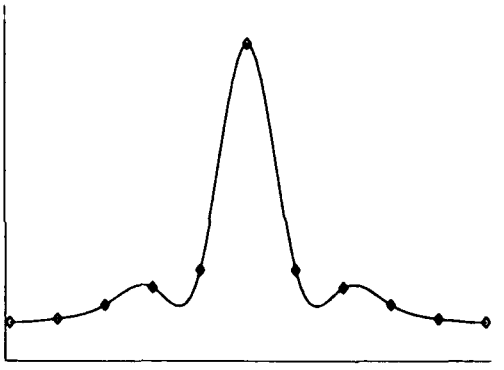


Figure 5 - Spline interpolation on sample data.

It can be seen that there are unwarranted undulations in the resulting curve. Akima [9] devised an algorithm which uses 5 nearest-neighbor points to eliminate the unwarranted undulations. Akima defined the slope at each data point as

$$t_i = \frac{|m_{i+1} - m_i|\, m_{i-1} + |m_{i-1} - m_{i-2}|\, m_i}{|m_{i+1} - m_i| + |m_{i-1} - m_{i-2}|} \quad (8)$$

with the special case of

$$t_i = \frac{m_{i-1} + m_i}{2} \quad (9)$$

when $m_{i-2} = m_{i-1} \neq m_i = m_{i+1}$, where m_k is the slope of the straight line from x_k to x_{k+1}. The cubic polynomial defined for the interval $x_j \leq x \leq x_{j+1}$ is

$$y = p_0 + p_1 (x - x_j) + p_2 (x - x_j)^2 + p_3 (x - x_j)^3 \quad (10)$$

where

$$p_0 = y_j$$

$$p_1 = t_j$$

$$p_2 = \frac{3 (y_{j+1} - y_j)/(x_{j+1} - x_j) - 2 t_j - t_{j+1}}{x_{j+1} - x_j} \quad (11)$$

$$p_3 = \frac{t_j + t_{j+1} - 2 (y_{j+1} - y_j)/(x_{j+1} - x_j)}{(x_{j+1} - x_j)^2}$$

Fig. 6 shows that the Akima algorithm applied to the same set of data as in Fig. 5 eliminated the undulations.

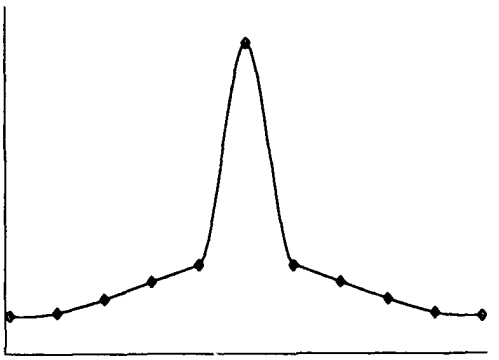


Figure 6 - Akima interpolation applied to the same set of data as in Fig. 5.

Local anomalies in the f-T characteristic of some devices require a local interpolation scheme. The Akima algorithm is the best of the ones investigated.

Other Interpolations Algorithms

Other interpolations algorithms investigated were 1) the osculatory algorithm[9] which uses three nearest neighbors and 2) a simple sliding parabola with no continuity conditions imposed. These perform almost as well as the Akima algorithm.

Interpolation Results

Fig. 7 shows the results of the Akima interpolation on the same data as was used in Figs. 1 and 2. It can be seen that the maximum residual is significantly lower than in Fig. 2.

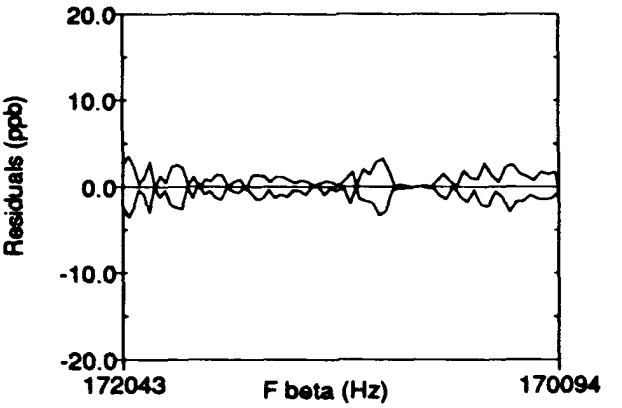


Figure 7 - Akima interpolation on the resonator of Figs. 1 and 2.

Regression with Multiple Segments

An oscillator which uses interpolation must store all of the calibration data in an ordered table, therefore, the memory requirements are larger than in a device which just stores regression coefficients. A technique which gives the superior results of interpolation with the ease of use of regression would be advantageous.

A compromise between interpolation and regression is to break up the temperature interval into segments and do a regression in each segment. This is similar to using a higher order polynomial but is more local and does not suffer from roundoff errors in the calculation. The calculation used in does not impose any continuity conditions at the breakpoints. This is not a problem as long as the residuals near the breakpoints are well within the required stability. Fig. 8 shows a three segment, 10th degree polynomial regression to the data used above. It can be seen that the residuals are similar to those of Fig. 7.

Fig. 9 shows the Akima interpolation and the three segment regression superimposed to show the similarity. An additional degree of freedom, which was not exercised in this example, is to vary the location of the breakpoints.

Segment No. vs. Degree vs. Algorithm

If the Akima algorithm is taken as the best maximum residual which can be obtained for a given f-T characteristic we can tabulate the degradation in the

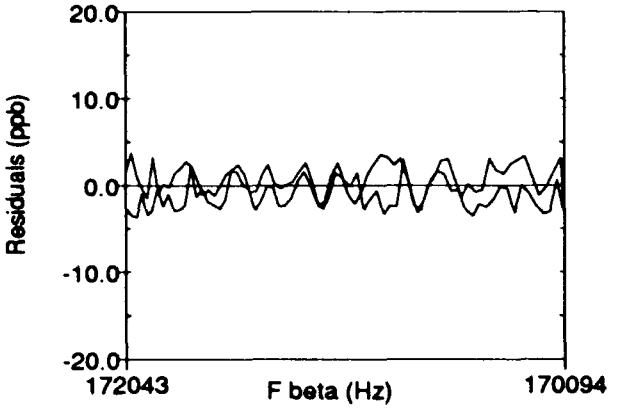


Figure 8 - Three segment, 10th degree polynomial, using Lawson's Algorithm to the data from Fig. 1.

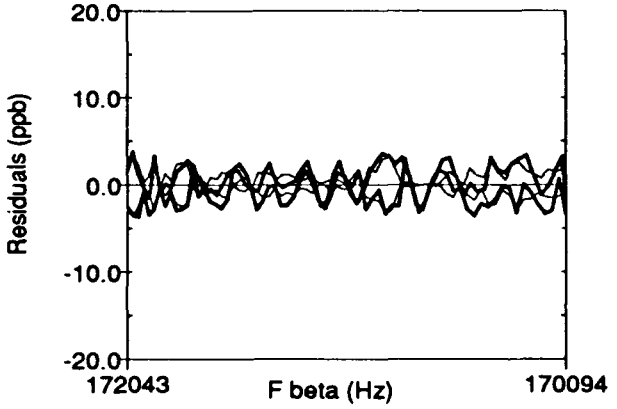


Figure 9 - Akima interpolation (light) and three segment Lawson regression (heavy).

result for other algorithms. Fig. 10 (following the references) shows data from 69 f-T runs. The 3 x 3 x 3 array shows polynomials of degree 6, 8 or 10, number of segments of 1, 2 or 3, and Least Squares, Lawson or Direct Minmax algorithms. As one would expect, the higher the degree and the greater the number of segments, the better the result. For the devices measured, an attractive tradeoff can be made between number of segments (2 or 3) and degree of the polynomial (8 or 10). The direct minmax algorithm did not give a significant improvement for the relatively large computing penalty.

Conclusions

There are many choices for f-T compensation algorithm which are superior to the usual least squares fit. These algorithms can be used with any digital compensation technique. The multisegment Lawson algorithm is the most efficient choice but any of the interpolation schemes will give optimum results at a cost of requiring more memory.

Smooth f-T characteristics and a good algorithm make it possible to achieve the limit set by the hysteresis of a device. Lower hysteresis is required to get the ultimate performance from digital temperature compensation.

References

[1] R. Bechmann, A. D. Ballato, and T. J. Lukaszek, "Higher-order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz," Proc IRE, 50, 1812,1822, 1962.

[2] S. Schowdoski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," Proc 43rd Annual Symposium on Frequency Control, pp. 2-7, 1989.

[3] R. Filler, "Measurement and Analysis of Thermal Hysteresis in Resonator's and TCXO's," Proc 42nd Annual Symposium on Frequency Control, pp. 380-388, 1988.

[4] MIL-O-55310, "Military Specification, Oscillators, Crystal Controlled, General Specification for," available from Military Specifications and Standards, 700 Robbins Ave., Bldg. D, Philadelphia, PA 19111-5094.

[5] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge, Cambridge University Press, 1986.

[6] J. R. Rice, The Approximation of Functions: Volume 1 - Linear Theory, Reading, Addison-Wesley, 1966.

[7] S. Ellacott and J. Williams, "Linear Chebyshev Approximation in the Complex Plane Using Lawson's Algorithm," Mathematics of Computation, 30,133, pp35-44, 1976.

[8] R. Filler, "Thermal Hysteresis in Quartz Crystal Resonators and Oscillators," Proc 44th Annual Symposium on Frequency Control, pp. 176-184, 1990.

[9] H. Akima, "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures," Journal of the Association for Computing Machinery, 17, 4, pp. 589-602, October 1970.

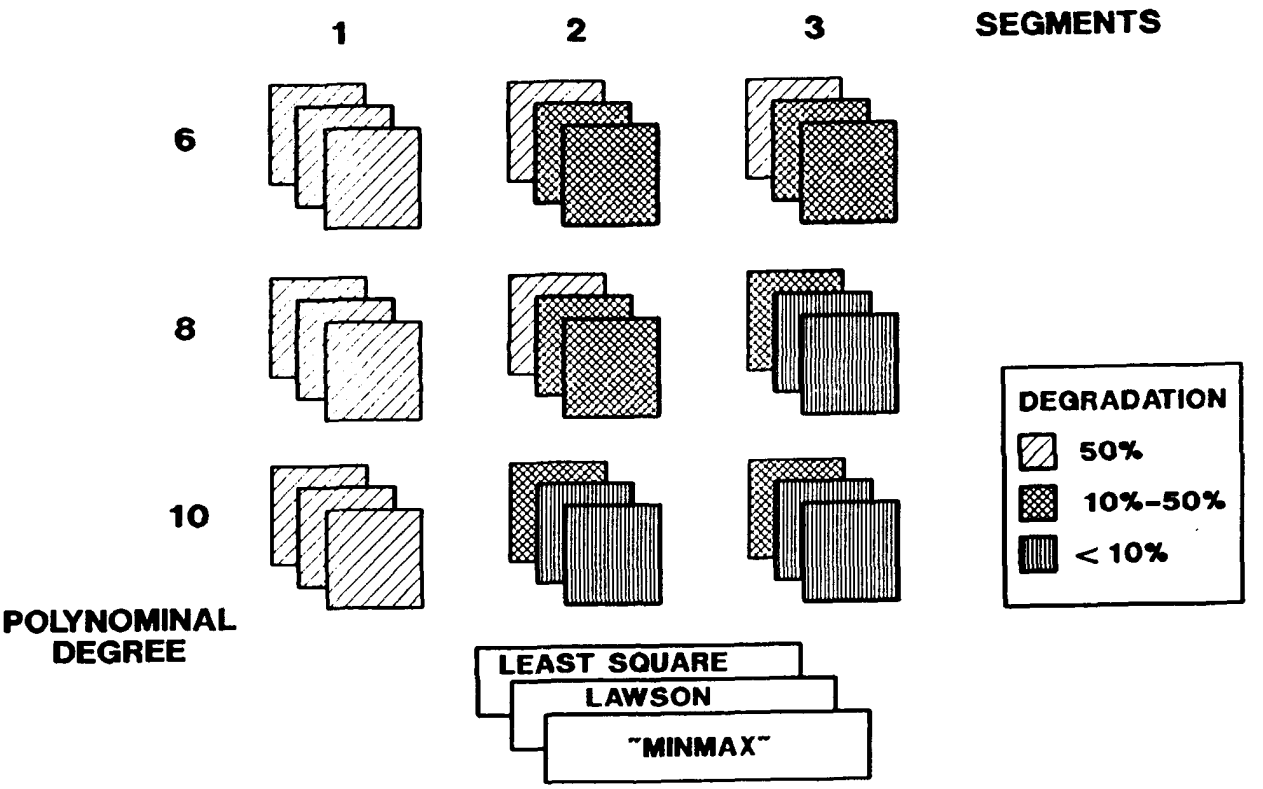


Figure 10 - Number of segments vs. Degree of Polynomial vs. Algorithm Relative to Akima Interpolation.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DIGITAL TCXO USING DELTA MODULATION

Makoto WATANABE* , Yukinori SAKUTA** AND Yoshifumi SEKINE**

* Nihon Dempa Kogyo Co., Ltd
3-2-4 Nishishinjuku, Shinjuku-ku, Tokyo 160, Japan

** Department of Electronic Engineering, College of Sience & Technology, Nihon University
7-24-1 Narashinodai, Funabashi-shi, Chiba 274, Japan

Summary

In Digital Temperature Compensated Crystal Oscillator, the information necessary to keep constant the oscillation frequency regardless of changes in temperature is already written in a memory. The information in the memory is stored as compensation information having 8 to 12 bits of data length per sampling temperature point. The memory size to store the amount of data that is needed to compensated for frequency variations over the total temperature range is enormous.

In order to reduce the memory size, we proposed a new method of using delta modulation to represent frequency compensation information with 1 bit of data length, which is a minimum unit of digital quantity, per sampling temperature point. The proposed method applies the delta modulation technique for encoding and decoding the frequency compensation information with respect to temperature changes.

Prototype D-TCXO has been produced in order to prove the effectiveness of the method. Our theory and experimental results agree with each other, and it has been confirmed that the proposed method is effective.

Introduction

To meet growing demands for more mobile radio stations, the bandwidth occupied by a single mobile radio station is reduced to increase the number of available channels. Smaller-size mobile radio stations are desired for better portability. As a result, crystal oscillators used in mobile radio units should be stabler in frequency and smaller in size.

Temperature compensated Crystal Oscillators (TCXO) are used as crystal oscillators in mobile radio units. The TCXOs can be classified into analog TCXOs (A-TCXO) and digital TCXO (D-TCXO) depending on how the frequency is controlled against temperature changes.

The A-TCXOs are composed of a temperature sensor, a resistor, and a varactor diode in combination, making up a circuit whose reactance is variable with temperature, and compensates for frequency vs. temperature characteristics of the crystal resonator. The frequency stability of mass-produced D-TCXOs is at most ± 0.5 ppm in the temperature range over 0 ℃ to +50 ℃.

The D-TCXOs digitally process temperature information, indicative of ambient temperature changes, from a temperature sensor to keep the oscillation frequency constant, thereby compensating for frequency vs. temperature characteristics of the crystal resonator.

Commercially available D-TCXOs have a frequency stability of ± 0.05 ppm over the temperature range from 0 ℃ to +50 ℃. Though the D-TCXOs have increased compensation accuracy for better frequency vs. temperature characteristics, they require a large memory size which is of 8 to 12 bits of data length per sampling temperature point for the compensation accuracy.

In this paper, we propose a D-TCXO which uses delta modulation for frequency compensation to reduce the frequency compensation information to 1 bit of data length, which is a minimum unit of digital information, per sampling temperature point.

CH2965-2/91/0000-405 $1.00 © 1991 IEEE

Conditions determined by temperature characteristics of a crystal resonator

The resonant frequency F of a typical AT-cut crystal resonator and the temperature T thereof are related as follows :

$$\Delta F/F_0 = A(T-T_0)^3+B(T-T_0)^2+C(T-T_0) \quad \cdots\cdot(1)$$

T_0 :reference temperature (+25℃)

F_0 :resonant frequency at T_0

A,B,C :coefficient determined mainly by the cut angle

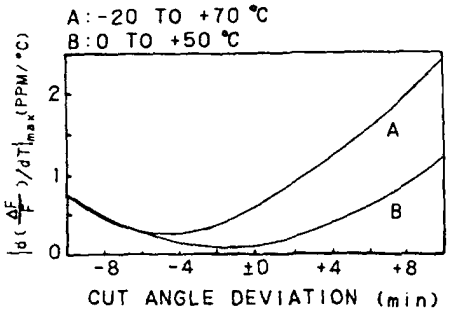


FIGURE1 MAXIMUM RATE OF CHANGE OF FREQUENCY VS. CUT ANGLE DEVIATION

FIGURE 1 shows the relationship between the maximum rate of change of the frequency and the cut angle of the AT-cut crystal resonator. The vertical axis represents the maximum rate of change of the frequency at each cut angle deviation, the rates of change being determined by differentiating the equation (1) with respect to temperature. The horizontal axis indicates the deviations from the cut angle deviation where the frequency temperature characteristics of the crystal resonator have a zero-temperature coefficient in the vicinity of normal temperature. If it is assumed that the maximum rate of change of the frequency with respect to a temperature change in the compensation temperature range is X (ppm/℃), then information per temperature step of Y/X (℃) will be necessary in order to obtain the frequency compensation accuracy Y (ppm), and the minimum number of temperature steps of information required within the entire compensation temperature range T (℃) will be T·X/Y. Specific examples are given on TABLE 1.

FREQUENCY COMPENSATION ACCURACY Y ≦ 0.1(PPM)

	TEMPERATURE RANGE (°C)	
	0 TO +50	-20 TO +70
CUT ANGLE DEVIATION (min)	-1.6	-2.4
$\left\|d\left(\frac{\Delta F}{F}\right)/dT\right\|_{max}$ (PPM/°C)	0.03	0.36
MIN. NUMBER OF STEPS	15	324

TABLE 1 MINIMUM NUMBER OF STEPS FOR COMPENSATION AND ITS CONDITIONS

Encoding principle of compensation information

A D-TCXO reads control information from a memory based on temperature compensation that is read out by a temperature sensor per clock pulse, and controls the oscillation frequency of the VCXO based on the control information.

A conventional D-TCXO approximates compensation information at each sampling temperature point with a staircase wave that is quantized into 8 to 12 bits of data length. The quantized compensation information is determined solely by its value at the temperature.

In this paper, the compensation information is encoded using delta modulation. The delta modulation adds a limitation that staircase approximation of compensation information can only have either one of two values, +W or -W (W represents the width of a quantizing step), as the difference of two quantized values between the neighboring temperature points. Thus, the compensation information at each temperature is approximated by staircase approximation which varies only by ±W, as shown in FIGURE 2.

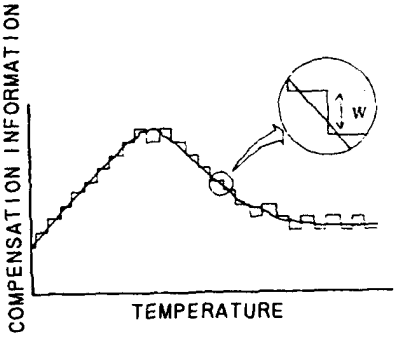


FIGURE 2 STAIRCASE APPROXIMATION OF CONPENSATION INFORMATION FOR DELTA MODULATION

By holding the difference between the neighboring steps to a constant of +W or -W, the compensation information can be encoded very simply.

In order to satisfy the target frequency compensation accuracy Y, the frequency quantizing step W has to be of a value equal to or less than Y.

The sampling time t (s) must satisfy :

$$t \leq Y/(X \cdot R) \quad \cdots\cdot(2)$$

where R is the rate of temperature change per time (℃/s). Even if the rate R is 0.02 ℃/s (1.2 ℃/min.),for example, over the temperature range from -20 ℃ to +70 ℃, the sampling time should be within 13.9 seconds, which may be 0.1 seconds or such in practice.

By setting the maximum frequency variations in a temperature step to the quantizing step W for frequency compensation, the frequency variation can be compensated for over the total temperature range.

Arrangement and operation of D-TCXO using delta modulation

FIGURE 3 is a block diagram of a circuit based on the above principle.

The D-TCXO is composed of a temperature information generator, a frequency compensation information generator, a D/A converter, and a VCXO.

FIGURE 4 is a block diagram of the temperature information generator.

The temperature information generator comprises a temperature sensor, a comparator, an up-down counter (U/D counter) , and a D/A converter. Temperature information is produced by generating a reference voltage which follows the output voltage of the temperature sensor, comparing the output voltage of the temperature sensor with the reference voltage in the comparator, and counting this temperature compared information $T_{(n)}$ with the U/D counter. The reference voltage is a DC voltage which is produced by counting the temperature compared information $T_{(n)}$ on the U/D counter in order to produce the temperature information $T_{x(n)}$, and by converting the latter information into an analog value. The voltage which corresponds to one step of the temperature compared information determines the resolution of the temperature information . The temperature information $T_{x(n)}$ serves as the reference voltage, and indicate in a memory address.

FIGURE 5 is a block diagram of the frequency compensation information generator. The frequency compensation information generator comprises a memory, a delay circuit, and a logic circuit which is shown in FIGURE 6.

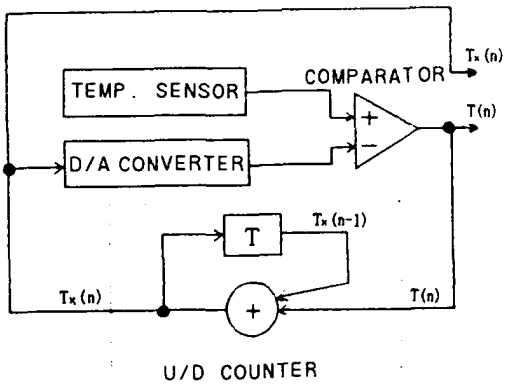


FIGURE 4 TEMPERATURE INFORMATION GENERATOR

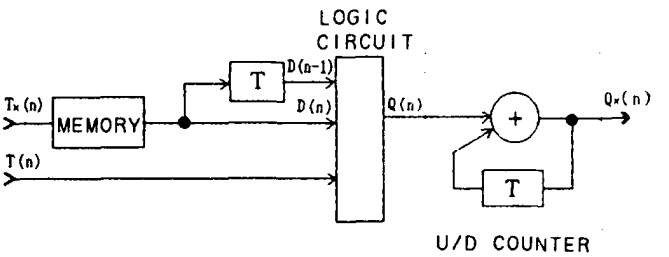


FIGURE 5 FREQUENCY COMPENSATION INFORMATION GENERATOR

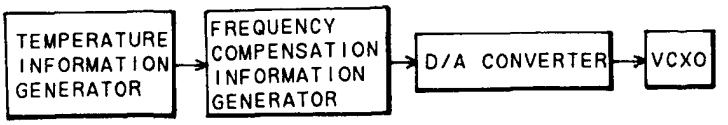


FIGURE 3 BLOCK DIAGRAM OF D-TCXO

T(n)	D(n)	D(n-1)	Q(n)	
0	0	0	1	$Q(n)=\overline{D(n-1)}$
0	0	1	0	
0	1	0	1	
0	1	1	0	
1	0	0	0	$Q(n)=D(n)$
1	0	1	0	
1	1	0	1	
1	1	1	1	

T(n) : TEMPERATURE COMPARED INFORMATION
D(n) : FREQUENCY COMPENSATION INFORMATION
D(n-1) : PRECEDING COMPENSATION INFORMATION
Q(n) : ARRANGED COMPENSATION INFORMATION

LOGIC FORMULA

$$Q(n)=T(n)\cdot D(n)+\overline{T(n)}\cdot\overline{D(n-1)}$$

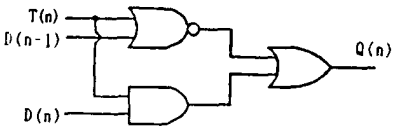


FIGURE 6 TRUTH TABLE AND ITS LOGIC CIRCUIT FOR ARRANGING THE COMPENSATION INFORMATION

In case that the memory stores the frequency compensation information for temperature rise, and that the temperature should fall, the frequency compensation information has to be output by inverting the memory output of the preceding control cycle through the logic circuit.

FIGURE 6 shows a truth table and a logic circuit for achieving the above operation. The logic circuit generates arranged compensation information $Q_{(n)}$ with temperature compared information $T_{(n)}$ and frequency compensation information $D_{(n)}$ at a time n, and $D_{(n-1)}$ at a time n-1.

The information $Q_{(n)}$ is counted by the U/D counter to produce control voltage information $Q_{x(n)}$ for the VCXO. The control voltage information $Q_{x(n)}$ is converted by the D/A converter into an analog voltage, which is applied to the control voltage terminal of the VCXO to control the frequency.

Design example

In order to confirm the operation of the D-TCXO under study, a prototype was designed to achieve a target frequency stability of ±0.2 ppm over the temperature range from 0 ℃ to +50 ℃.

The frequency vs. temperature characteristics of a VCXO used in this case is shown in FIGURE 7.

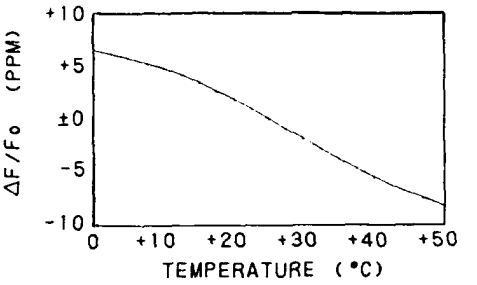


FIGURE 7 FREQUENCY VS. TEMPERATURE CHARACTERISTICS OF VCXO BEFORE COMPENSATION

FIGURE 8 illustrates the characteristics of a control voltage for keeping constant the oscillation frequency of the VCXO. The measured frequency control voltage sensitivity of the VCXO was 2.58 ppm/V.

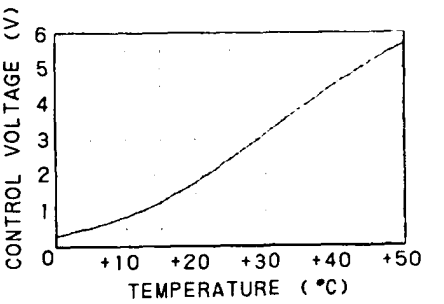


FIGURE 8 CONTROL VOLTAGE FOR COMPENSATION

In designing a prototype, a temperature step to achieve the target frequency stability was determined, first of all. The frequency variation per temperature step was designed to 0.1 ppm or less in consideration of margin to be added to the target frequency stability of 0.2 ppm.
Therefore, calculating from the above meintioned control voltage sensitivity, a change in the control voltage per temperature step has to be 0.1 (ppm)/2.58 (ppm/V) = 0.039 (V) or less.
The change rate of the control voltage of the VCXO with respect to temperature changes was 0.14 V/℃ at maximum. Then, the temperature step has to be 0.039 (V)/0.14 (V/℃) = 0.27 (℃) or less.

The prototype circuit designed by the above conditions had a control voltage step of 0.039 V and a temperature step of 0.25 ℃.

Consequently, the frequency change per temperature step was 0.1 ppm at maximum.

Experimental Results of the prototype

The prototype D-TCXO exhibits frequency vs. temperature characteristics as shown in FIGURE 9.

It can be seen from FIGURE 9 that the frequency stability is maintained within ±0.15 ppm over the temperature range from 0 ℃ to +50 ℃.

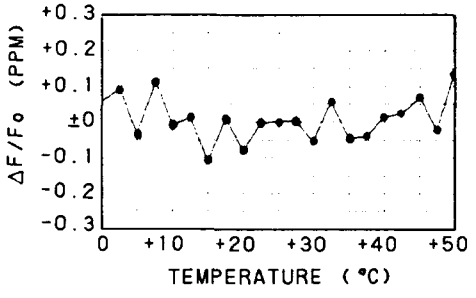


FIGURE 9 FREQUENCY VS. TEMPERATURE CHARACTERISTICS OF VCXO AFTER COMPENSATION

Conclusions

As described above, the memory size, which has conventionally required 8 to 12 bits of data length per sampling temperature point, is reduced to 1 bit by using delta modulation in the frequency compensation technique for D-TCXOs.

Our research proved that the delta modulation technique is so effective especially for downsizing D-TCXOs .

Delta modulation method makes possible to provide the flexible design technique in variation of uses.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Temperature Compensated Crystal Oscillator Employing New Shape GT Cut Quartz Crystal Resonator

Hirofumi KAWASHIMA and Kenji SUNAGA

Seiko Electronic Components Ltd.
Tochigi-shi 328, Japan

Abstract

This paper describes a temperature internally compensated crystal oscillator employing a new shape GT cut quartz crystal resonator formed by an etching method. It was reported[1] [6] by one of the authors that the new shape GT cut quartz crystal resonator has very small frequency deviation in a wide temperature range, because an absolute value of its third order temperature coefficient is much smaller than that of the conventional GT plate.

An object of this paper is to clarify if drain output resistance Rd and resistance Rqs connected in series to the resonator in elements constructing a CMOS oscillator circuit which suppress spurious vibrations, influence oscillation frequency and frequency temperature behavior, and to obtain a temperature compensated quartz crystal oscillator, which is called "TCXO・MGQ", with suppression of the spurious vibrations versus temperature.

In analysis procedure, first, a CMOS quartz crystal oscillator circuit is transformed into an equivalent circuit and an equation which gives oscillation frequencies is derived from Kirchhoff's law.

Second, since the equation is given as a function of Rd and Rqs, it is theoretically forecast and experimentally examined that Rd and Rqs influence the oscillation frequency and the frequency temperature behavior. Consequently, an increase of Rd and Rqs is found to raise the oscillation frequency for principal vibration a little bit and to cause the first order temperature coefficient for principal vibration to shift to a further negative value respectively.

Finally, a temperature compensated crystal oscillator which is called "TCXO・MGQ" is found to be obtained with frequency deviation within +/- 0.5*ppm* to nominal frequency for principal vibration and +/-2.5*ppm* over a wide temperature range of −30℃ to +85℃ without any external temperature compensation by thermister and so on by employing the present GT cut quartz crystal resonator.

§1 Introduction

Recently, a quartz crystal oscillator (TCXO) which is used in communication equipment and consumer products of principally handy type as a frequency standard, is excellent in frequency temperature behavior, simultaneously, a further miniaturized oscillator with low consumption power is earnestly desired and expected. It was reported[7] by one of the authors that, as a quartz crystal oscillator satisfying the desire and expectation, a CMOS oscillator circuit employing a new shape GT cut quartz crystal resonator is very suitable. The oscillation frequency and the frequency temperature behavior for principal vibration was then found to vary remarkably with particularly, load capacitance C_L in elements of the circuit construction. However, spurious vibrations occur under some conditions in constructing an oscillator, a quartz crystal oscillator with suppression of spurious vibration is, therefore, desired and expected.

An object of this paper is to provide a temperature compensated crystal oscillator which is called "TCXO・MGQ" with suppression of spurious vibrations over a wide temperature range. Spurious vibration is mainly classified into two vibrations which are piezoelectrically excited and caused by reflection of elastic wave at the edge of a resonator. The former shows strong spurious vibration and the latter shows comparatively weak response. The strong spurious vibration can be eliminated by selection of resonator's shape and size, but, the weak spurious response becomes a problem. This spurious vibration is closely related to level to drive a resonator and can be suppressed by taking small drive level. It is possible to obtain the small drive level by inserting drain output resistance Rd and resistance Rqs connected in series to a resonator in an oscillator circuit. However, it is predicted that oscillation frequency and frequency temperature behavior vary with Rd and Rqs, because a new shape GT cut quartz crystal resonator is a coupling resonator[3].

Another object of this paper is clarify if Rd and Rqs in elements constructing an oscillator circuit influence oscillation frequency and frequency temperature behavior.

In analysis procedure, first, since a CMOS inverter, as is already known, can be linearly expressed by using a voltage source and drain resistance for a small signal, the present oscillator circuit is transformed into an equivalent circuit. Next, an equation of the condition which gives oscillation frequencies is theoretically derived from Kirchhoff's law. Furthermore, an imaginary part of impedance for the present quartz crystal resonator is calculated. From these results an equation of oscillation frequencies is derived. Since the obtained oscillation frequencies are given as a function of Rd and

CH2965-2/91/0000-410 $1.00 © 1991 IEEE

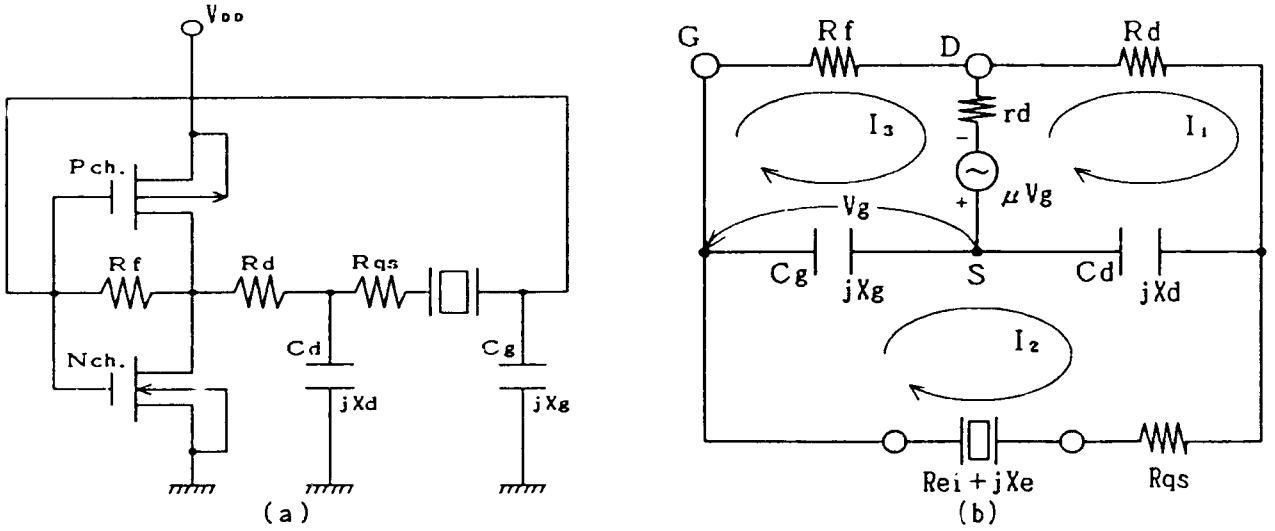


Fig. 1 Construction of CMOS quartz crystal oscillator circuit (a) and its electrical equivalent circuit (b).

Rqs, it is theoretically forecast and experimentally examined that those influence oscillation frequency and frequency temperature behavior. As a result, an increase of Rd and Rqs raises oscillation frequency of principal vibration a little bit, and with respect to temperature coefficient, they cause the first order temperature coefficient α, to shift to a further negative value.

Therefore, by selection of Rd, Rqs and load capacitance C_L which is given as a function of gate load capacitance Cg and drain load capacitance Cd, it is shown that a temperature compensated crystal oscillator (TCXO・MCQ) with oscillation frequency of principal vibration finely adjusted to the nominal frequency, and suppression of spurious response and extremely stable frequency in a wide temperature range of −30℃ to +85℃ is obtained without any external temperature compensation by thermister and so on. Let us describe concretely below.

§ 2 Analysis Procedure

2.1 Construction of CMOS quartz crystal oscillator circuit

Figure 1 shows a construction diagram (a) of a CMOS quartz crystal oscillator circuit and its electrical equivalent circuit (b). The circuit (a) comprises a CMOS inverter, drain load capacitance Cd, gate load capacitance Cg, feedback resistance R_f, drain output resistance Rd, power source V_{DD}, a new shape GT cut quartz crystal resonator and resistance Rqs connected in series to the resonator. Since the CMOS inverter, as is already known, can be linearly expressed by employing a voltage source μVg (μ:Amplification factor, Vg:Gate side voltage) and drain resistance rd for a small signal, the circuit (a) can be equivalently expressed by the circuit (b). In addition, the impedance of the new shape GT cut quartz crystal resonator which is a coupling resonator is expressed by a summation of real part Rei (Resistance part, i=1.2, i=1:Principal vibration, i=2:Subvibration) and imaginary part Xe (Reactance part). Accordingly, an equation of the condition to give oscillation frequency is calculated from the circuit (b).

2.2 Derivatives of oscillation frequency

In the circuit of Fig. 1 (b), let us take electric currents I_1, I_2, I_3 which flow in each loop. Simultaneous equations are obtained by applying Kirchhoff's second law in each loop. Furthermore, the electric currents I_1, I_2, I_3 are not zero, therefore, in order to obtain a nontrivial solution in the equations, the following relation must be satisfied:

$$\begin{vmatrix} rd+Rd+jXd & j(\mu Xg-Xd) & -(rd+j\mu Xg) \\ -jXd & Rei+Rqs+j(Xe+Xg+Xd) & -jXg \\ -rd & -jXg(1+\mu) & Rf+rd+jXg(1+\mu) \end{vmatrix} = 0 \quad (1)$$

Accordingly, as equation (1) which gives the oscillation frequency condition and the amplitude continuation condition, is expressed by a summation of a real part and an imaginary part, from the imaginary part, Xe is given as

$$Xe = \frac{\left[-\left(1+\frac{rd\,Rd}{Rf\,(rd+Rd)}\right)(Xg+Xd) - \frac{Rei+Rqs}{rd+Rd}\left(1+\frac{rd}{Rf}\right)Xd - \frac{Rei+Rqs}{Rf}\left(1+\frac{Rd}{rd+Rd}-\mu\right)Xg \right]}{\left[1+\frac{rd\,Rd}{Rf\,(rd+Rd)} - \frac{Xd\,Xg\,(1+\mu)}{Rf\,(rd+Rd)} \right]} \quad (2)$$

In addition, when angular frequency ω is taken, the signs Xg and Xd are given as $Xg=-1/\omega Cg$ and $Xd=-1/\omega Cd$. Substituting these relations into equation (2), an

equation of the condition to give oscillation frequency is further simplified as follows, by taking into account R_i, rd, Rd.

$$Xe = \frac{1}{\omega C_L} \cdot \frac{1}{1 - \frac{1+\mu}{Rf\,(rd+Rd)\,\omega^2\,CgCd}} \tag{3}$$

$$\frac{1}{C_L} = \frac{1}{Cg} + \frac{1}{Cd}\left(1 + \frac{Rei+Rqs}{rd+Rd}\right) \tag{4}$$

Therefore, by calculating reactance Xe of the quartz crystal resonator, the oscillation frequency can be derived from equation (3). Since Xe[7] has been already calculated when a coupling between principal vibration and subvibration of the quartz crystal resonator is taken as capacitance coupling, let us employ the result, consequently, Xe and Rei are, respectively, given as the following equations, taking into account that the coupling between principal vibration and subvibration is extremely weak in the calculation of Rei.

$$Xe = -\frac{1}{\omega Co}\cdot\frac{\left(\frac{\omega^2}{\omega_1^2}-1\right)\left(\frac{\omega^2}{\omega_2^2}-1\right)-\frac{C_1C_2}{Cm^2}}{\left[\left(\frac{\omega^2}{\omega_1^2}-1\right)\left(\frac{\omega^2}{\omega_2^2}-1\right)-\frac{C_1C_2}{Cm^2}-\left\{\frac{C_2}{Co}\left(\frac{\omega^2}{\omega_1^2}-1+\frac{C_1}{Cm}\right)+\frac{C_1}{Co}\left(\frac{\omega^2}{\omega_2^2}-1+\frac{C_2}{Cm}\right)\right\}\right]} \tag{5}$$

$$Rei = Ri\left(1+\frac{Co}{C_L}\right)^2 \qquad (i=1,2) \tag{6}$$

where ω_1, ω_2 : Uncoupled angular frequencies of principal and subvibrations
C_o : Shunt capacitance
Cm : Coupling capacitance
C_1, C_2 : Motional capacitance of principal and subvibrations
R_1, R_2 : Series resistance of principal and subvibrations

Accordingly, the oscillation frequency can be easily derived by substituting equation (5) into equation (3). In addition, substituting equation (6) into equation (4) and taking $(Co/Cd)^2 \ll 1$ into consideration, load capacitance C_L is obtained. First, oscillation frequencies $f_\pm$ are as

$$f_\pm^{\,2} = \frac{1}{2}\left[(f_1'^2+f_2'^2) \pm \sqrt{(f_1'^2-f_2'^2)^2+4Ks^2f_1^2f_2^2}\right] \tag{7}$$

where f_1, f_2: Resonant frequencies of principal vibration and subvibration when uncoupled

$$f_1'^2 = \left(1+\frac{1}{r_1}\cdot\frac{a}{1+a-h}\right)f_1^2 \qquad h = \frac{1+\mu}{Rf(rd+Rd)\omega^2 Cg\,Cd}$$

$$f_2'^2 = \left(1+\frac{1}{r_2}\cdot\frac{a}{1+a-h}\right)f_2^2 \qquad Km^2 = \frac{C_1C_2}{Cm^2}$$

$$Ks^2 = Km^2\left(1+\frac{b}{1+a-h}\right)^2$$

$$a = Co/C_L \qquad b = Cm/C_L \qquad r_1 = Co/C_1 \qquad r_2 = Co/C_2$$

Equation (7) gives two oscillation frequencies, namely, the oscillation frequency f_+ of the principal vibration for plus sign in front of the root and the oscillation frequency f_- of the subvibration for minus sign in front of the root. On the other hand, load capacitance C_L is given as

$$\frac{1}{C_L} = \frac{\frac{1}{Cg}+\frac{1}{Cd}\left(1+\frac{Ri+Rqs}{rd+Rd}\right)}{1-\frac{Co}{Cd}\cdot\frac{2Ri}{rd+Rd}} \tag{8}$$

As is apparent from equation (7), f_+ and f_- are given as a function of load capacitance C_L and from equation (8), since C_L is given as a function of Rd and Rqs, it is, therefore, predicted and understood that oscillation frequencies $f_\pm$ and their frequency temperature behavior vary with Rd and Rqs.

§ 3 Results and Discussion

Let us explain the results according to the following procedure. First, Rd and Rqs are found to suppress spurious vibrations. Second, a relationship of oscillation frequency deviation $\triangle f_+/f_+$ for principal vibration versus Rd and Rqs is shown. Third, a relationship of variation $\triangle\alpha_+$, $\triangle\beta_+$, $\triangle\gamma_+$ of the first, second and third order temperature coefficients α_+, β_+, γ_+ versus Rd and Rqs is also shown, particularly, $\triangle\alpha_+$ is found to vary markedly. As a result, a specification of frequency temperature behavior is shown for a quartz crystal resonator to be manufactured when Rd and Rqs are taken into account. Finally, it is shown that a high stable temperature compensated crystal oscillator is obtained with oscillation frequency of principal vibration finely adjusted to the nominal frequency, and excellent frequency temperature behavior without any external temperature compensation and the suppression of spurious response over a wide temperature range. Moreover, a CMOS quartz crystal oscillator circuit is so designed that h becomes approximately zero from characteristics of the CMOS inverter and the circuit construction, in addition, since Cg and Cd are connected in series when looking at the circuit from both terminals of the quartz crystal resonator, one replaces capacitance of in-series connection CgCd/(Cg+Cd) with C_{Lo}. Let us explain concretely below.

Table 1 shows the resonant frequency and the electrical equivalent circuit parameters of the principal vibration and subvibration for eleven new shape GT cut quartz crystal resonators used in the present experiments. As is obvious from Table 1, the resonators show the very satisfactory values as a resonator. In addition, capacitance ratio r_i is defined as $r_i = Co/C_i$ $(i=1,2)$. Further, C_{11} of uncoupled vibrations between principal vibration and subvibration is obtained from coupling capacitance Cm calculated by employing the coupling factor[7] derived from the frequency temperature behavior. A specification on frequency temperature behavior to be manufactured will be described later.

3.1 Spurious vibration

Figure 2 shows an example of frequency temperature

Table 1 Electrical equivalent circuit parameters of the new shape GT cut resonators used in the present experiments.

No.	Frequency (MHz)		$R_{1,2}$ (ohm)	$L_{1,2}$ (mH)	$C_{1,2}$ (fF)	C_0 (pF)	$r_{1,2}$	Q ($\times 10^3$)
1	Prin.	2.0997	51.0	990	5.80	2.44	420	256
	Sub.	1.8934	196	2786	2.54		961	169
2	Prin.	2.0996	65.1	984	5.84	2.45	420	199
	Sub.	1.8927	235	2530	2.79		878	128
3	Prin.	2.0996	56.4	962	5.97	2.46	412	225
	Sub.	1.8937	595	2767	2.55		965	55
4	Prin.	2.1035	75.2	967	5.92	2.43	410	170
	Sub.	1.8948	753	2831	2.49		976	45
5	Prin.	2.1040	65.8	973	5.88	2.48	422	196
	Sub.	1.8953	651	2857	2.48		1004	52
6	Prin.	2.0997	49.5	985	5.59	2.48	443	262
	Sub.	1.8928	222	2679	2.59		957	143
7	Prin.	2.0997	51.0	990	5.57	2.44	439	256
	Sub.	1.8934	196	2786	2.49		981	169
8	Prin.	2.0996	58.1	964	5.71	2.46	431	219
	Sub.	1.8937	595	2767	2.51		982	55
9	Prin.	2.0997	58.1	964	5.71	2.49	418	219
	Sub.	1.8936	1235	2585	2.68		931	25
10	Prin.	2.0996	80.3	985	5.83	2.47	424	162
	Sub.	1.8934	235	2651	2.67		925	134
11	Prin.	2.0996	73.4	972	5.91	2.45	414	175
	Sub.	1.8937	1250	2548	2.77		884	24

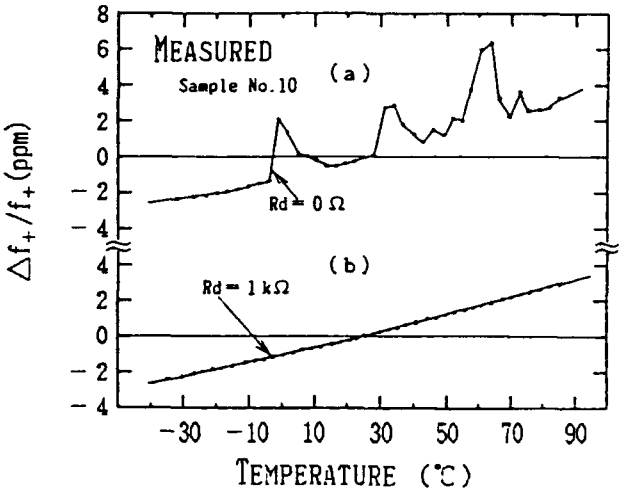


Fig. 2 Influence of spurious vibration(a) and its suppression by drain output resistance Rd (b).

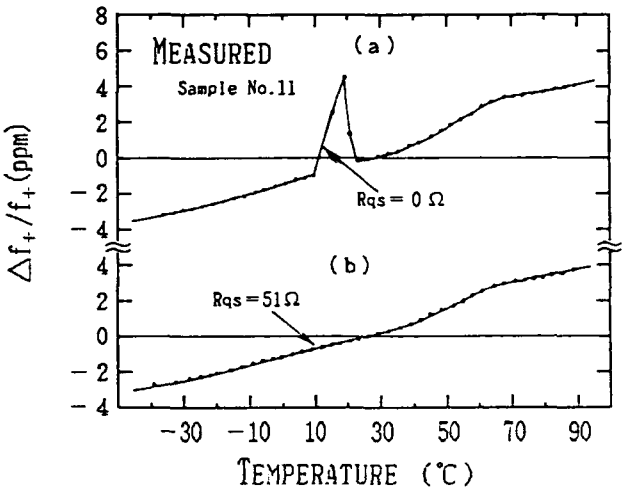


Fig. 3 Influence of spurious vibration (a) and its suppression by resistance Rqs connected in series to the resonator (b).

behavior. Fig. 2 (a) exhibits frequency temperature behavior in a CMOS quartz crystal oscillator circuit consisting of a resonator of sample No. 10, C_{LO}=15*pF*, R_f=10 *MΩ* and V_{DD} 5.0 (*V*). Frequency jump occurs in the vicinity of 0 ℃, 35 ℃ and 60 ℃ in temperature, because the resonator is affected by spurious vibration, while Fig. 2 (b) exhibits frequency temperature behavior when drain output resistance Rd with a value of 1*kΩ* is further connected in the circuit construction of Fig. 2 (a). A stable frequency versus temperature is obtained without the occurrence of spurious vibration in the temperature range of −30 ℃ to +85 ℃. Thus, Rd can suppress spurious vibration.

Figure 3 shows another example of frequency temperature behavior. Fig. 3 (a) exhibits frequency temperature behavior in a CMOS quartz crystal oscillator circuit consisting of a resonator of sample No. 11, C_{LO}=14.8*pF*, R_f =10*MΩ* and V_{DD}=5.0 (*V*). Frequency jump occurs in the vicinity of 15℃ in temperature, it is easily understood from the above result that the resonator is influenced by spurious vibration, while Fig. 3 (b) exhibits frequency temperature behavior when Rqs with a value of 51 *Ω* is further connected in the circuit construction of Fig. 3 (a). As well as the connection of Rd, a stable frequency versus temperature is obtained with the suppression of spurious vibration. Thus, it is sufficiently understood that Rd and Rqs have the function of suppressing spurious vibration.

3.2 Oscillation frequency

Figure 4 shows the relationship between oscillation frequency deviation $\Delta f_+/f_+$ and drain output resistance Rd in a CMOS oscillator circuit with C_{LO} 15*pF* and the

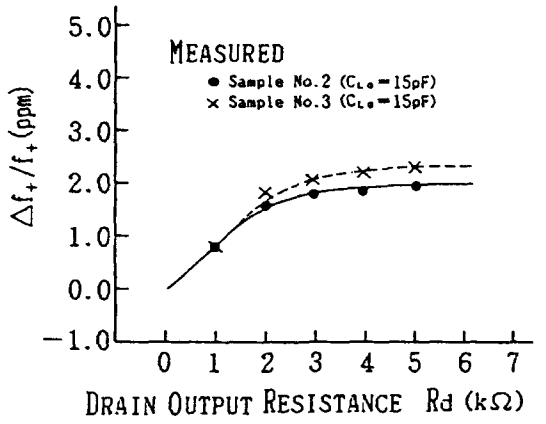


Fig. 4 Oscillation frequency deviation $\Delta f_+/f_+$ for principal vibration versus drain output resistance Rd.

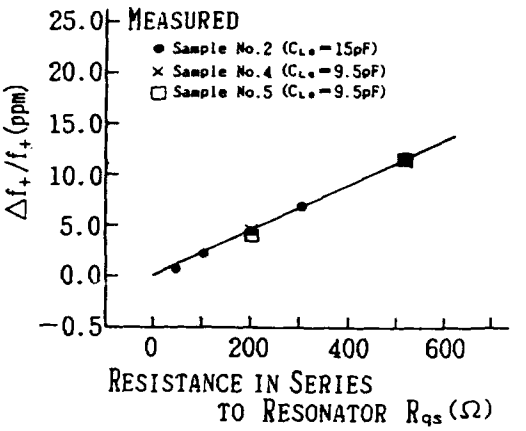


Fig. 5 Oscillation frequency deviation $\Delta f_+/f_+$ for principal vibration versus resistance Rqs connected in series to resonator.

resonator of samples No. 2 or No. 3. According to increase of Rd, the oscillation frequency becomes somewhat high, but, frequency deviation $\Delta f_+/f_+$ is approximately $2ppm$ even if Rd varies from 0 to $5k\Omega$ and its variation is very small.

Figure 5 shows the relationship between oscillation frequency deviation $\Delta f_+/f_+$ of principal vibration and resistance Rqs connected in series to a resonator in a CMOS oscillator circuit comprising a new shape GT cut quartz crystal resonator. In the oscillator circuit, the resonators of samples No. 2, 4, 5 are used. C_{Lo}=$15pF$ for No. 2 and C_{Lo}=$9.5pF$ for No. 4, 5. As Rqs increases, the oscillation frequency becomes somewhat high, but, frequency deviation $\Delta f_+/f_+$ is approximately $4.5ppm$ independent on C_{Lo} value, even if Rqs varies from 0 to 200 Ω and as well as Rd, its variation is very small.

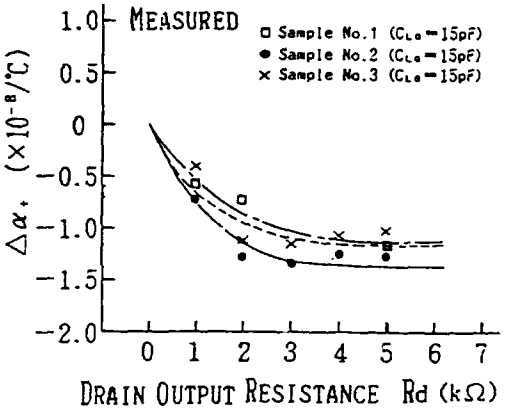


Fig. 6 Change of the first order temperature coefficient($\Delta\alpha_+$) versus drain output resistance Rd.

3.3 Frequency temperature coefficients

Figure 6 shows the relationship of variation $\Delta\alpha_+$ of the first order temperature coefficient α_+ for principal vibration versus drain output resistance Rd when a resonator of samples No. 1, 2, 3 is used in the same construction as the CMOS oscillator circuit of Fig. 4. As is apparent from Fig. 6, $\Delta\alpha_+$ decreases, as Rd increases. When Rd varies from 0 to $5k\Omega$, $\Delta\alpha_+$ has a negative value of approximately -1.0×10^{-8}/℃ to -1.5×10^{-8}/℃. This result shows that frequency deviation $\Delta f_+/f_+$ of 0.55 to $0.825ppm$ occurs in a temperature range of −30℃ to +85℃ in the clockwise direction. Thus, it is shown that Rd raises the oscillation frequency a little bit and changes the frequency temperature behavior. But, since Rd is employed in the vicinity of the first fixed value, the about third of the frequency change described above may be actually considered.

Figure 7 shows the relationship of variation $\Delta\alpha_+$ of the first order temperature coefficient α_+ for principal vibration versus resistance Rqs connected in series to a resonator, when the resonator of samples No. 2, 4 is

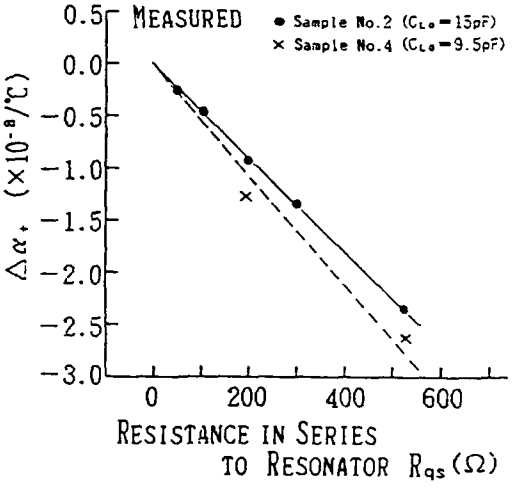


Fig. 7 Change of the first order temperature coefficient($\Delta\alpha_+$) versus resistance Rqs in series connected to quartz crystal resonator.

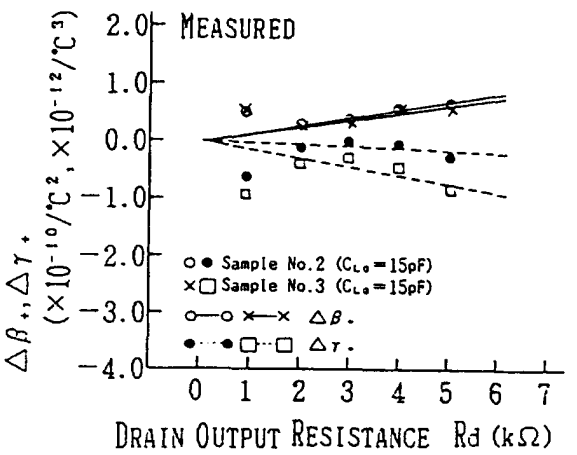


Fig. 8 Change of the second and third order temperature coefficient ($\Delta\beta_+$, $\Delta\gamma_+$) versus drain output resistance Rd.

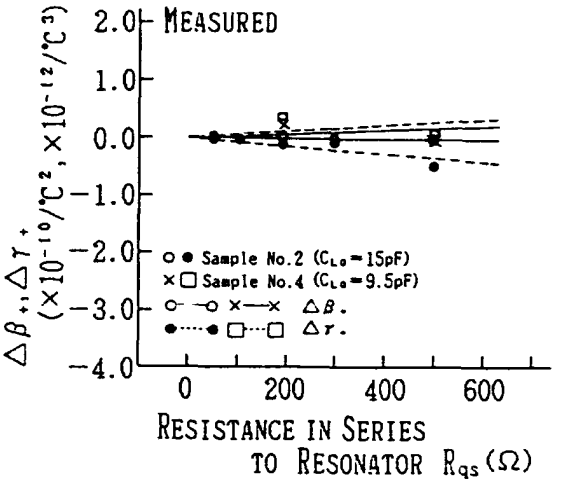


Fig. 9 Change of the second and third order temperature coefficient ($\Delta\beta_+$, $\Delta\gamma_+$) versus resistance Rqs in series connected to quartz crystsl resonator.

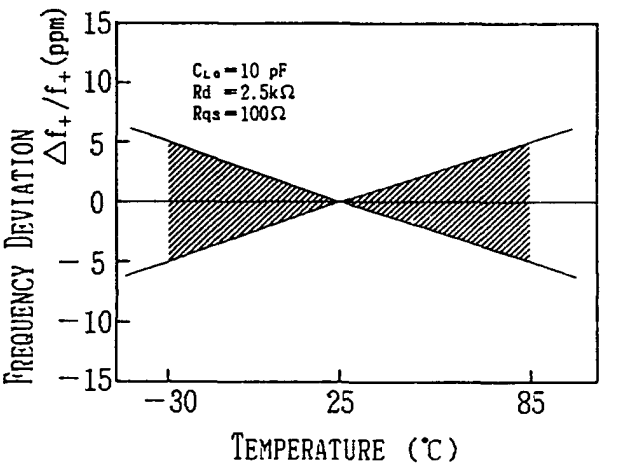


Fig. 10 Region of frequency temperature characteristics for new shape GT cut quartz crystal resonators to be manufactured.

employed in the same construction as the CMOS oscillator circuit of Fig. 5. When Rqs varies from 0 to 200Ω, $\Delta\alpha_+$ has a negative value of approximately -0.94×10^{-8}/℃ to -1.27×10^{-8}/℃ as well as Rd. In this case, frequency deviation $\Delta f_+/f_+$ of 0.52 to 0.70*ppm* takes place in the temperature range of −30℃ to +85℃ in the clockwise direction. Accordingly, frequency deviation of approximately 1.07 to 1.53*ppm* occurs in the above-mentioned temperature range by a change of both Rd and Rqs described previously. In addition, when resonators are manufactured in an oscillator circuit wherein a value of Rd and Rqs is properly determined, this result suggests that the oscillator circuit can be so designed that the first order temperature coefficient α_+ hardly varies with selection of a value of Rd and Rqs, even if they are changed.

Figure 8 shows the relationship of variation $\Delta\beta_+$, $\Delta\gamma_+$ of the second and third order temperature coefficients β_+, γ_+ for principal vibration versus drain output resistance Rd when a resonator of samples No. 2, 3 is employed in the same construction as the CMOS oscillator circuit of Fig. 4. As shown in Fig. 8, when Rd varies from 0 to 5kΩ, $\Delta\beta_+$ and $\Delta\gamma_+$ has a value of $+7.02\times10^{-11}$/℃² to $+6.67\times10^{-11}$/℃² and 2.5×10^{-13}/℃³ to 8.2×10^{-11}/℃³ respectively and their variation is extremely small. In addition, $\Delta\beta_+$ and $\Delta\gamma_+$ are found to be negligibly small, even if frequency deviation within +/− 2.5*ppm* is discussed in a temperature range of −30℃ to +85℃, because Rd is employed in the vicinity of the first fixed value as described in Fig. 6. Accordingly, as shown in Fig. 6, a change of the first order temperature coefficient α_+ must be considered only versus that of Rd.

Figure 9 shows the relationship between Rqs and variation $\Delta\beta_+$, $\Delta\gamma_+$ of the second and third order temperature coefficients β_+, γ_+ for principal vibration. When Rqs varies from 0 to 200Ω, $\Delta\beta_+$ and $\Delta\gamma_+$ has a value of $+2.09\times10^{-13}$/℃² to -2.2×10^{-11}/℃² and 1.04×10^{-13}/℃³ to $+3.74\times10^{-13}$/℃³ respectively and their variation is found to be negligibly small as well as Rd. In this case, a change of the first order temperature coefficient α_+ must be also considered.

3.4 Frequency temperature behavior

Figure 10 shows the region of frequency temperature behavior for new shape GT cut quartz crystal resonators which are manufactured by using a CMOS oscillator circuit comprising C_{Lo}=10*pF*, Rd=2.5kΩ and Rqs=100Ω. The region is drawn by oblique lines. The resonators are so manufactured that the frequency deviation for principal vibration has a value of −/+5*ppm* to +/−5*ppm*. Simultaneously, since it was reported[7] that frequency temperature behavior varies with gate load capacitance Cg and drain load capacitance Cd, this result is here employed. Therefore, when oscillation frequency for principal vibration is finely adjusted within +/−0.5*ppm* in

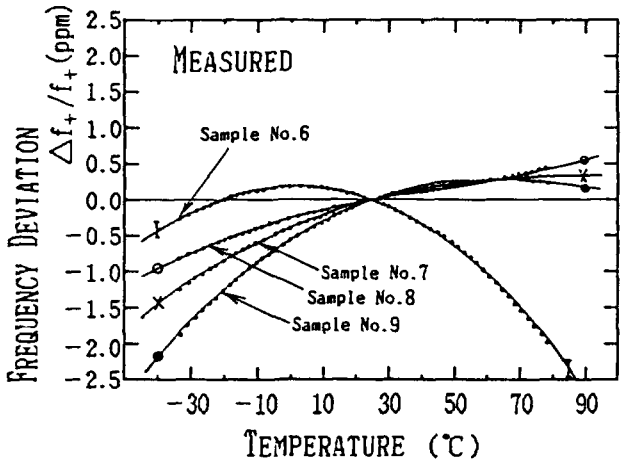


Fig. 11 Examples of frequency temperature behavior for the resonators adjusted by Rd, Rqs and Cg, Cd.

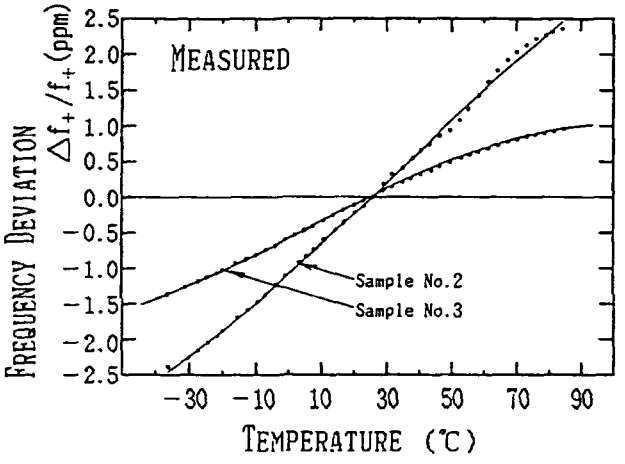


Fig. 12 Another examples of frequency temperature behavior.

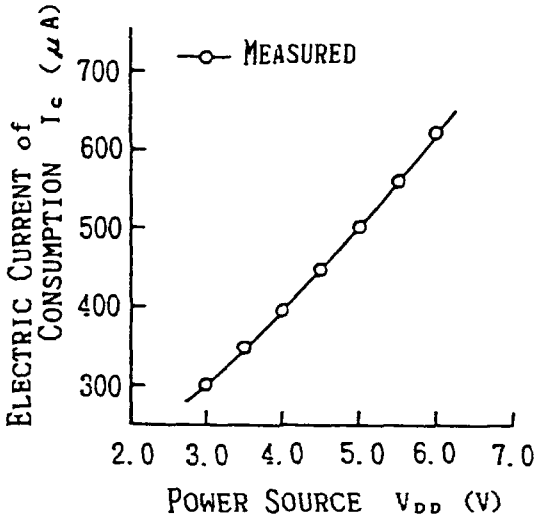


Fig. 13 Relationship between electric current of consumption I_c and power source V_{DD}.

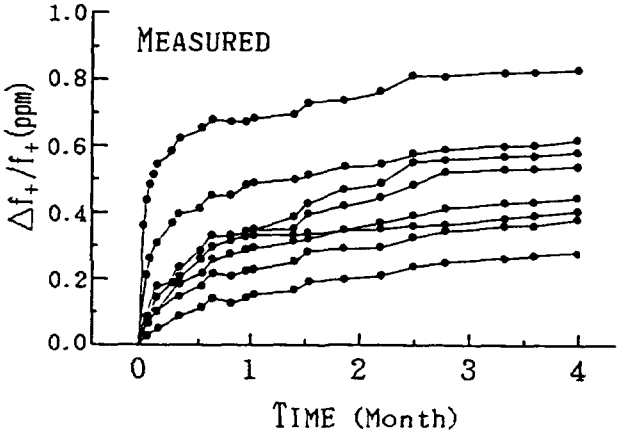


Fig. 14 Aging characteristics.

frequency deviation versus the nominal frequency by Cg and Cd. the resonators are so manufactured that frequency deviation of 1.5 to 2.5*ppm* in an absolute value occurs in a temperature range of −30℃ to +85℃. Needless to say, as is evident from the reference[7], resonators are so manufactured that oscillation frequency becomes lower for positive α, and higher for negative α, than the nominal frequency.

Figure 11 shows examples of the frequency temperature behavior obtained by selection of Rd, Rqs and Cg, Cd in the oscillator circuits comprising new shape GT cut quartz crystal resonators manufactured under specification of Fig. 10 and with electrical equivalent circuit parameters of Table 1. It is fully understood from Fig. 11 that stable frequency characteristics are obtained with the suppression of spurious vibrations, in addition, frequency deviation is within +/− 2.5*ppm* (actually, +0.48*ppm* to -2.45*ppm*) over a wide temperature range of −30℃ to +85℃ without any external temperature compensation, thus, extremely stable frequency versus temperature is obtained. In order to obtain more excellent frequency temperature behavior, the second order temperature coefficient β, of a resonator must be small. It is possible by selection of cut angle and this will be the coming subject.

Figure 12 shows another examples of frequency temperature behavior. It is sufficiently understood that temperature compensated crystal oscillators (TCXO・MCQ) are obtained with extremely small frequency change versus temperature without any external temperature compensation as well as Fig. 11.

3.5 Electric current consumption

Figure 13 shows the relationship between electric current of consumption I_c and power source V_{DD}. The electric current of consumption increases proportionally



Photo 1 TCXO・MGQ developed by us and compared with DIME.

to the power source V_{DD}. When V_{DD} 5.0(V), the electric current of consumption in the present oscillator circuit has a value of approximately 500 μA. This is a value of one half to one third of the conventional TCXO.

3.6 Aging characteristics

Figure 14 shows examples of aging characteristics although dependent on samples, frequency change is 0.26*ppm* to 0.81*ppm* in aging of about four months. Frequency change is probably less than 1*ppm* per the first year, judging from the aging of about four months.

3.7 Other

Photo 1 shows TCXO・MGQ developed by us and compared with DIME.

§ 4 Conclusions

In this paper, a study was performed with a view to clarifying if drain output resistance Rd and resistance Rqs connected in series to the resonator in elements constructing a CMOS oscillator circuit employing a new shape GT cut quartz crystal resonator, influence oscillation frequency and frequency temperature behavior and obtaining a temperature internally compensated crystal oscillator (TCXO・MGQ) with the suppression of spurious vibration. First, in analysis procedure, since a CMOS inverter, as is already known, can be linearly expressed by using a voltage source and drain resistance for a small signal, the present oscillator circuit was transformed into an equivalent circuit. Next, an equation of condition to give oscillation frequencies was theoretically derived from Kirchhoff's law. Furthermore, an imaginary part of impedance for the present quartz crystal resonator was calculated. From these results the equation of oscillation frequencies was derived. The obtained oscillation frequencies was found to be given as a function of Rd and Rqs. in addition, an increase of Rd and Rqs was found to raise the oscillation frequency for principal vibration a little bit and also influence the frequency temperature behavior. As a result, it was shown that a temperature compensated crystal oscillator which is called "TCXO・MGQ", is obtained with the suppression of spurious vibration, frequency deviation within +-2.5*ppm* (actually, +0.48*ppm* to 2.45*ppm*) in a temperature range of -30 ℃ to +85 ℃ without any external temperature compensation and the oscillation frequency for principal vibration finely adjusted within +-0.5*ppm* to the nominal frequency, by selection of a value for Rd, Rqs and Cg, Cd.

Finally, the coming subject will be to clarify the influence on oscillation frequencies and frequency temperature behavior due to another elements constructing a CMOS oscillator circuit.

References

[1] H. Kawashima: "A Variational Analysis of a GT Cut Quartz Crystal Resonator with the Supporting Portions at the Ends", Trans. IECE(A), J68-A, 8, pp.771-778 (August 1985)(in Japanese).

[2] H. Kawashima: "Variational analysis of GT cut quartz crystal resonators with the supporting portions at the ends", Proc. the 40th Annual Symp. on Frequency Control, pp.193-200(1986).

[3] H. Kawashima, H. Sato and O. Ochiai: "New frequency temperature characteristics of miniaturized GT cut quartz resonators", Proc. the 34th Annual Symp. on Frequency Control, pp.131-139(1980).

[4] H. Kawashima: "A vibration analysis of a GT cut quartz crystal resonator taking account of dissipation term", J. IEICE(A), J72-A, 4, pp.633-639 (April 1989) (in Japanese).

[5] H. Kawashima and M. Matsuyama: "An Analysis of Electrical Equivalent Circuit Parameters for a GT Cut Quartz Crystal Resonator with the Supporting Portions at Both Ends", J. IEICE(A), J71-A, 6, pp.1217-1224 (June 1988)(in Japanese).

[6] H. Kawashima: "An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New Shape GT Cut Quartz Crystal Resonator Formed by an Etching Method". IEEE Ultrasonics Symp., Proc. pp.465-470 (1989).

[7] H. Kawashima: "An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using Coupling Quartz Crystal Resonator", Proc. the 44th Annual Symp. on Frequency Control, pp.585-592(1990).

[8] H. Kawashima: "Influence on Frequency Temperature Behavior Due to the Supporting Portions for GT Cut Quartz Crystal Resonators Formed by an Etching Method", J. IEICE(A), J73-A, 6, pp.1031-1038(June 1990)(in Japanese).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

VOLTAGE CONTROLLED TEMPERATURE COMPENSATED CRYSTAL OSCILLATOR USING 2-PORT CRYSTAL RESONATOR

YOSHIFUSA UENO
NIHON DEMPA KOGYO CO., LTD.
SHINWA BLDG. 3-2-4, NISHI-SHINJUKU
SHINJUKU-KU, TOKYO, JAPAN 160

HIROSHI SHIMIZU
UNIVERSITY OF ELECTROCOMMUNICATIONS
CHOFU, TOKYO, JAPAN 182

ABSTRACT

Generally, the frequency versus temperature characteristic of voltage controlled temperature compensated crystal oscillators(VCTCXO) varies when their oscillation frequency is shifted. This problem is well known in the art, and difficulty has been experienced in solving the problem insofar as the conventional crystal resonator is used.

In an effort to solve the problem, our research has led to the development of a VCTCXO using a 2-port crystal resonator with a special electrode structure.

Typical example of a 2-port crystal resonator is an MCF(Monolithic Crystal Filter). The crystal resonator of an MCF is however not suitable for use as a crystal resonator for an oscillator because it has two resonant responses (symmetric and anti-symmetric modes). The developed 2-port crystal resonator achieves a single resonant response with an electrode structure for suppressing the anti-symmetric mode.

The VCTCXO with the 2-port crystal resonator can have two load capacitances, i.e., a load capacitance for temperature compensation and a load capacitance for shifting the oscillation frequency. Therefore, the VCTCXO can compensate for temperature changes and shift the oscillation frequency independently.

In our research, a new electric equivalent circuit for the 2-port crystal resonator is introduced, and fundamental equations are derived from the equivalent circuit for determining the extent to which the oscillation frequency varies with the load capacitances, with comparison made between measured and calculated values.

The article also refers to the frequency versus temperature characteristic of VCTCXO and DTCXO using the 2-port crystal resonator.

1. INTRODUCTION

In a temperature compensated crystal oscillator(TCXO), the load capacitance of the crystal oscillator is varied depending on its frequency versus temperature characteristic so as to compensate for changes in the oscillation frequency and minimize the frequency deviation. Figure 1(a) shows a typical TCXO circuit in which a voltage generated by a temperature-compensation network is applied to a variable-capacitance diode D, and the temperature-dependent characteristic of the oscillation frequency is corrected as flat as possible by the change in the capacitance of the diode D. Figure 1(b) illustrates the frequency versus temperature characteristic of a crystal oscillator which is not temperature compensated and that of a crystal oscillator which is temperature compensated.

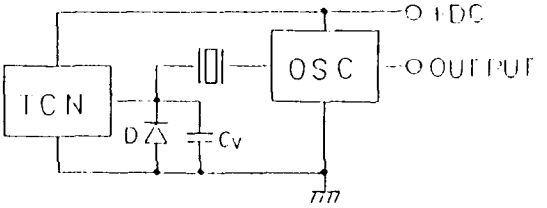


Figure 1(a) : Typical circuit of TCXO

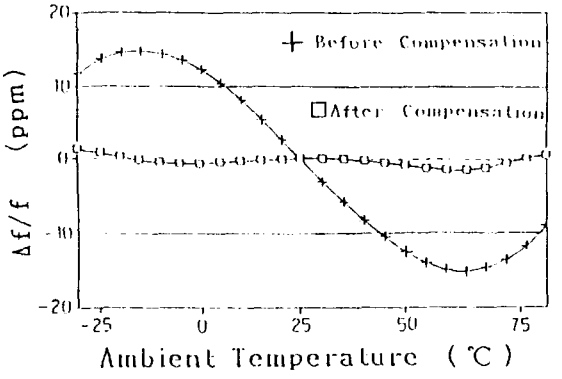


Figure 1(b) : Frequency versus temperature characteristics before and after temperature compensation

The fractional change of the oscillation frequency of the crystal resonator is expressed by the following equation (1):

$$\frac{\Delta f}{f_1} = \frac{1}{2r\left(1+\frac{C_L}{C_0}\right)} \qquad (1)$$

CH2965-2/91/0000-418 $1.00 © 1991 IEEE

where r is the capacitance ratio, C_0 the shunt capacitance , C_L the load capacitance , f the oscillation frequency, f_1 the series resonance frequency, and $\Delta f=f-f_1$. The extent to which the oscillation frequency of the crystal oscillator can vary is approximately calculated according to the equation (1).

Since the capacitance ratio r and the shunt capacitance C_0 are fixed by the crystal resonator used, the oscillation frequency is varied by only the load capacitance C_L. Inasmuch as the frequency difference Δf and the load capacitance C_L are of hyperbolic nonlinear relationship to each other, as can be seen from the equation(1), the rate of change of Δf caused by the load capacitance C_L varies depending on the value of C_L. With the conventional crystal oscillators, the temperature compensation of frequencies, the shifting of the oscillation frequencies, and the aging compensation of frequencies are all effected by varying the load capacitance C_L. Therefore, when the load capacitance C_L is varied in order to compensate for an aging-dependent change in the frequency, for example, the temperature compensation of the frequency which has previously been carried out is spoiled. Should this happen, it may be possible for the compensated frequency to shift out of the allowable error range of crystal oscillators having higher accuracy requirements. As the allowable oscillation frequency error range will get more tight in the future, this problem will require more attention but it will be difficult to be soluved by the conventional TCXOs.

One example for TCXO in terms of the problem mentioned above be described below.

It is assumed that an AT-CUT crystal resonator is used with the capacitance ratio r=250 and the shunt capacitance C_0=3pF, and the load capacitance C_L is varied to correct the frequency versus temperature characteristic of a crystal oscillator which has a maximum frequency deviation of 15ppm. (The crystal oscillator has a circuit arrangement as shown in Figure 1(a), and $C_L=C_{(diode)}+C_V$.) Figure 2 shows frequency versus temperature characteristic curves before and after temperature compensation according to the equation(1) with C_L=20pF. The frequency versus temperature characteristic with C_L=20pF (constant) is indicated by the mark + in Figure 2, and the frequency versus temperature characteristic that is compensated by varying C_L at each temperature is indicated by the mark □ in Figure 2. Furthermore, C_L is varied to a certain extent to shift the oscillation frequency by +10ppm and -10ppm, with the resultant frequency versus temperature characteristic curves being indicated by the marks + and ×, respectively, in Figure 3. Study of Figure 3 indicates that the crystal oscillator, though it has been temperature compensated to reduce the frequency error almost to zero at the central frequency f_0, produces a frequency error of about ±2.5ppm when oscillation frequency is shifted by ±10ppm.

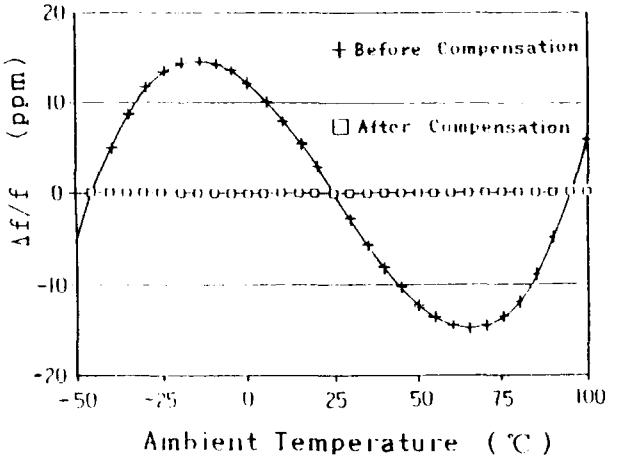


Figure 2 : Frequency versus temperature characteristics before and after optimum temperature compensation

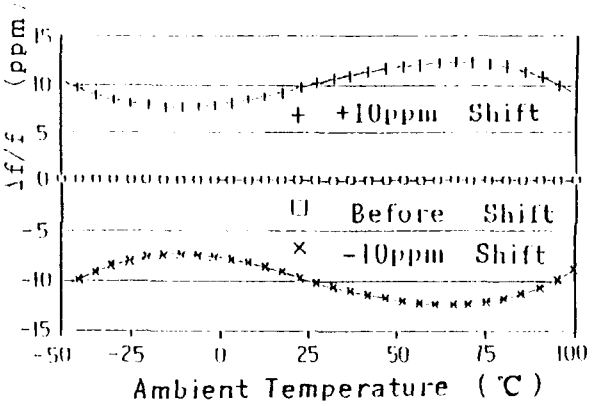


Figure 3 : Frequency versus temperature characteristics before and after shifting the frequency of an optimally temperature compensated oscillator by ±10ppm.

We have developed a 2-port crystal resonator with a view to solving the above problem. The conventional MCF is not suitable for use as a crystal resonator for an oscillator, since the conventional MCF substantialω has two resonant responses of almost equal magnitudes, as shown in Figure 4. (f_1 indicates a resonant frequency in a symmtric mode and f_2 a resonant frequency in an anti-symmetric mode) The developed 2-port crystal resonator has an electrode structure, as shown in Figure 5, in which the anti-symmetric mode is suppressed in order to provide a single resonant response for an oscillator. Figure 6 shows the resonant response.

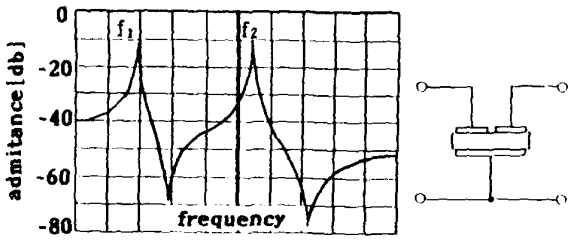


Figure 4 : The resonant responses

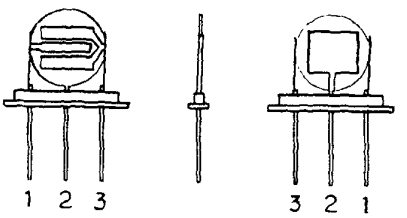


Figure 5 : The developed 2-port crystal resonator with an electrode structure for suppressing the anti-symmetric mode.

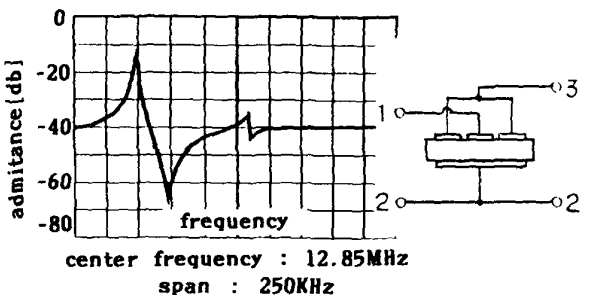


Figure 6 : The resonant response of the developed 2-port crystal resonator

Figure 7 shows an oscillator which incorporates the 2-port crystal resonator developed in our research. Figure 8 shows the dependence C_L with C_V as the parameter. Figure 10 in comparison with Figure 8, shows results obtained using an oscillator as shown in Figure 9 which incorporates a conventional crystal resonator.

The crystal oscillator which employs the developed 2-port crystal resonator has two load capacitances that can be varied independently of each other to vary the oscillation frequency.

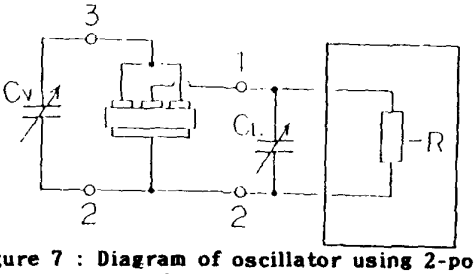


Figure 7 : Diagram of oscillator using 2-port crystal resonator developed in our research

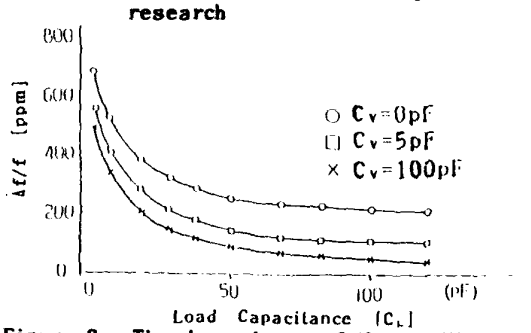


Figure 8 : The dependence of the oscillation frequency on the load capacitance C_L with C_V as the parameter, in case of an oscillator shown in Figure 7

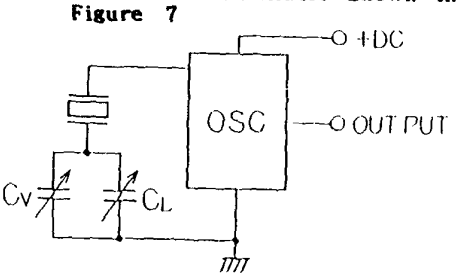


Figure 9 : Diagram of oscillator using a conventional crystal unit.

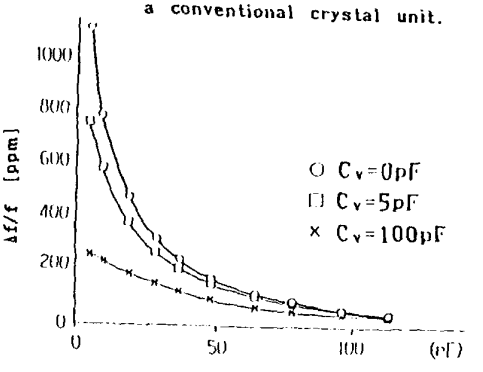


Figure 10 : The dependence of the oscillation frequency on the load capacitance C_L with C_V as the parameter, in case of an oscillator shown in Figure 9

2. ANALYSIS OF OSCILLATOR USING 2-PORT CRYSTAL RESONATOR

Since the developed 2-port crystal resonator has a single mode, its equivalent circuit may be represented by Figure 11.

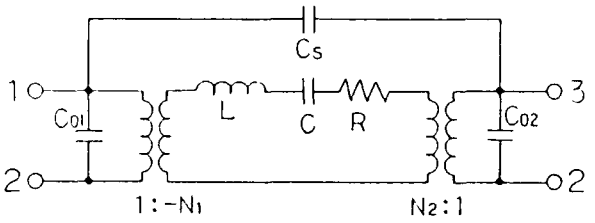


Figure 11 : Equivalent circuit of the developed 2-port crystal resonator

In Figure 11, the various reference characters are defined as follows:

- N_1, N_2 : force factors of respective terminal pairs.
- L : equivalent mass of the crystal resonator.
- C : equivalent compliance of the crystal resonator.
- R : equivalent mechanical resistance of the crystal resonator.
- C_s : stray capacitance between terminals 1 and 3.
- C_{01} : shunt capacitance between terminals 1 and 2.
- C_{02} : shunt capacitance between terminals 3 and 2.

and

$\omega_1^2 = 1/CL$

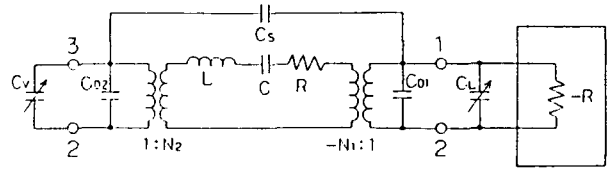


Figure 12 : Equivalent circuit of the oscillator using 2-port crystal resonator.

The impedance Z viewed from the terminals 1 and 2 in Figure 11 is given by the following equation(2):

$$Z = \frac{1}{j\omega} \cdot \frac{C_{02}+C_s}{C_{01}C_{02}+C_{02}C_s+C_sC_{01}} \cdot X \frac{\frac{\omega^2}{\omega_1^2}-1-\frac{N_2^2C}{C_{02}+C_s}}{\frac{\omega^2}{\omega_1^2}-1-\frac{N_2^2C_{01}+N_1^2C_{02}+(N_1+N_2)^2C_s}{C_{01}C_{02}+C_{02}C_s+C_sC_{01}}C_1} \quad (2)$$

An equivalent circuit of the oscillator using the 2-port crystal resonator is shown in Figure 12. The frequency condition of self oscillation (imaginary part=0) is given by

$$Z + \frac{1}{j\omega C_L} = 0$$

provided that C_{02} is replaced with $C_{02}+C_v$. From this relation the following equation(3) that gives the oscillation frequency $f(=\omega/2\pi)$ is obtained through approximation:

$$\frac{\omega}{\omega_1} = 1 + \frac{N_1^2C}{2(C_L+C_{01}+C_s)} + \frac{N_2^2C}{2(C_v+C_{02}+C_s)} \quad (3)$$

by putting $f=f_1+\Delta f$, $\omega_1=2\pi f_1$, $N_1^2C=C_1$, $N_2^2C=C_2$, $r_1=C_{01}/C_1$, and $r_2=C_{02}/C_2$, the equation(3) is rewritten as follows:

$$\frac{\Delta f}{f_1} = \frac{1}{2r_1\left(1+\frac{C_L+C_s}{C_{01}}\right)} + \frac{1}{2r_2\left(1+\frac{C_v+C_s}{C_{02}}\right)} \quad (4)$$

The equation (4) indicates that the oscillation frequency can be varied independently when the capacitances C_L, C_v are varied.

3. MEASURED AND CALCULATED VALUES OF CHANGES IN THE OSCILLATION FREQUENCY CAUSED BY THE LOAD CAPACITANCES

The equivalent constants of a 2-port crystal resonator used in the experiment are as follows: $r_1=357$, $r_2=1697$, $C_{01}=4.50$pF, and $C_{02}=1.55$pF. The relationship between the oscillation frequency and the load capacitance has been measured of the circuit arrangement shown in Figure 7. In Figure 13, the relationship between the frequency and the load capacitance C_L is plotted with C_v as a parameter. Figure 14 shows calculated values of the equation(4) which correspond to the measured values in the experiment. In the calculations, the stray capacitance C_s is regarded as zero since it is very small. The results shown in Figure 13 and 14 are slightly different from each other probably due to the stray capacitance developed by lead wires. Figure 15 shows values of the equation(4) calculated assuming that the stray capacitance of 1pF for each of C_L and C_v. The calculated values shown in Figure 15 well match those in Figure 13.

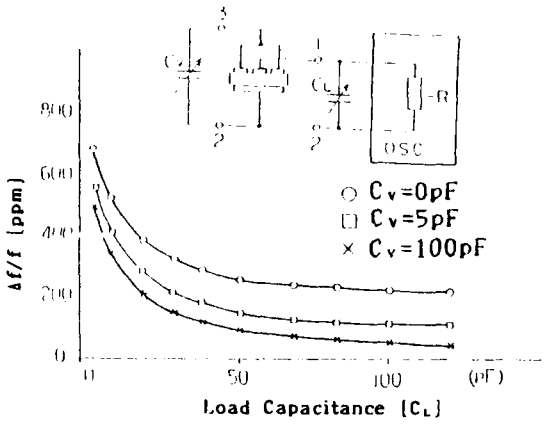


Figure 13 : Measured values of oscillation frequency as a function of load capacitance C_L, C_V being a parameter

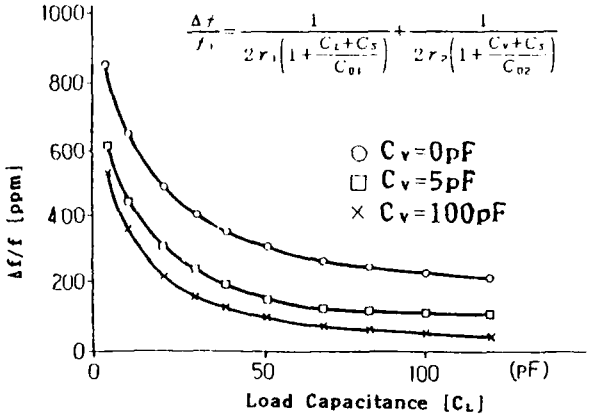


Figure 14 : Calculated values of oscillation frequency as a function of load capacitance C_L, C_V being a parameter

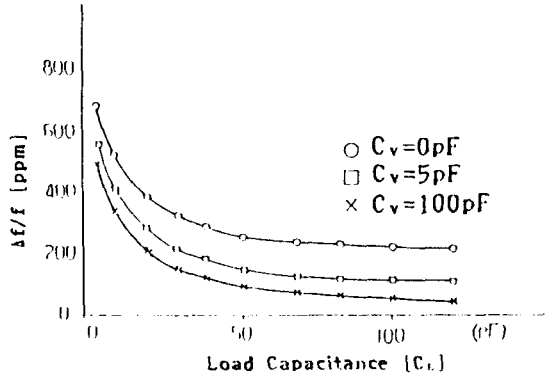


Figure 15 : Fractional frequency variation calculated compensating the influence of stray capacitances between lead wires, that is, taking C_L+1pF, C_V+1pF in place of C_L and C_V, respectively.

4. EQUIVALENT CONSTANTS OF 2-PORT CRYSTAL RESONATOR

In the experiment, AT-CUT crystal resonators were fabricated, whose dimension ratios between the central and side electrodes were different form each other. Specific electrode structures are shown in Figure 16 at (A), (B), (C), and (D). The respective equivalent constants were measured. Table 1 shows the results of the measurement. Capacitance ratios r_1 and r_2 of respective terminal pairs, and various constants are as follows:

$r_1=C_{01}/N_1^2C$, $r_2=C_{02}/N_2^2C$, $C_1=N_1^2C$, $C_2=N_2^2/C$, $L_1=L/N_1^2$, $L_2=L/N_2^2$.

These constants were measured accoding to the following three methods, and determined by calculations:

(i) The terminals 3 and 2 are short-circuited, measurements are made from the port 1-2, and L_1, C_1, R_1, $C_{01}+C_S$, are determined as equivalent circuit constants of an ordinary one-port crystal resonator.

(ii) The terminals 1 and 2 are short-circuited, measurements are made from the port 3-2, and L_2, C_2, R_2, $C_{02}+C_S$, are determined.

(iii) The terminals 1 and 3 are short-circuited, and the shunt capacitance is measured from the port 1-2, thus $C_0(=C_{01}+C_{02})$ is determined.

The equivalent constants of a 2-port crystal resonator depend on the electrode dimensions. The values of r_1 and r_2 are important factors for a crystal resonator for VCTCXOs because these dominate to which the frequency is variable. The area ratio between the central and side electrodes is minimum for the electrode structure(A), and maximum for the electrode structure(D). It can be seen from table 1 that the greater the area ratio, the smaller the capacitance ratio r_1, and greater the capacitance ratio r_2. The electrode structure(D) may be suitable if the extent to which the frequency is variable by C_V may be small. The electrode structure(A) may be suitable if the extent to which the frequency is variable by C_V is to be increased. The electrode structure (B) or (C) may be suitable if the extent to which the frequency is variable by C_V is to be equalized to the extent to which the frequency is variable by C_L.

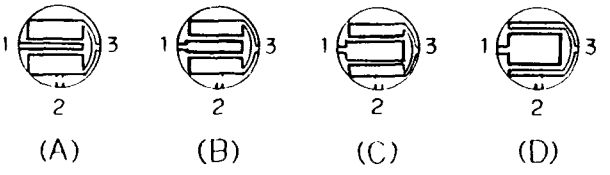


Figure 16 : Electrode structures of 2-port crystal resonator.

Table 1 : Equivalent constants of 2-port crystal resonator

ELECTRODE	Electrode A	Electrode B	Electrode C	Electrode D
L_1(mH)	168.5	33.39	15.3	10.26
C_1(fF)	0.915	4.62	10.00	15.04
R_1(Ω)	557.5	96.1	36.6	36.9
L_2(mH)	12.22	28.05	82.4	448.7
C_2(fF)	12.6	5.49	1.87	0.344
R_2(Ω)	42.3	81.3	198	2005
$C_{o1}+C_s$	1.61(pF)	2.41	3.30	4.07
$C_{o2}+C_s$	4.57(pF)	3.76	2.91	2.02
$C_{o1}+C_{o2}$	5.68(pF)	5.72	5.73	5.68
N_2^2/N_1^2	13.7	1.19	0.186	0.00229
C_s(pF)	0.25	0.23	0.24	0.21
C_{o1}(pF)	1.36	2.19	3.06	3.87
C_{o2}(pF)	4.32	3.54	2.67	1.82
r_1	1486	473	306	257
r_2	343	644	1428	5286

5. FREQUENCY VERSUS TEMPERATURE CHARACTERISTIC OF VCTCXOs USING 2-PORT CRYSTAL RESONATOR AND CONVENTIONAL CRYSTAL RESONATOR

Figure 17 shows how the frequency versus temperature characteristic varies when the frequency is shifted by ±10ppm in a conventional VCTCXO, as shown in Figure 18. Figure 19 shows how the frequency versus temperature characteristic varies when the frequency is shifted by ±10ppm in a VCTCXO using a 2-port crystal resonator, as shown in frequency 20.

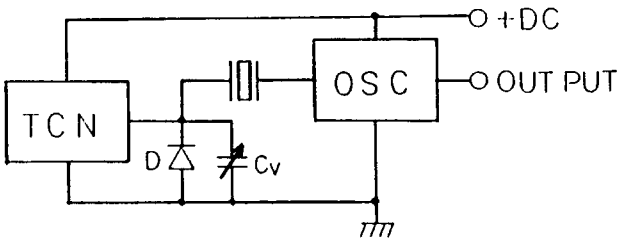


Figure 18 : Measurement diagram for a conventional VCTCXO.

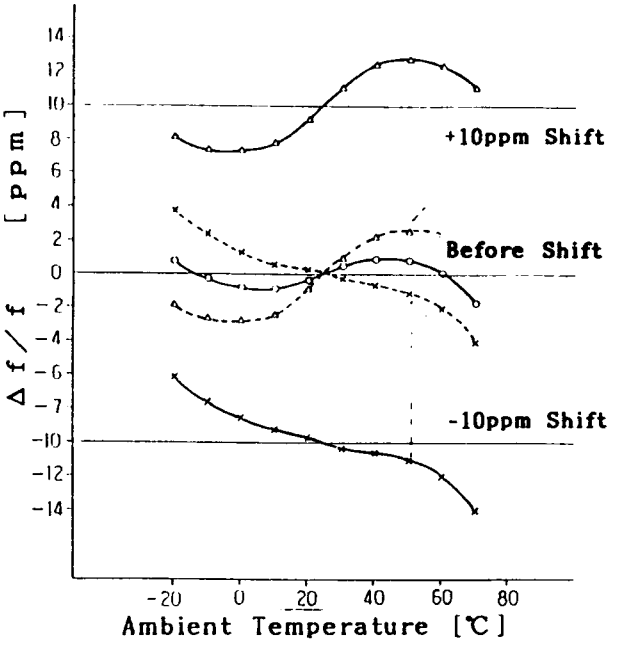


Figure 17 : Variation of the frequency vs. temperature characteristic when the frequency is shifted by ±10ppm in a conventional VCTCXO.

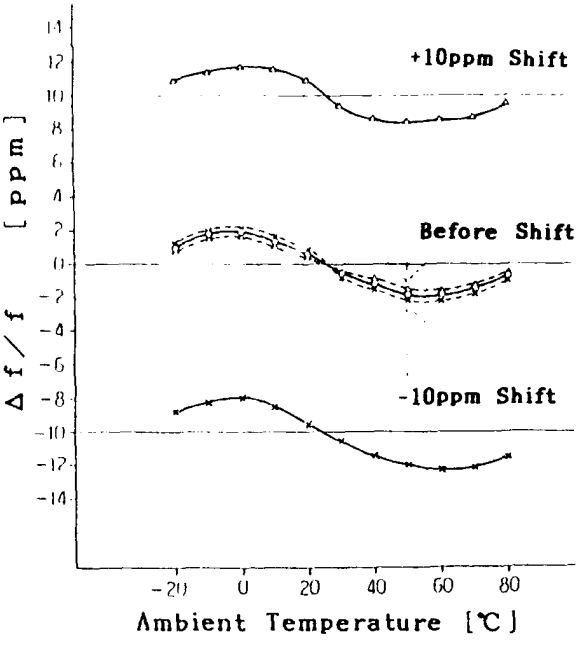


Figure 19 : The frequency vs.temperature characteristic varies when the frequency is shifted ±10ppm in a VCTCXO.using a 2-port crystal resonator.

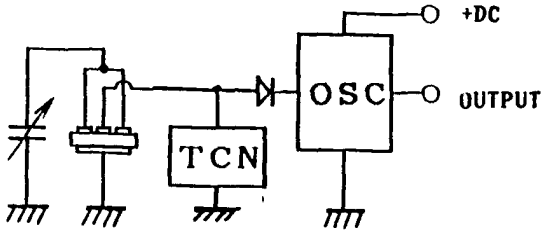


Figure 20 : Measurement diagram for a VCTCXO using a 2-port crystal resonator.

6. FREQUENCY VERSUS TEMPERATURE CHARACTERISTIC OF D-TCXOs USING 2-PORT CRYSTAL RESONATOR AND CONVENTIONAL CRYSTAL RESONATOR

Figure 21 shows how the frequency versus temperature characteristic varies when the frequency is shifted by ±10ppm in a conventional D-TCXO, as shown in Figure 22. Figure 23 shows how the frequency versus temperature characteristic varies when the frequency is shifted by ±10ppm in a D-TCXO using a 2-port crystal resonator, whose diagram is shown in Figure 24.

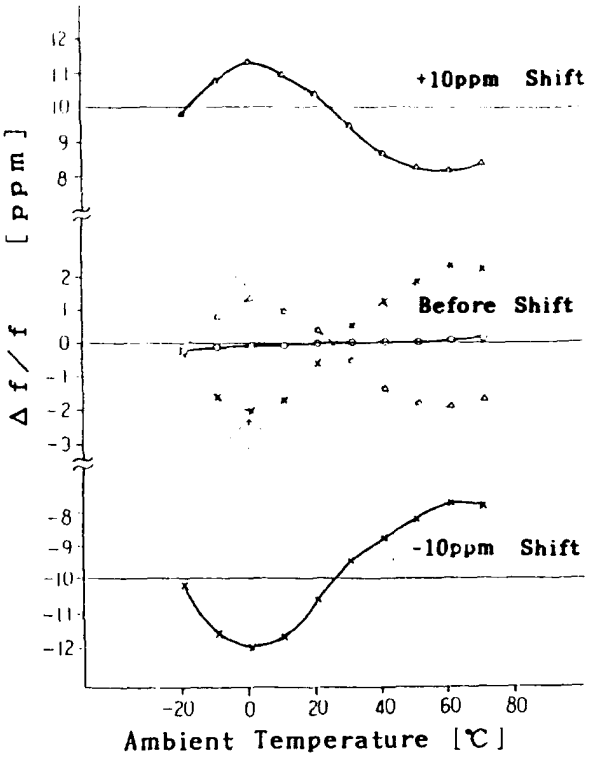


Figure 21 : The frequency vs.temperature characteristic varies when the frequency is shifted ±10ppm in a conventional D-TCXO.

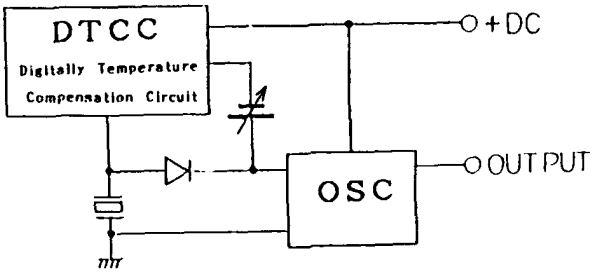


Figure 22 : Measurement diagram for a conventional D-TCXO.

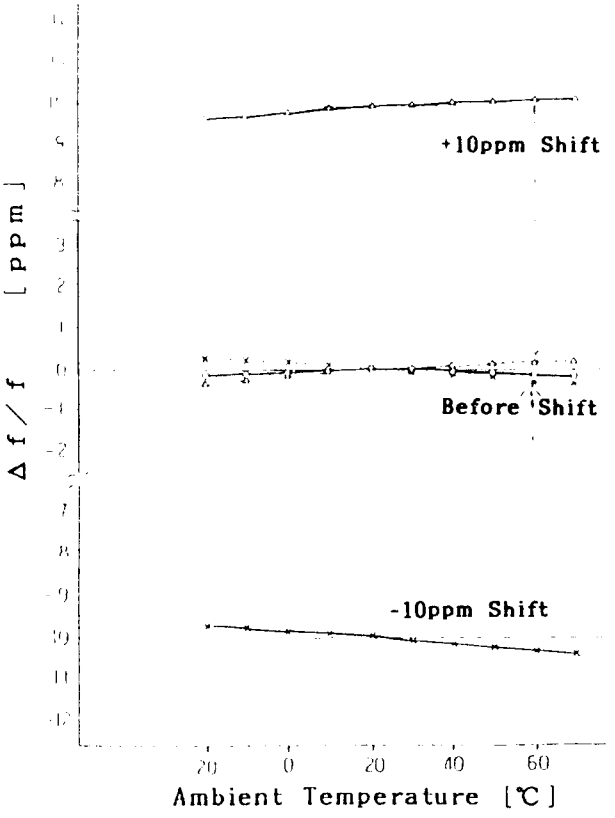


Figure 23 : Variation of the frequency vs. temperature characteristic when the frequency is shifted by ±10ppm in a D-TCXO using a 2-port crystal resonator.

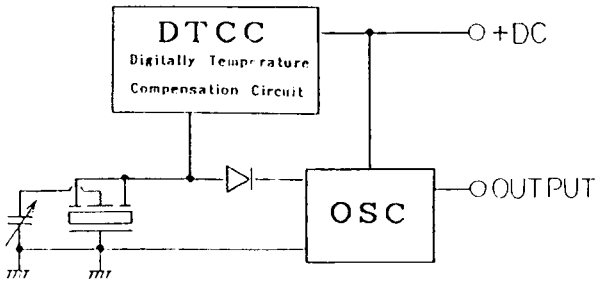


Figure 24 : Measurement diagram for a D-TCXO using a 2-port crystal resonator.

7. CONCLUSIONS

The problem of the conventional TCXO that the frequency versus temperature characteristic varies when the frequency is shifted has been considered to be a very tough problem to solve insofar as the crystal resonator is used. Our research gives one solution by using the 2-port crystal resonator.

The equation(4) for estimating the oscillation frequency depending on the resonator constants and the load capacitances has been derived from the oscillation frequency condition by using the equivalent circuit shown in Figure 11. The oscillation frequencies calculated according to the equation(4) have well matched the measured values by taking the stray capacitances due to lead wires into consideration. Therefore, the equation(4) can be said to be a guiding principle in the design of a VCTCXO using the 2-port crystal resonator.

The equation(4) is derived also taking account of the stray capacitance C_s between the divided electrodes. However, the value of C_s is about 0.2pF in actual crystal resonators, and much smaller than the values of C_{o1}, C_{o2}, C_L and C_v. Thus, the C_s is generally negligible.

ACKNOWLEDGEMENT

We are greatful to Mr.T.Tanzawa, Mr.H. Ishii, Mr.K.Kubo, and Mr.T.Yamakawa of NDK in assisting us in our research.

Special acknowledgement is extended to Mr.M.Takeuchi, president of NDK, Mr.T.Takeuchi, Exective Vice president of NDK, and Mr.H. Takeuchi, Director of NDK for their giving us environments useful for our research and for their helpful advices and suggestions.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

BVA-TYPE QUARTZ OSCILLATOR FOR SPACECRAFT

Jerry R. Norton

The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723

Abstract

The outstanding performance of an oscillator using a BVA-type stress-compensated cut resonator is presented. This oscillator design has demonstrated 24-h aging rates of <2E-12, 10-s Allan variance of 7E-14, and phase noise of <-130 dBc/Hz at 1 Hz from the carrier. BVA-type resonators operating at both 5 MHz and 10 MHz have been evaluated. In most performance categories, oscillators using BVA resonators perform better than those using conventional resonators. Abrupt frequency changes were observed in some BVA resonators; however, no pattern to these frequency changes, either in magnitude or shape, was observed. The oscillator was designed to survive the rigorous environment encountered during launch and operation in a space environment. A test program qualifying the oscillator for space flight has been successfully completed. The environmental testing included radiation, thermal vacuum, magnetics, shock, acceleration, and vibration.

Introduction

Detailed performance data from a newly developed high-performance oscillator that uses a BVA quartz resonator as its frequency control element is presented. This new oscillator design is the most recent in a series of oscillators developed at The Johns Hopkins University Applied Physics Laboratory (JHU/APL), dating back to the late 1950s. In each successive design, oscillator performance was improved, usually by taking advantage of improvements in quartz resonator technology. A description of the electrical and mechanical design is given.

The BVA Resonator

A BVA-design 5-MHz, 3rd overtone, SC (stress compensated) cut quartz resonator determines the output frequency and the frequency stability of this oscillator. The unique feature of the BVA design is the removal of the deposited metal electrodes from the critical vibrating area of the quartz resonator [1]. This reduces the chance of trapping contaminating material between the quartz and the electrodes and reduces stress created by depositing dissimilar material on the quartz. The net result of the BVA design is a resonator with improved frequency stability and reduced radiation-induced frequency changes.

Performance capability of this oscillator design using BVA resonators is outstanding with selected resonators. However, variations exceeding factors of 10 in resonator performance occurred while screening resonators in test oscillators. Some resonators were tested in different oscillators, and the performance for a given resonator was similar in each oscillator. This finding indicates that resonator quality dominates oscillator performance.

Oscillator Performance

Frequency Stability and Phase Noise Results

Allan variance data for the 5-MHz oscillator are shown in Figure 1. The Allan variance data for 10 s, 100 s, and 1000 s are all <1E-13. The 1000-s Allan variance is dependent on the aging rate of the oscillator. There is little difference between the 100-s and the 1000-s Allan variance data for oscillators with aging rates in the low parts in E-11. The frequency stabilities presented are not isolated cases that cannot be repeated. Allan variance data collected continuously will have periods of 8 to 10 h without a single data set exceeding 1E-13 (Fig. 2). Figure 3 presents a strip chart showing frequency measurements made at a 100-s sample rate for two oscillators. Oscillator S/N 01 has an Allan variance of 6.9E-14, oscillator S/N02 an Allan variance of 2.1E-13. The difference in the numerical data is <3 to 1, but there is a striking difference in the strip chart recordings of the two oscillators. Oscillator S/N01 has a smooth, consistent character with fewer random perturbations, whereas oscillator S/N02 is much noisier. The Allan variance data presented in this paper does NOT assume equal noise sources (not reduced by 3 dB).

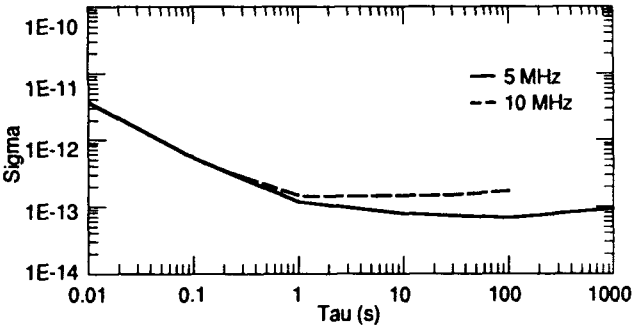


Figure 1. Allan variance of BVA oscillators. Measurement bandwidth = 1000 Hz. Measured data does not assume equal noise sources.

Phase noise ($\mathcal{L}_f$) performance of these oscillators is also very good, particularly near the carrier. The phase noise 1 Hz from the carrier is -131 dBc/Hz descending to a noise floor of -160 dBc/Hz at 100 kHz (Fig. 4).

Twenty-four hour aging rates for the oscillators varied between 2E-12 and 4E-11. The oscillators perform better in a vacuum than in atmospheric conditions, particularly in relation to elapsed time, i.e., aging rate and 1000-s Allan variance.

Environmental Effects

Figure 5 presents oscillator performance as a function of temperature. The total frequency change is -1.1E-11 for a temperature change of 80°C, with the maximum temperature coefficient of -4.3E-13/°C occurring between 37°C and 50°C. The temperature coefficient between 15°C and 30 °C, which is the nominal operating range, is 1.9E-13/°C.

The maximum accumulated radiation dose encountered by these oscillators during any given orbit is 0.6 rd (Si). A radiation test that simulates this

CH2965-2/91/0000-426 $1.00 © 1991 IEEE

```
 93.      3.009          42.27999
SIGMA=                 7.566E-014
RECORD #                    14
101.     3.142          42.27992
SIGMA=                 4.764E-014
SIGMA=                 6.756E-014
 29.     4.009          42.28001
SIGMA=                 5.599E-014
SIGMA=                 7.396E-014
SIGMA=                 7.796E-014
SIGMA=                 7.300E-014
 64      5.010          42.27989
SIGMA=                 4.466E-014
SIGMA=                 8.144E-014
SIGMA=                 6.763E-014
100.     6.010          42.28012
RECORD #                    15
101.     6.026          42.27999
SIGMA=                 6.297E-014
SIGMA=                 4.368E-014
SIGMA=                 7.701E-014
 75.     7.010          42.28040
SIGMA=                 5.963E-014
SIGMA=                 6.761E-014
SIGMA=                 5.941E-014
 71      8.010          42.28071
SIGMA=                 6.509E-014
SIGMA=                 5.205E-014
SIGMA=                 9.204E-014
RECORD #                    16
101      8.510          42.28050
SIGMA=                 7.212E-014
  6.     9.010          42.28032
SIGMA=                 6.656E-014
SIGMA=                 5.729E-014
SIGMA=                 6.146E-014
 42.    10.011          42.28094
SIGMA=                 1.044E-013
SIGMA=                 7.374E-014
SIGMA=                 9.241E-014
 72     11.011          42.28070
SIGMA=                 7.400E-014
SIGMA=                 5.459E-014
RECORD #                    17
101.    11.394          42.28085
SIGMA=  .              4.986E-014
SIGMA=                 6.786E-014
 13.    12.011          42.28103
SIGMA=                 7.148E-014
SIGMA=                 8.984E-014
SIGMA=                 5.353E-014
 48.    13.011          42.28135
SIGMA=                 7.255E-014
SIGMA=                 6.440E-014
```

Figure 2. Allan variance of oscillator over a 10-h period. Mean, 6.9 E-14.

condition caused a radiation-induced frequency change for these resonators of 1 to 3 E-11/rd (Si). Similar results using BVA resonators have been reported [2]. The radiation-induced frequency changes are generally smaller for BVA resonators than for conventional resonators using electrodes plated on the vibrating quartz resonator [3]. Preconditioning is an effective method of reducing resonator sensitivity to radiation. These resonators were preconditioned with a radiation dose of 20 krd (Si) from a cobalt 60 source. Figure 6 is a typical response to radiation exposure for BVA resonators and clearly shows the effect of preconditioning.

The initial design of this oscillator included a radiation shield made of tantalum to reduce the level of radiation reaching the quartz resonator. The inclusion of this shield would increase both the assembly complexity of the oscillator and its weight. The design implemented after a more thorough analysis used a thick copper oven housing along with the other metal assemblies of the

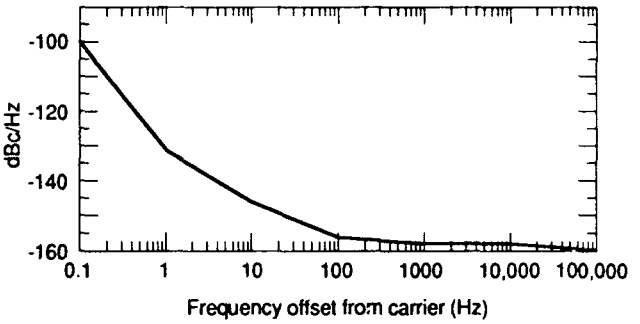


Figure 4. Phase noise ($\mathcal{L}_f$) of 5-MHz BVA oscillator.

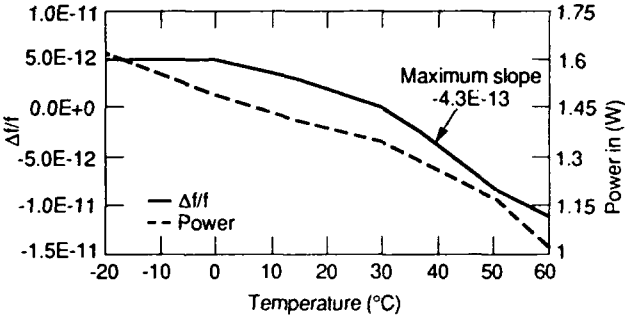


Figure 5. Oscillator performance versus temperature in vacuum.

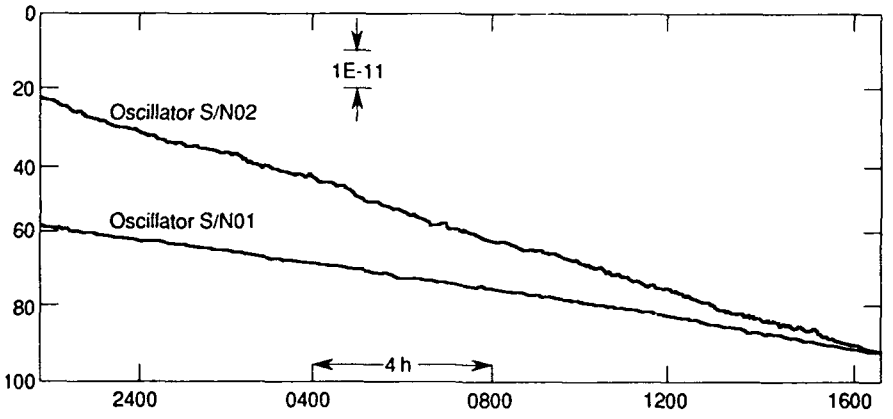


Figure 3. Oscillator Δf/f data, 100-s sample, for oscillators S/NO1 and S/NO2.

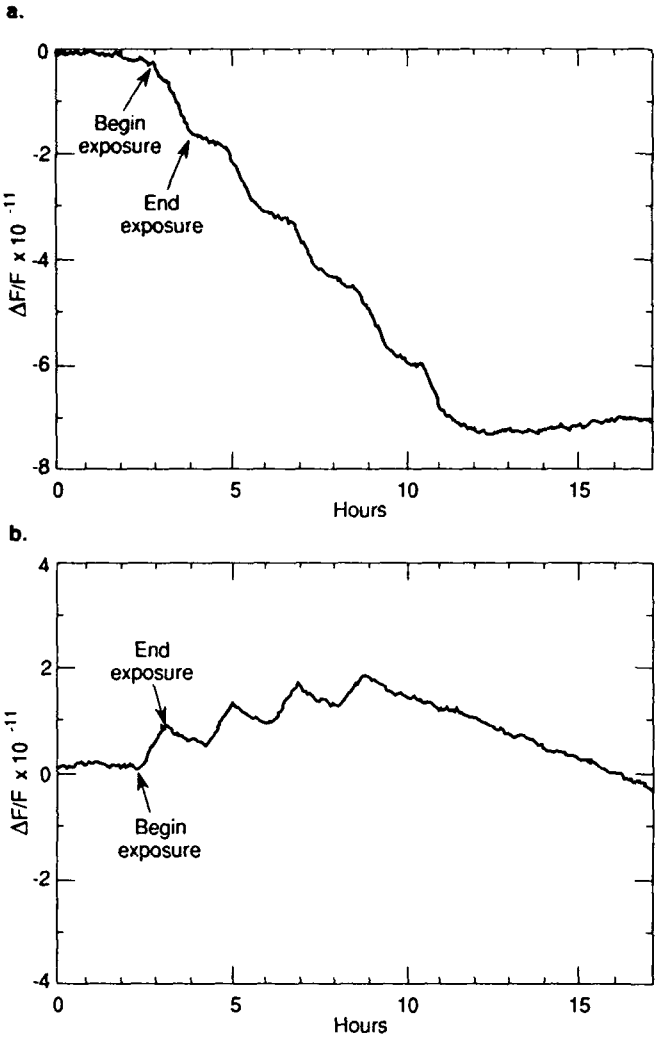


Figure 6. Frequency response of BVA-SC quartz S/N23 resonator to radiation before (a) and after (b) pre-conditioning with 20 Krd. Exposure was 0.6 rd for 42 min at a dose rate of 0.014 rd (Si)/min for both.

oscillator to provide a shield that is equivalent to 5 g/cm² of aluminum without the use of tantalum. Protons are the major radiation threat to this oscillator in orbit. A radiation shield with a density of 5 g/cm² is very effective for protons with energy levels up to 80 MeV. In a trade-off between radiation shield weight versus shielding effectiveness, 5 g/cm² is optimum. Additional shielding adds weight at a linear rate, but the effectiveness of the added radiation shielding is limited.

This oscillator uses a magnetic shield to reduce the effects of magnetic fields on the oscillator output frequency. Frequency changes produced by applying a ±1.0 G magnetic field to the oscillator were difficult to distinguish from the normal noise in the oscillator output frequency. The worst case frequency change was 2E-12/G.

The oscillator response to pressure changes is presented in Figure 7. Before this test started, the oscillator had been operating in a vacuum of 1E-6 torr. The oscillator response to being vented to atmospheric pressure is depicted in Figure 7, which also shows that the pressure is reduced from atmospheric pressure to a vacuum of 1E-6 torr. A frequency change occurs immediately, due to a thermal transient produced by a decrease in the effective thermal insulation in the oscillator. A corresponding change in DC power to the oscillator further illustrates a rapid change in the thermal characteristics of the oscillator. The net frequency change from the beginning to the end of this test was less than 2E-10.

The effect of acceleration on the output frequency of this oscillator is not an important parameter; therefore, a simple worst-case 2G tip-over test was conducted. The measured frequency change varied between 6.2E-11 and 2.5E-10, with a mean of 1.5E-10.

A unique approach was taken to isolate the BVA resonator from potentially damaging shock and vibration expected during launch of the spacecraft into orbit. An existing and proven APL-designed elastomeric frustoconical-shaped isolator molded from a low-Q RTV compound was modified to accommodate the increased weight and size of this oscillator. These isolators are carefully processed to eliminate air bubbles during molding and then baked in a vacuum to remove volatile components, which reduces outgassing to acceptable levels. A flight qualification test program for sine and random vibration and pyro shock was successfully completed using this isolation system. Data from a vibration survey of the BVA resonator suggest that the resonator probably would survive the launch environment with the resonator hard-mounted in the oscillator without vibration isolation. If additional tests confirm that the resonator could withstand the vibration environment, the size and weight of the oscillator could be reduced.

Frequency Retrace

Frequency retrace of the oscillator was characterized by removing power from the oscillator for a 24-h period and then turning power on again. Measurements from this test are depicted in Figure 8. The oscillator output frequency was –8.7E-6 when power was first applied and gradually increased

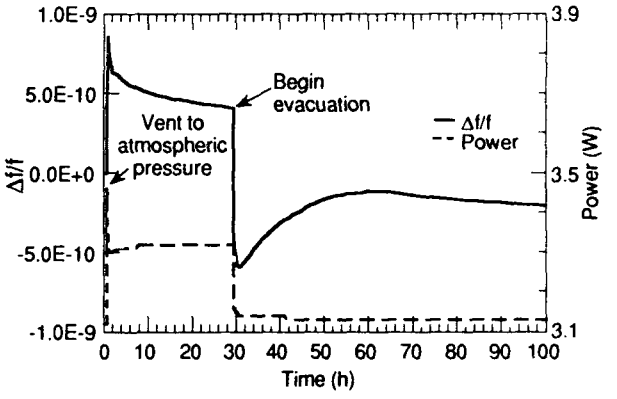


Figure 7. Oscillator response to pressure change.

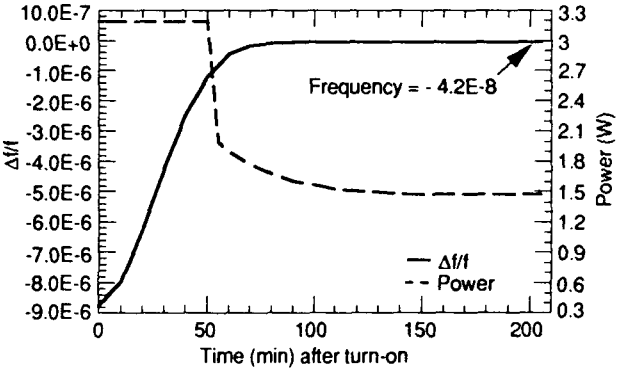


Figure 8. Oscillator turn-on characteristics in vacuum from turn-on to 60–70 min after turn-on. Ambient temperature, 32°C.

during the first 60 to 70 min of the test until the oscillator oven reached its operating temperature. At this time, the output frequency was ≈ -2E-7 and still increasing. Figure 9 shows the oscillator output frequency from 100 min to 1500 min after turn-on. After 24 h (1440 min) the output frequency or retrace was 2E-10 below the oscillator frequency when it was turned off.

DC input power to the oscillator at turn-on was 3.12 W and 1.48 W after the oscillator oven stabilized. The ambient temperature was 32° C during this test.

Abrupt Frequency Changes

Abrupt frequency changes or frequency jumps were observed in some BVA resonators. Strip chart data taken at a 100-s sample rate (Fig. 10) illustrates one of these events. The oscillator frequency had been very stable with an Allan variance of <2E-13 before the erratic behavior shown in Figure 10. After 6 h of erratic performance, the oscillator again became quiet and the net frequency change near zero, taking into account the established aging rate. The magnitude and duration of the frequency jumps occur over a wide range. Frequency jumps occur more often during the first few days after oscillator turn-on and usually decrease as operating time accumulates; however, frequency jumps are not restricted to the early phase of oscillator operation. The data shown in Figure 10 is for an oscillator that had been operating for more than eight months.

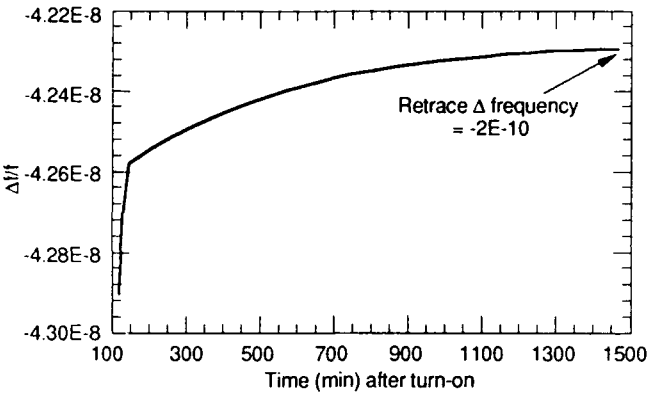


Figure 9. Oscillator turn-on characteristics in vacuum from 100 to 1500 min after turn-on. Ambient temperature, 32°C.

After frequency jumps were observed in several resonators, a screening procedure was implemented to select the best resonators for installation in flight oscillators. With this precaution, the performance of the flight oscillators was not compromised. The same test oscillators were used to evaluate several resonators. Some resonators did not have frequency jumps, whereas others produced multiple jumps; it is therefore safe to conclude that frequency jumps originate in the resonators, not in the oscillator circuitry.

Oscillator Design

Because this oscillator is critical to the spacecraft mission, parts with the lowest failure rates are used in this oscillator design. To enhance the reliability of the oscillator, the voltage, current, and power ratings of the parts chosen were heavily derated. A minimum useful life of three years was the design goal for this oscillator, with a potential extended mission life of five years. The spacecraft has two completely independent oscillators to provide redundancy if a problem should occur with either oscillator during the life of the mission.

The oscillator circuit is a modified Colpitts type that has both DC and AC negative feedback to stabilize the circuit operating point and to reduce flicker noise. The BVA resonator and critical oscillator components are housed in a high-stability, single, proportionally controlled oven. The oven maintains a constant temperature over the ambient operating temperature range to ensure high performance. Output amplifiers with high isolation (>70 dB) provide two independent output signals at a power level of +7 dBm. Dual high-stability DC voltage regulators isolate critical oscillator and amplifier circuits from input power supply variations and noise.

Oscillator performance and its mechanical design are closely interrelated. Because the thermal and mechanical design of the oscillator are highly interdependent, the mechanical design of the oscillator requires more care than simply designing a box to enclose circuits. The oscillator chassis, or outer housing, provides mechanical support for the various oscillator components and is part of the vibration isolation system. The chassis and its internal structures reduce the probability of electromagnetic interference (EMI) signal from entering or exiting the oscillator. Methods used to reduce EMI problems include placement of lips on the seams between metal covers and the oscillator housing, use of partitions to separate and isolate circuits, and placement of EMI filters on all wiring between electromagnetic clean compartments and EMI dirty areas. The flask assembly—consisting of the quartz resonator, the low-level oscillator circuit, the oven-control circuit, and the oven—forms the heart of the oscillator. Because these components are sensitive to vibration and

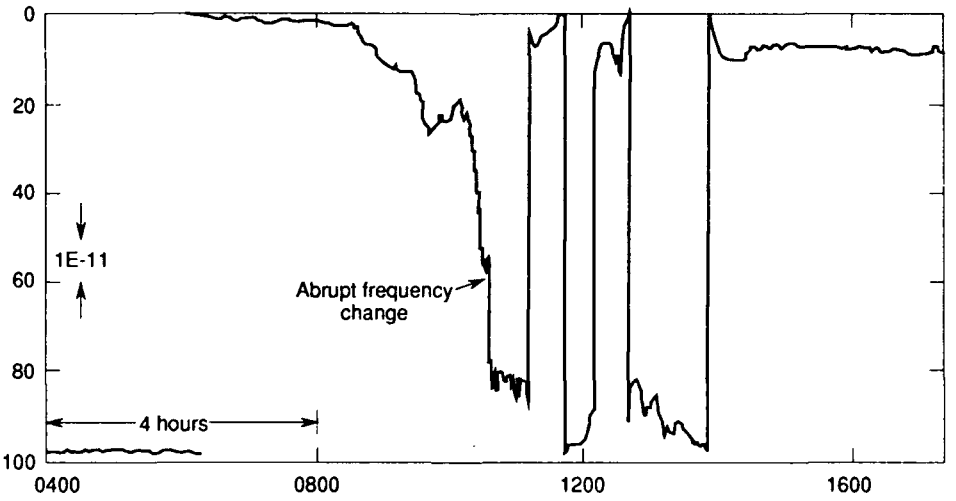


Figure 10. Abrupt frequency changes in 5-MHz BVA resonators.

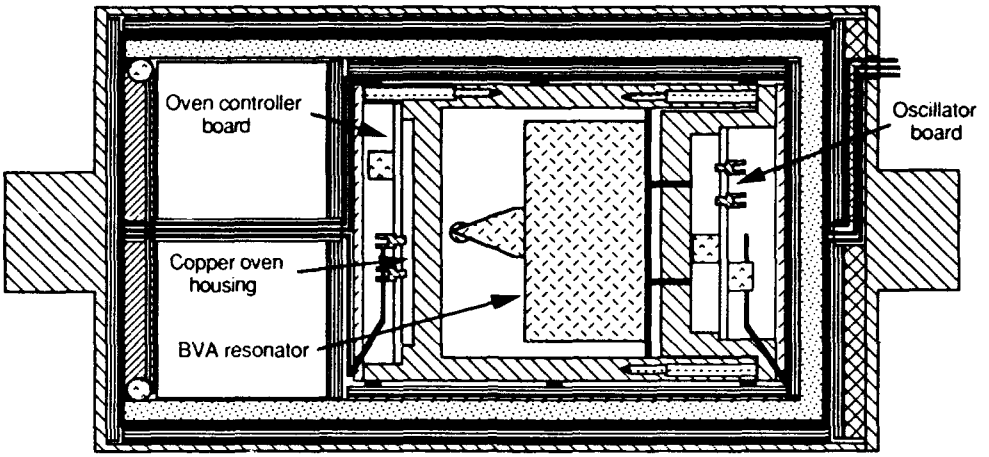


Figure 11. Cross section of oscillator flask assembly.

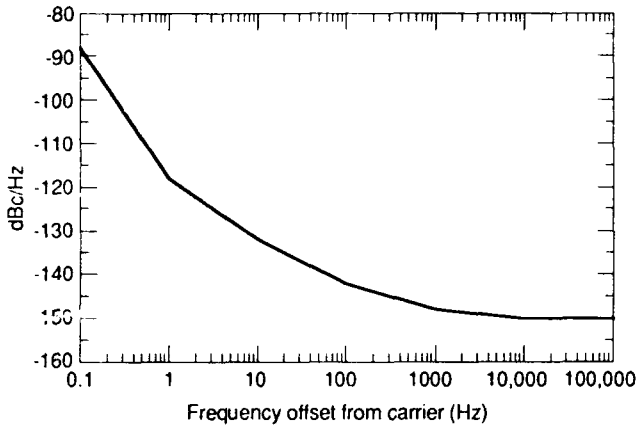


Figure 12. Phase noise ($\mathcal{L}_f$) of 10-MHz BVA oscillator. Measured data does not assume equal noise sources.

shock, the vibration isolation system described previously isolates the entire flask assembly from the oscillator chassis. Figure 11 depicts a cross section of the flask assembly.

The oscillator weighs 3.88 lb and occupies a volume of 118 in^3.

10 MHz BVA Resonators

Limited testing of 10-MHz BVA resonators was conducted. The Allan variance and phase noise ($\mathcal{L}_f$) are presented in Figures 1 and 12, respectively. Although performance is not as good as 5-MHz BVA resonators, especially the phase noise near the carrier, the results are still very good. Aging rates of the 10-MHz resonators are 2E-11/24 h.

The 10-MHz oscillator circuit, oven, and mechanical assemblies are not as well developed as the 5-MHz oscillators and may contribute to a decrease in performance. One very positive aspect of the 10-MHz resonators is the absence of the frequency jumps observed in the 5-MHz resonators. The number of 10-MHz resonators evaluated is small and, when additional units are tested, frequency jumps may be found.

Conclusion

Superior performance by a newly developed, space-qualified oscillator using a BVA quartz resonator has been demonstrated. This oscillator design minimizes frequency changes caused by the environmental effects of temperature, magnetics, power-supply noise, load changes, and EMI. Probably the greatest environmental effect on oscillator frequency in orbit will be radiation. Oscillator performance continues to be dominated by resonator quality.

Acknowledgments

This work was supported by the TOPEX/POSIDON Project Office at The Jet Propulsion Laboratory. The author is grateful to many colleagues, especially J. M. Cloeren and L. E. Stillman for valuable discussions and contributions to the development of this oscillator.

References

[1] R. J. Besson, "A New Electrodeless Resonator Design," Proceedings of the 31st Annual Symposium on Frequency Control, June 1977, pp. 147–152.

[2] J. J. Suter, A. G. Bates, J. M. Cloeren, J. R. Norton, B. Schlueter, and U. Peier, "The Susceptibility of BVA-SC Resonators to Proton Ionization Effects," Proceedings of the Third European Frequency and Time Forum, March 1989, pp. 11–21.

[3] J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results From Gamma Ray and Proton Beam Radiation Testing Of Quartz Resonators," Proceedings of the 38th Annual Symposium on Frequency Control, May 1983, pp. 63–372.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A SPACE OSCILLATOR WITH CYLINDRICAL OVEN AND SYMMETRY

Marc Mourey, Raymond J. Besson

Ecole Nationale Supérieure de Mécanique et des Microtechniques
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - France

ABSTRACT

A new 10 MHz BVA quartz crystal oscillator for severe environments was presented last year [1]. Drift rates of less than 10^{-11} per day and $\sigma_y(\tau)$ of 5 to 7 x 10^{-13} at 10 seconds were obtained together with spectral purity of 115 dB at 1 Hz and 160 dB at 1000 Hz. 10 MHz third overtone tactical BVA cut resonators were used.

However, it was found necessary to improve the oven design especially for space uses in satellites. For such uses the oscillator circuit had also to be modified to meet magnetic field effect requirements. In this paper improvements are described and results on the new oscillator are given and commented.

A cylindrical oven whose axis is the very axis of the quartz resonator (in HC 40 can) is studied. Thermal flux is canalized on its axis at a place where thermistor is located. Thermal exchanges include conduction and radiation effects (however reduced by use a dual enveloppe). The oven is theoretically studied through a scheme which uses electric analogs and results are obtained for external temperature between - 25°C and + 60°C. The thermistor is located onto the resonator can where temperature is very close to the turn over point. Analytical modelization uses oven symmetry.

Thermal regulation design includes a frequential analysis between 0.001 Hz and 10 Hz with a 60 dB gain. A response analysis to temperature steps and to linear temperature variation (0.5°C/mn) is also performed. The thermal gain and dynamical behaviour of oven electronics is deduced. Insertion of a correction network with phase advance in oven electronics yields a fast oven response.

The oscil' r improvements in space conditions are described and co..nented.

INTRODUCTION

This paper presents improvements needed to turn the oscillator previously described [1] into a space oscillator for use in satellite environment. A miniature BVA_2 10 MHz, SC cut 3rd overtone crystal is used. The resonator, which comes inside a regular HC 40/U enclosure, yields low pressure sensitivity and a good compromise between shock and acceleration sensitivities. The unit is carefully prepared from swept material and colwelded under ultra low pressure (less than 5 10^{-8} mbar). The main crystal characteristics are :

Frequency : 9'999'983 (- 10 Hz + 6 Hz)
T.O.P. : 80°C (- 2°C + 3°C)
Q factor : 1.35 10^6 (typical)
R_1 : 95 Ω (typical)
G maximum sensitivity : less than 10^{-10}/g
External pressure sensitivity : 5 10^{-11}/bar (typical)
Magnetic field sensitivity : 4.4 x 10^{-12}/Gauss

A new oven devoted to use in space conditions has been studied whose mechanical first resonant frequency is higher than 2000 Hz. Thermal gain is higher than 2400. Under those conditions, for an external temperature between - 25°C to + 60°C the crystal unit is submitted to a temperature variation of 3.5 x 10^{-2}°C. Oscillator is designed to satisfy European Space Agency shock and vibration norms. Automatic gain control is used together with isolation amplifier of 100 dB or more. A selecting resonant circuit with a Q factor of approximately 30 is used to select the C mode frequency when oscillations are started. This circuit is automatically damped by use of AGC signal when oscillations are established so as to avoid instabilities due to the selecting circuit. The total volume of oscillator is less than 300 cm^3.

I - OVEN DESIGN

Preliminary theoretical study of the thermal behaviour of this particular type of BVA resonator has been done in past years. As a consequence it was found useful to preserve the cylindrical symmetry of resonator and HC 40 enclosure and to design a cylindrical oven whose axis is the axis of the resonator crystal and mounting. Since the oscillator is operating in space several reflectors polished and gilded are used to limit thermal exchanges by radiation. The oscillator is fixed by a titanium spacer so as to limit exchanges by heat conduction.

CH2965-2/91/0000-431 $1.00 © 1991 IEEE

The oven control is proportionnal and uses a thermistor located at the place of maximum thermal flux. An electrical equivalent network of the oven, using electric analog, has been studied with a P.spice simulation together with the oven electronic control. Result appear on Fig. 1. Thermal regulation design includes a frequential analysis between 0.001 Hz and 10 Hz (Fig. 2), a response analysis to linear temperature variation of 0.5°C/minute (Fig. 3). Warm up may be simulated for various external temperatures (Fig. 4 and Fig. 5). The thermal gain and dynamical behaviour of oven is deduced.

Of course a mechanical analysis of the oven is performed through a finite element method. The final design correspond to a first resonant frequency higher than 2000 Hz.

II - OSCILLATOR DESIGN

The oscillator is basically the oscillator previously described [1] and is represented on Fig. 6. The oscillating loop uses a single very low noise transistor previously "noise selected". A CE configuration is used for the amplifier together with π phase shifting transformer. A true AGC is used which involves an F.E.T. and allows oscillation levels of the Xtal between 60 and 100 μW. The final level is choosen for a good compromise between short term stability and aging. The frequency control has been designed to allow frequency variations in the 10^{-6} range. Two cascode output amplifiers yield an isolation factor larger than 100 dB.

Oscillator circuit previously described needed a C mode selection obtained through a parallel resonant circuit mounted on transistor collector. This circuit was sensitive to external magnetic field thus leading to frequency instabilities (which was too bad since BVA tactical resonators turn out to be at least 3 times less sensitive to external magnetic field [2]). After starting, the selectivity of the resonant circuit is reduced. So, the slope of $\Delta f_{out} / \Delta_{\varphi filter}$ is lowered and the stability improved.

III - MEASUREMENTS AND CONCLUSIONS

Measurements have been performed to check oven thermal gain and characteristics under vacuum yielding good agreement with theory. No resonant mechanical frequency has been found lower than 2200 Hz. Preliminary phase noise and short term stability measurements have been performed according to Fig. 7 showing some improvement at 1 Hz from the carrier and a better $\sigma_y(\tau)$ ($\leqslant 5\ 10^{-13}$ at 10 s). However additional tests must be conducted on the final Engineering model in space conditions. Nevertheless, as pointed out by other authors [3] most oscillator performances seem to be improved by use of BVA type resonators. For this reason a space qualification of these resonator is now undertaken in our laboratory.

ACKNOWLEDGEMENTS

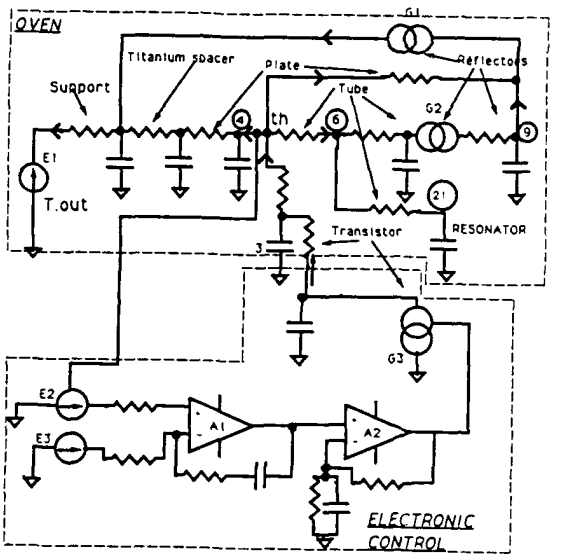
The authors would like to thank Compagnie Industrielle de Radioélectricité (C.I.R.) in Gals (Switzerland) for fruitful cooperation.

REFERENCES

[1] R.J. Besson, M. Mourey, "A B.V.A. quartz crystal oscillator for severe environments", 44th Annual Symposium on Frequency Control (A.S.F.C.), Proceedings pp 593-596, 1990.

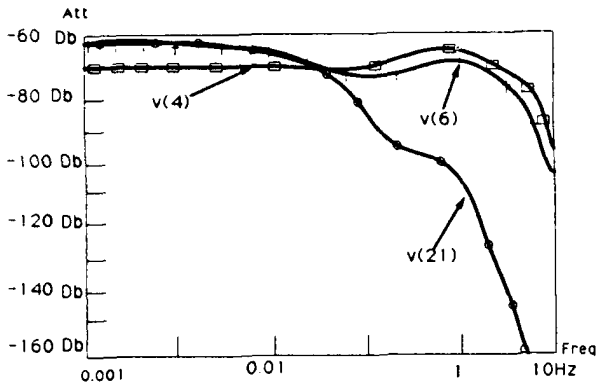
[2] R. Brendel, C. Hassani, P. Krux, "Magnetic sensitivity of oscillator components", 5th European Frequency and Time Forum, Besançon, 1991.

[3] J.R. Norton, "BVA-type quartz oscillator for spacecraft", 45th Annual Symposium on Frequency Control, Los Angeles, 1991



OVEN EQUIVALENT NETWORK

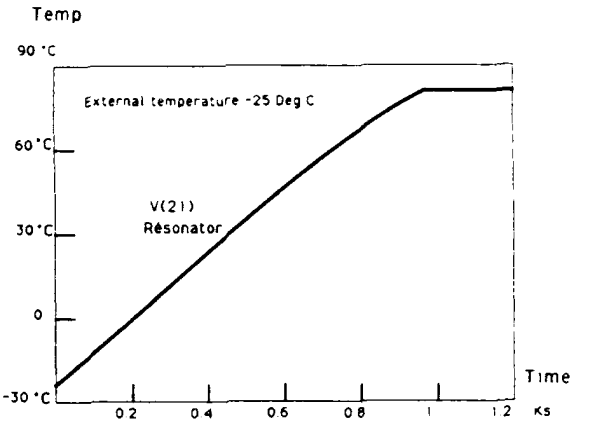
Fig. 1



A.C ANALYSIS

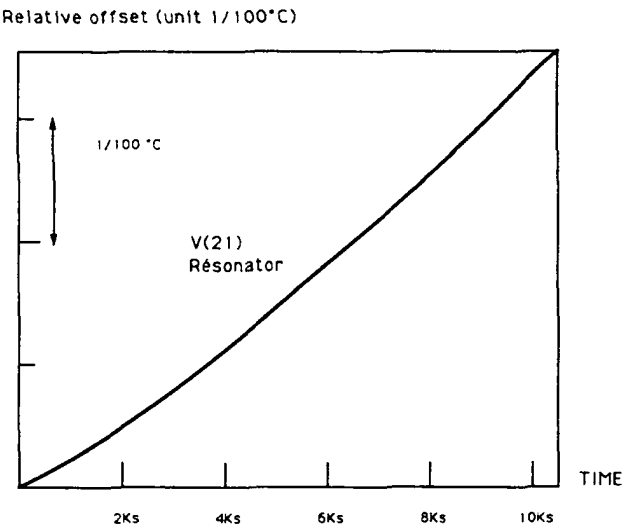
v(4):thermistor
v(21):résonator
v(6):tube

Fig. 2



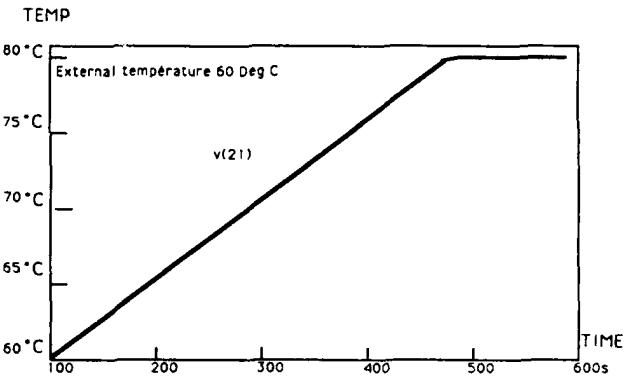
WARM UP

Fig. 4



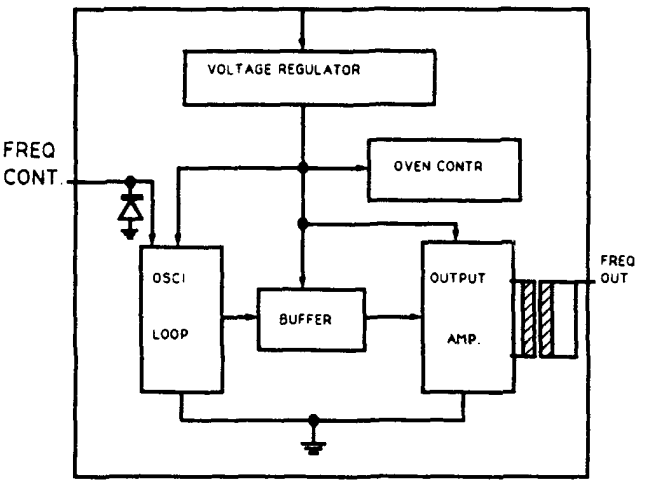
TEMPERATURE RAMP RESPONSE (0.3°C/MN
-25°C +60°C)

Fig. 3



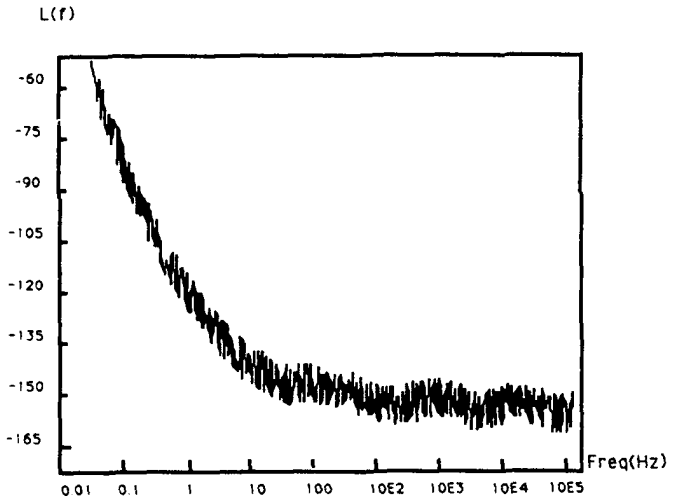
WARM UP

Fig. 5



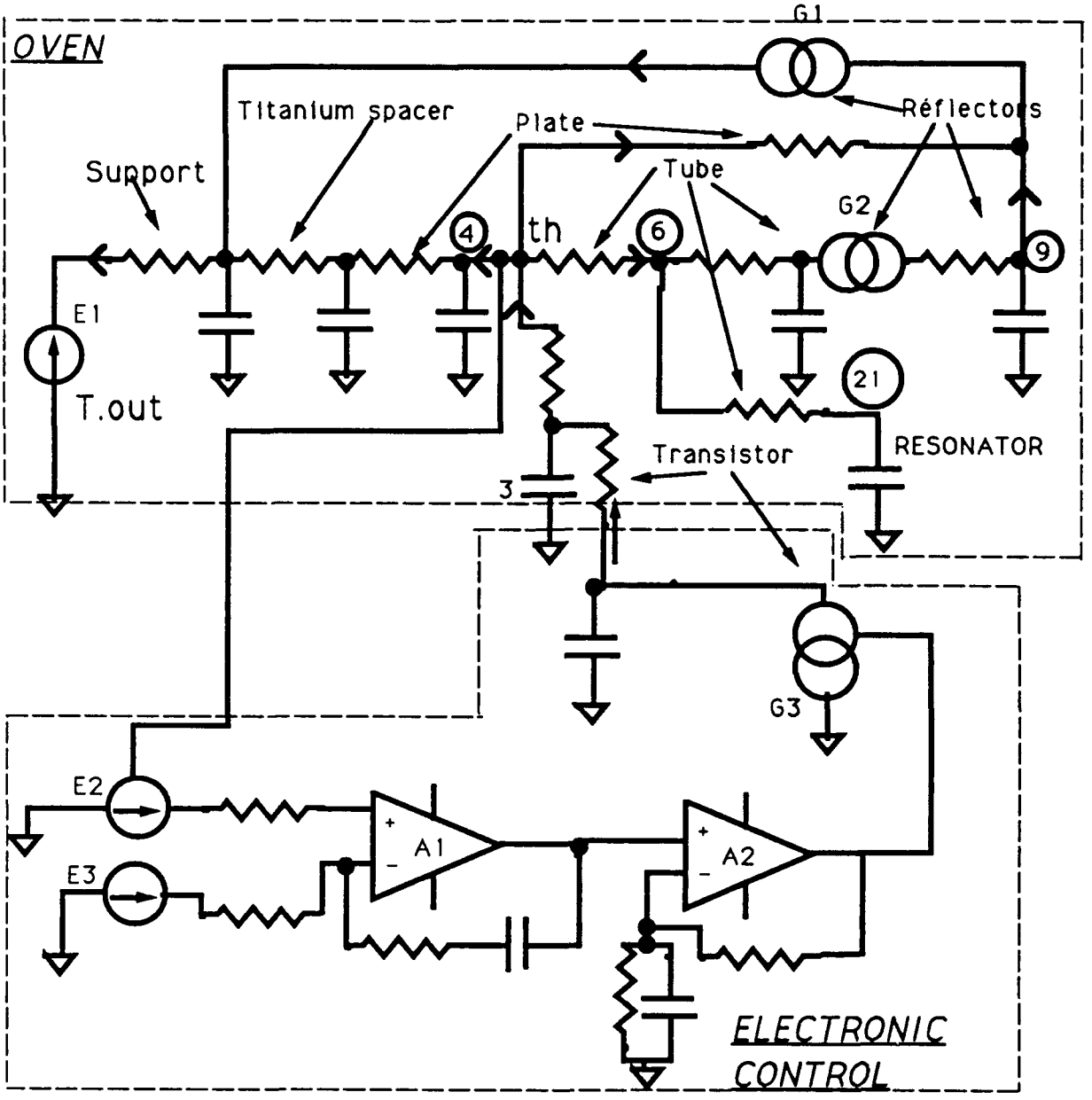
Block DIAGRAM

Fig. 6



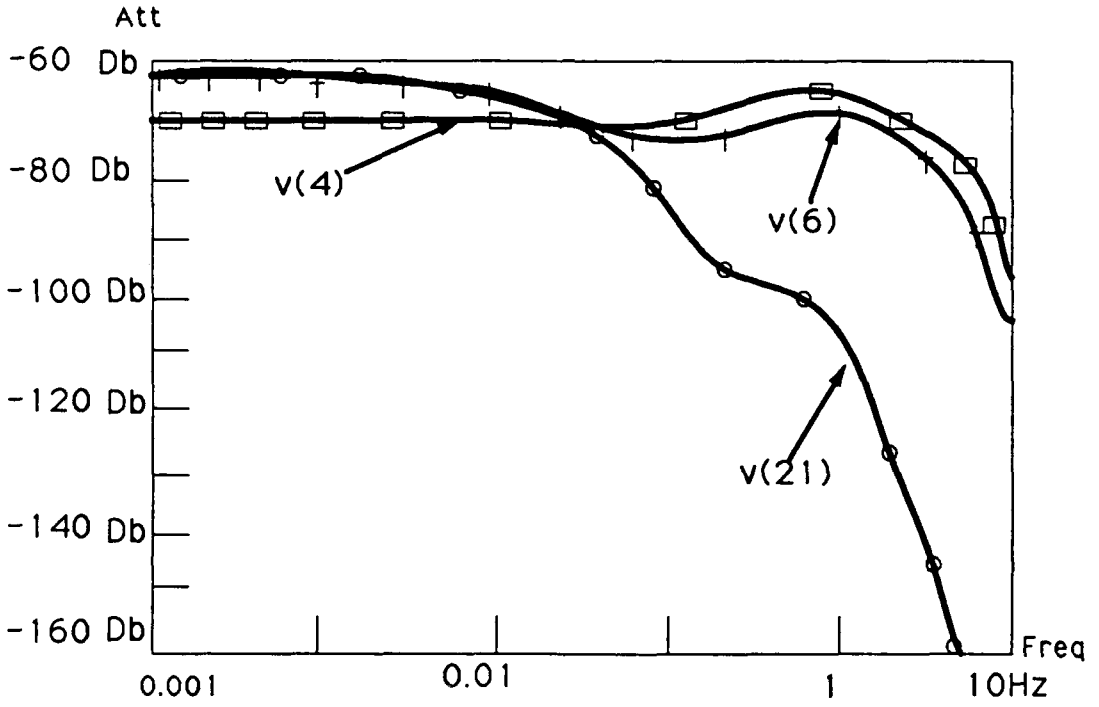
PHASE NOISE

Fig. 7



OVEN EQUIVALENT NETWORK

Fig. 1



A.C ANALYSIS

v(4):thermistor
v(21):résonator
v(6):tube

Fig. 2

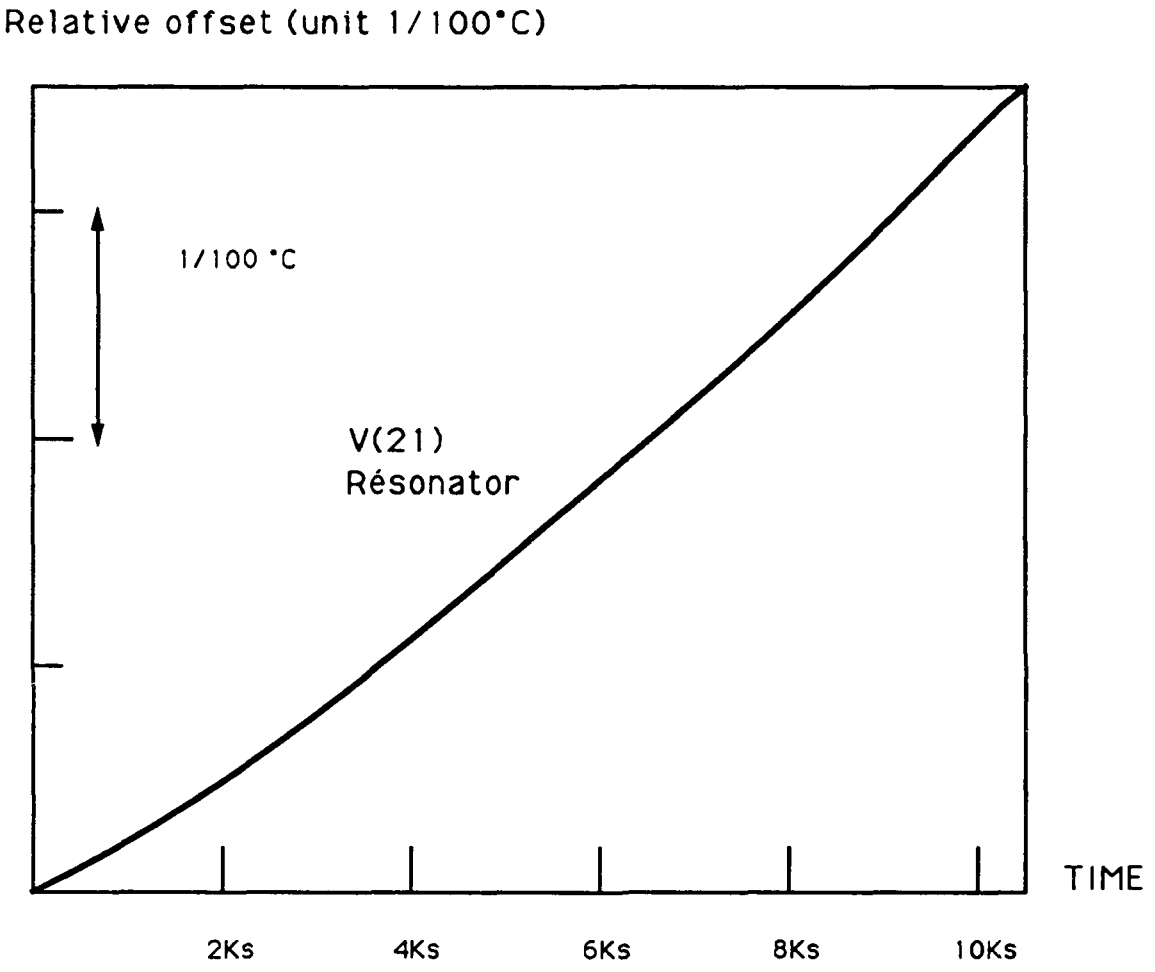

Relative offset (unit 1/100°C)
1/100 °C
V(21)
Résonator
TIME
2Ks
4Ks
6Ks
8Ks
10Ks
TEMPERATURE RAMP RESPONSE (0.5°C/MN
-25°C +60°C)

Fig. 3

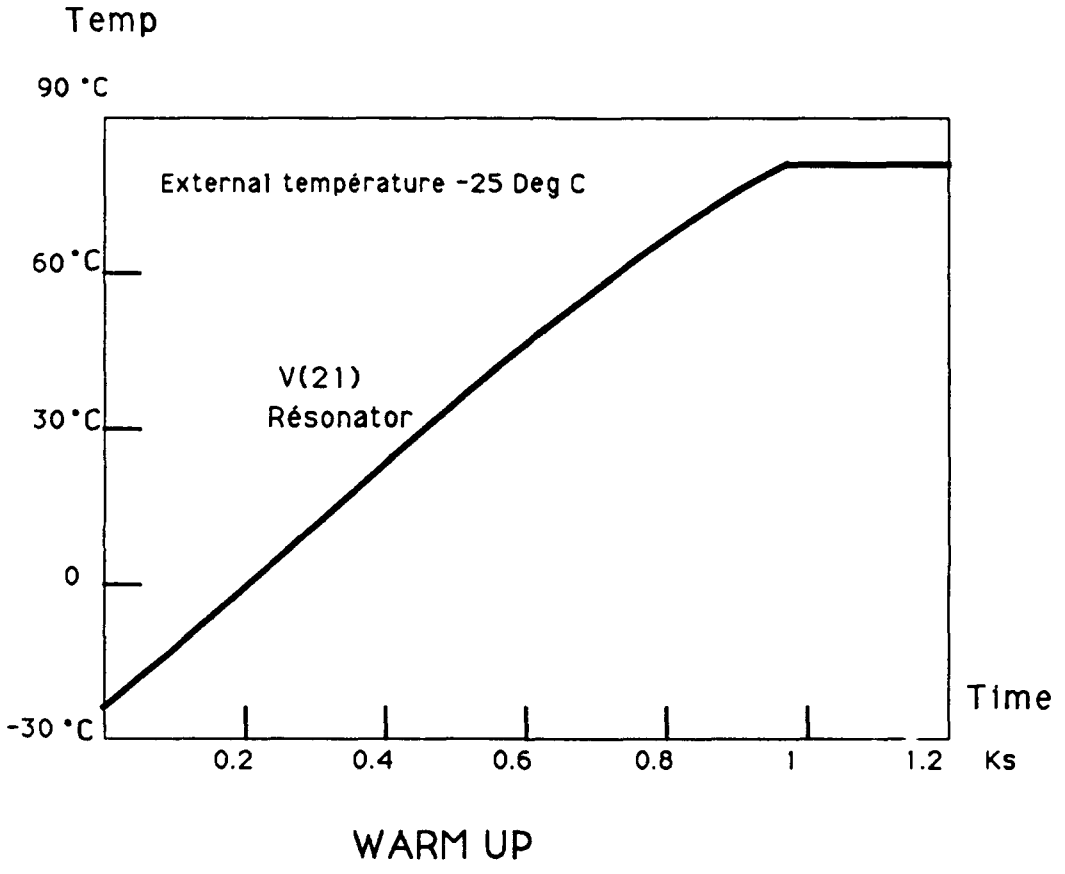
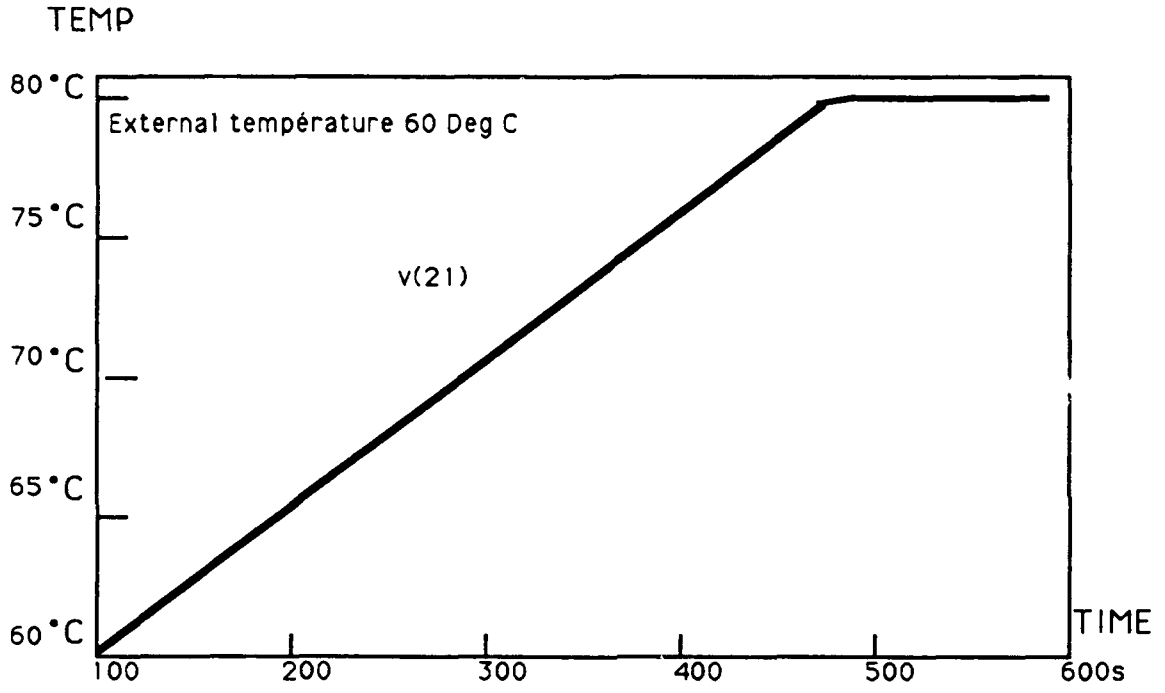
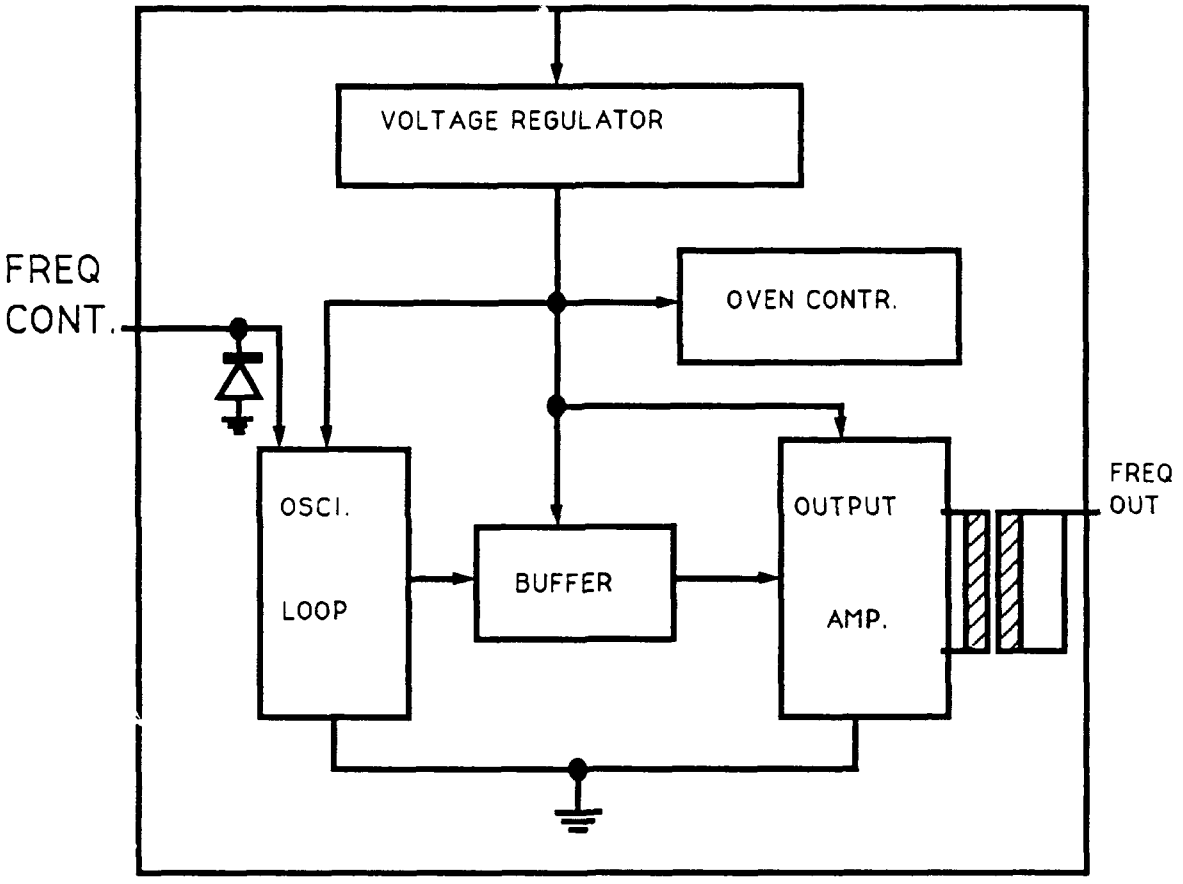

Temp
90 °C
External température -25 Deg C
60 °C
V(21)
Résonator
30 °C
0
-30 °C
Time
0.2
0.4
0.6
0.8
1
1.2
Ks
WARM UP

Fig. 4


TEMP
80°C
External température 60 Deg C
75°C
v(21)
70°C
65°C
60°C
TIME
100
200
300
400
500
600s

WARM UP

Fig. 5



Block DIAGRAM

Fig. 6

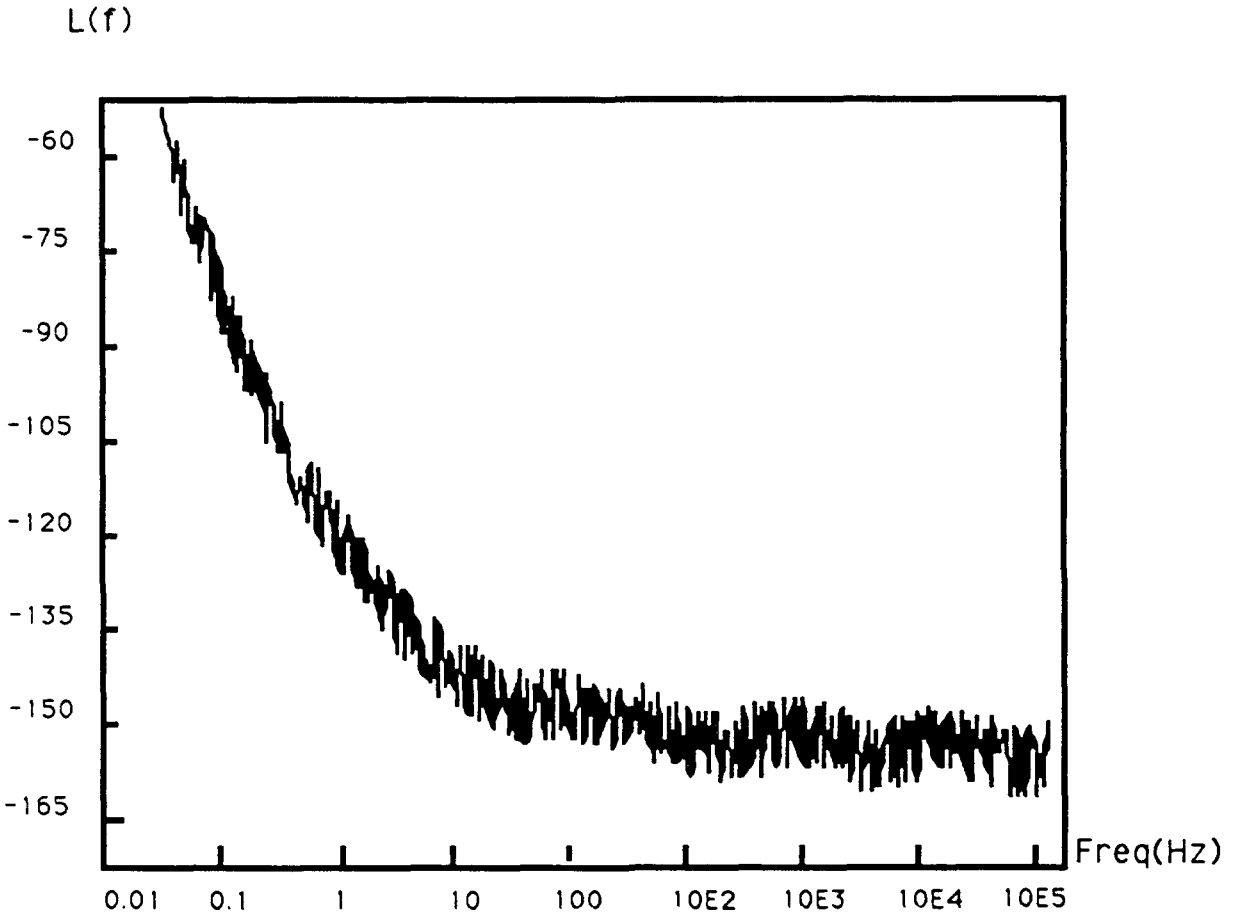

L(f)
-60
-75
-90
-105
-120
-135
-150
-165
0.01
0.1
1
10
100
10E2
10E3
10E4
10E5
Freq(Hz)
PHASE NOISE

Fig. 7

A SPACE-QUALIFIED FREQUENCY SYNTHESIZER

Matthew J. Reinhart, P.E.
The Johns Hopkins University
Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099

Abstract

This paper describes the performance and some of the design features of a dual-frequency synthesizer developed and built for the Topography Experiment (TOPEX) program. The space-qualified synthesizer generates high-quality 4.096 MHz and 19.056391 MHz radio-frequency signals from a 5 MHz reference source for use by Global Positioning Satellite and TDRSS transponder equipment aboard the TOPEX-Poseidon satellite.

Performance

The measured performance of the flight synthesizer approaches that of a good reference oscillator. The spurious amplitude spectrum (excluding harmonics) is less than –80 dBc from 1 Hz to 1.5 GHz offset from the carrier of each output frequency. The only observed spurious output that results from the synthesis process appears during phase noise tests as a spur at the –100 dBc/Hz level and 8 kHz offset. The single sideband (SSB) phase noise of the 4.096 MHz output as shown in Figure 1 is –110 dBc/Hz at 1 Hz offset from the carrier and –147 dBc/Hz at 1 kHz offset. The 19.056391 MHz output also shown in Figure 1 measures –95 dBc/Hz at 1 Hz offset and –130 dBc/Hz at 1 kHz offset. The short-term stability (Allan Variance) of each frequency output in 1-second and longer measurement intervals is less than 5 parts in 1×10^{-12} fractional stability as shown in Figure 2. The stability of the synthesizer improves with longer measurement periods, and thus the outputs gradually acquire the long-term stability of the reference oscillator.

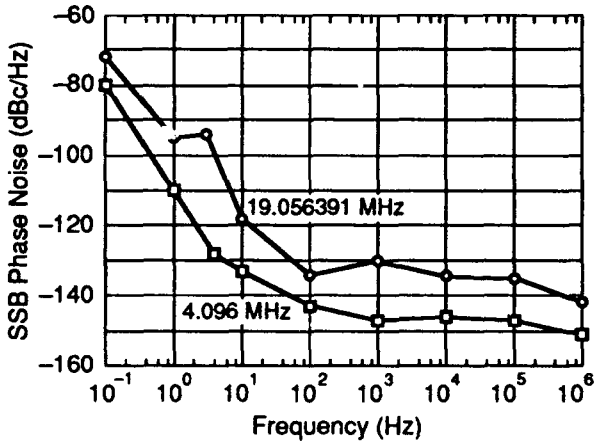


Figure 1. TOPEX Synthesizer Phase Noise.

Space Qualification

The synthesizer has been fully qualified for space flight use. Extensive thermal vacuum tests have verified unimpaired performance over a temperature range of –20° to +60° C. The phase noise and stability had no measurable dependence on temperature within the test range. The qualification also included vibration tests to ensure launch and deployment survival. Synthesizer operation remained stable during shock, sine, and random vibration tests. The synthesizer was exposed to cobalt 60 radiation to simulate in-orbit radiation. During two sequential 42-minute exposures of 0.6 Rad (silicon), the synthesizer maintained phase lock with only a few millivolts of change in the 7-volt PLL control signal.

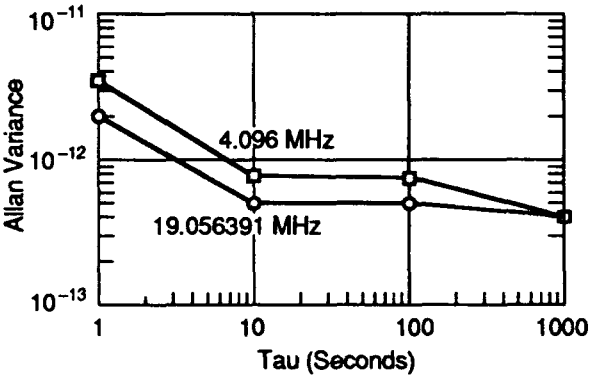


Figure 2. TOPEX Synthesizer Fractional Frequency Stability.

CH2965-2/91/0000-442 $1.00 © 1991 IEEE

Circuit Description

The synthesizer produces the 4.096 MHz and 19.056391 MHz output signals from a 5 MHz reference signal through the process depicted in Figure 3. Two high-quality voltage-controlled crystal oscillators (VCXO's) phase locked to the reference oscillator are employed to ensure excellent long-term stability. The VCXO's provide a low spurious amplitude spectrum and low phase noise performance with high short-term stability outside the bandwidth of the phase-locked loops (PLL's).

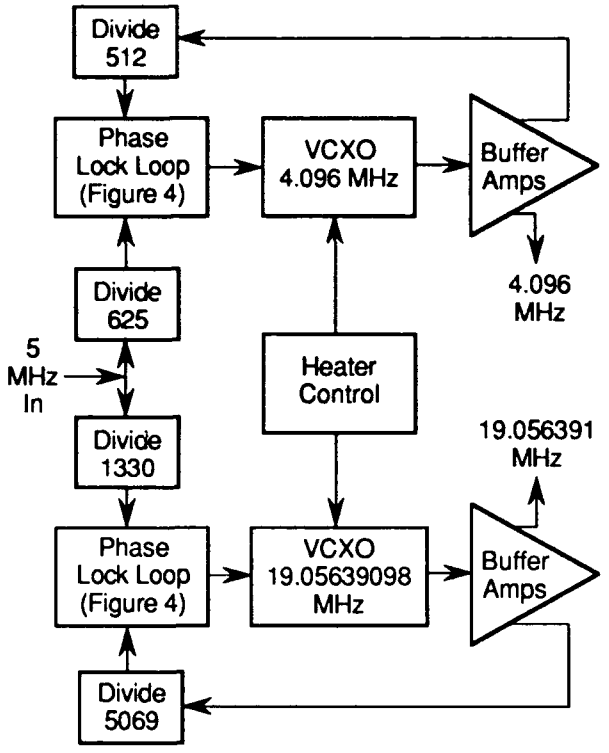


Figure 3. TOPEX Synthesizer Functional Block Diagram.

Each output signal and the 5 MHz reference signal are digitally divided to a common frequency used to drive a unique high-gain phase detector (see Figure 3). A high-gain detector is necessary to counteract the large division ratios in the loop and the low VCXO tuning range. A conventional double-balanced mixer would have required a DC amplifier with high gain to obtain the desired loop parameters and so was deemed unsuitable because of the inherent noise contributions of a high-gain DC amplifier and mixer.

The 4.096 MHz and 5 MHz signals, as shown in Figure 4, which illustrates the 4.096 MHz synthesizer loop, are divided to 8 kHz and fed into an analog switch phase detector that provides the entire 2 π phase slope in one 5 MHz cycle rather than an 8 kHz cycle. The phase detector operates by closing and opening an analog switch in one 5 MHz cycle once in every 8 kHz cycle (5 MHz /625 = 8 kHz). The input signal to the analog switch is a 50% duty square wave at an 8 kHz rate (4.096 MHz /512 = 8 kHz). The loop locks with the leading edge of the square wave approximately centered within the pulse from the 5 MHz divider, which closes the switch. The analog switch samples the 8 kHz square wave during its state change, giving an average analog representation of the relative phase. In the presence of slight phase errors, the analog switch will pass a disproportionate amount of either the high or low state of the 8 kHz square wave. This will result in a control voltage offset towards one of the

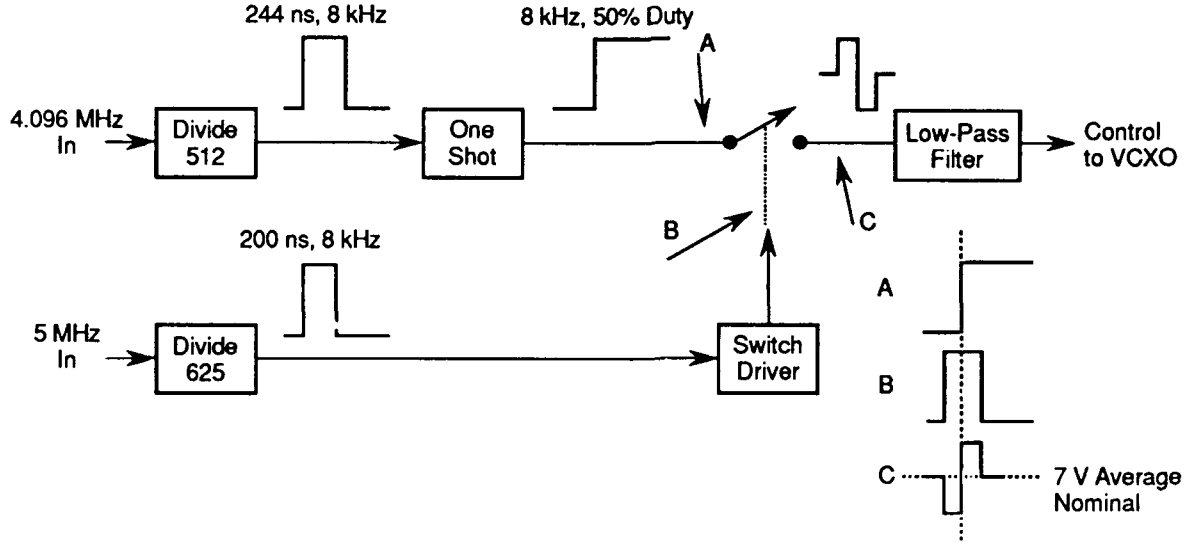


Figure 4. TOPEX Synthesizer Phase-Locked Loop Block Diagram.

square wave states, depending on the sign of the phase error. The analog switch output is low-pass–filtered with a lag–lead filter to attenuate loop noise and 8 kHz ripple with unity open loop gain at approximately 0.2 Hz. The gain of the phase detector, including the dividers, is 12 volts per 2 π radians of 5 MHz, or 2.4 volts per one radian at 4.096 MHz.

The operation of the 19.056391 MHz synthesizer loop is similar to that of the 4.096 MHz loop, but the signals are divided to approximately 4 kHz, and a wider loop bandwidth of approximately 2 Hz is used. The wider loop bandwidth improves both the short-term stability and close-in phase noise of the 19.056391 MHz VCXO by tracking the excellent characteristics of the 5 MHz reference oscillator. In the case of the 4.096 MHz synthesizer, the 4.096 MHz VCXO provides lower phase noise at 1 Hz than the internal loop noise. Therefore, if a 2 Hz bandwidth were used for the 4.096 MHz loop (as in the 19.056391 MHz loop), the close-in phase noise performance would be degraded because of excessive loop noise. The performance difference between the VCXO's results from the much higher quality factor of the 4.096 MHz resonator with respect to the 19.056391 MHz resonator.

The VCXO's employ FC-cut quartz crystal resonators housed in temperature-controlled ovens. The temperature control of the resonators minimizes the effects of environmental changes on stability and also reduces the required tuning range. The VCXO's are based on a modified Colpits design with a fractional frequency tuning range of 1 part in 1×10^6 to accommodate long-term aging of the resonators. The resonator used in the 4.096 MHz VCXO was screened for phase noise performance at 1 Hz offset from the carrier, where a variation of 20 dB was measured among resonators tested from the same manufacturer with the same test oscillator. This clearly demonstrates that the 4.096 MHz synthesizer phase noise at low-frequency offsets is dominated by the resonator. The resonator quality factor as supplied with the devices by the manufacturer ranged from 0.9 million to 1.8 million. As expected, the higher quality factors provided the better phase noise performance. The resonator drive level was optimized for low phase noise performance and is maintained by a feedback loop to improve the stability.

The synthesizer outputs are buffered with tuned amplifiers (see Figure 3) that provide 0 dBm of output power into 50 ohms and over 60 dB of isolation from load variations. The total input power to the unit is 4.0 watts in a vacuum at 25° C and 6.0 watts at turn-on until the resonator ovens reach their operating temperature of 70° C. The synthesizer as shown in Figures 5 and 6, is housed in a double-sided chassis with ten isolated compartments and has a mass of 1150 grams fully assembled with side covers. The isolated compartments and the use of filtered feedthroughs for control and supply lines control interference among the circuit boards. Every circuit board has a linear regulator to reduce supply line interference between circuit boards and external sources further. One of the chassis compartments is a dedicated "dirty box" where all the DC and analog inputs and outputs are filtered before entering the unit.

Analysis

The phase-locked loops were analyzed for frequency and time domain response. The 4.096 MHz open loop response as shown in Figure 7 has unity gain at 0.2 Hz with over 100 dB and 45° of stability margin (see Figure 8). Essentially, the loop is unconditionally stable with its lag–lead filter, since the 180° phase shift occurs as a result of additional higher-frequency filtering. The closed loop response and time domain step response are shown in Figures 9 and 10, and the damping factor was determined to be 0.55 with a natural frequency of 0.34 Hz.

The 19.056391 MHz loop was similarly analyzed. As shown in Figure 11 , the open loop unity gain occurs at 1.7 Hz with 80 dB of gain margin and 60° of phase margin (see Figure 12). The 19 MHz loop uses simpler

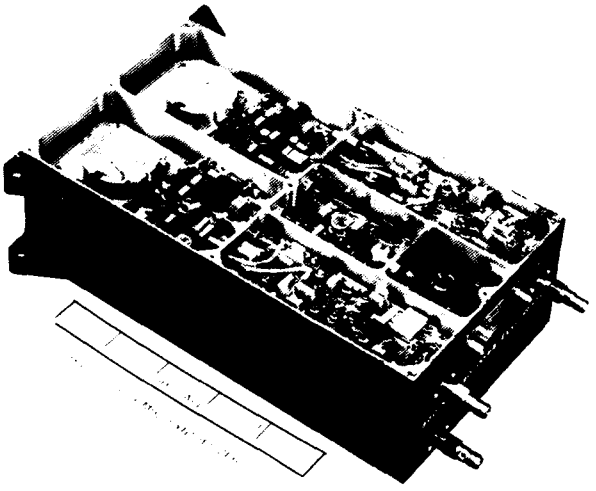


Figure 5. TOPEY Synthesizer (Top View).

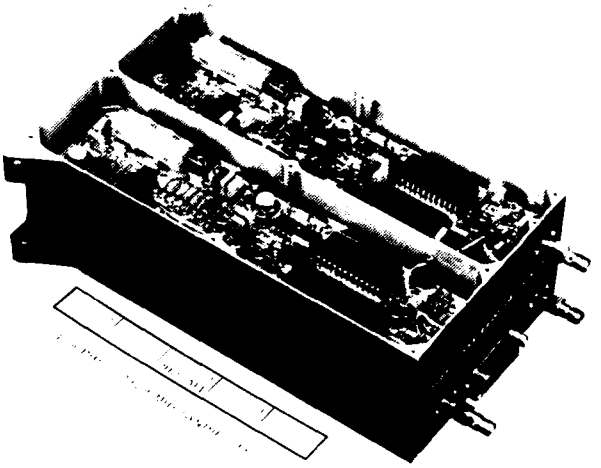


Figure 6. TOPEX Synthesizer (Bottom View).

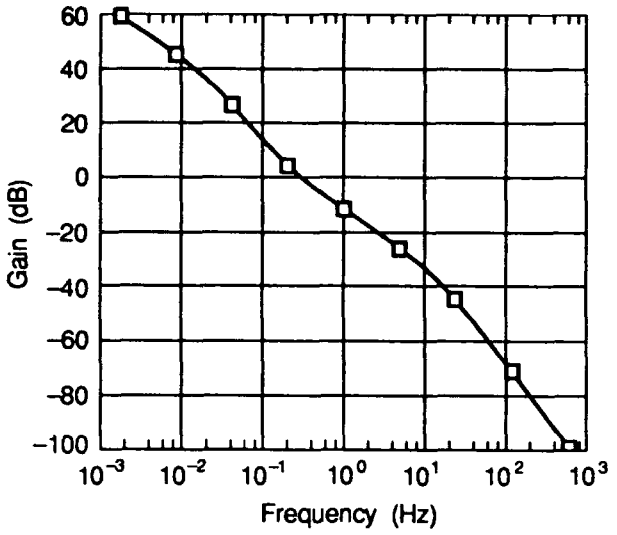


Figure 7. 4.096 MHz Synthesizer Open Loop Gain.

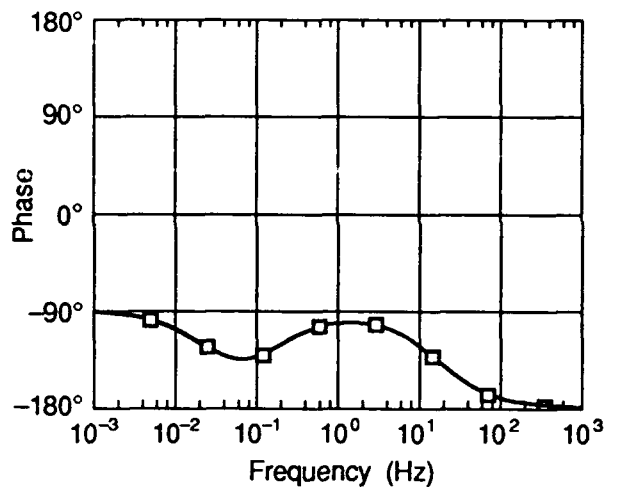


Figure 8. 4.096 MHz Synthesizer Open Loop Phase.

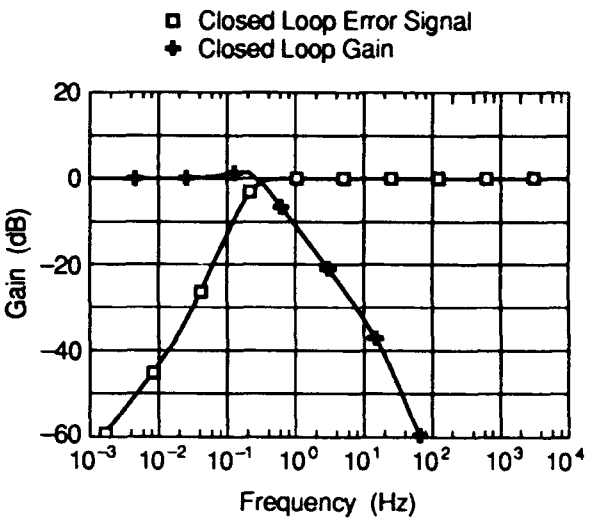


Figure 9. 4.096 MHz Synthesizer Closed Loop Response.

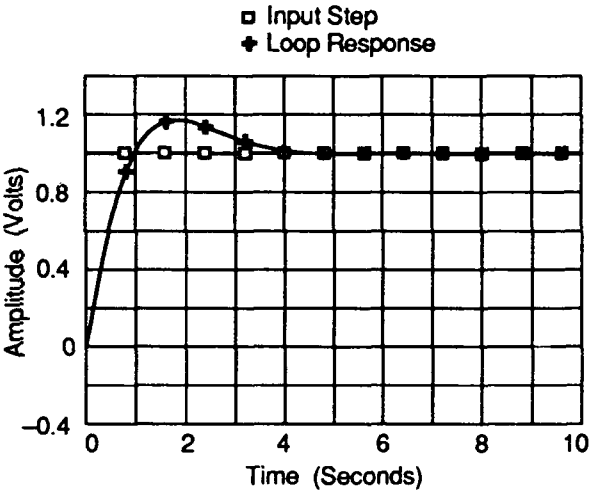


Figure 10. 4.096 MHz Synthesizer Step Response.

filtering than the 4.096 MHz loop but is clearly stable. The closed loop response is shown in Figure 13. Note that the error signal hump at 3 Hz coincides with the hump in the phase noise response shown earlier in Figure 1. The time domain step response is shown in Figure 14, and the damping factor was determined to be 0.56 with a natural frequency of 4.8 Hz.

Acknowledgments

The author gratefully acknowledges the contributions of J. Cloeren, J. Norton, P. Gilbert (deceased), and F. Wall (National Institute of Standards and Technology) in developing the synthesizer design and in fabricating the qualified units.

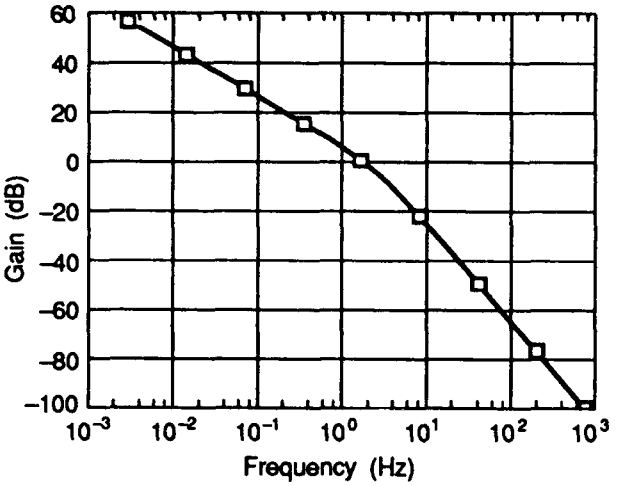


Figure 11. 19.056391 MHz Synthesizer Open Loop Gain.

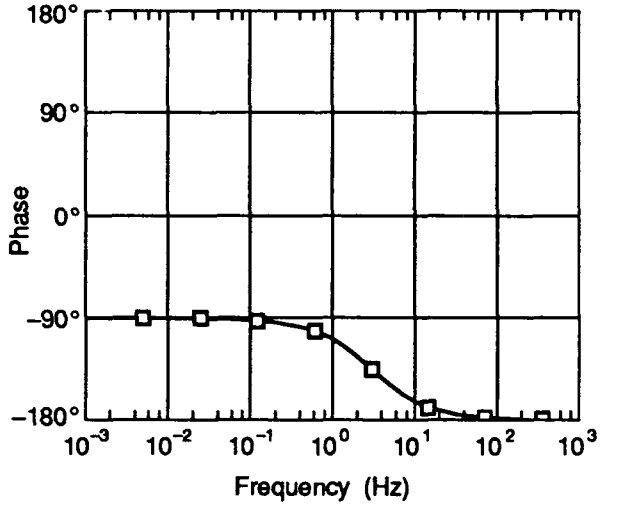


Figure 12. 19.056391 MHz Synthesizer Open Loop Phase.

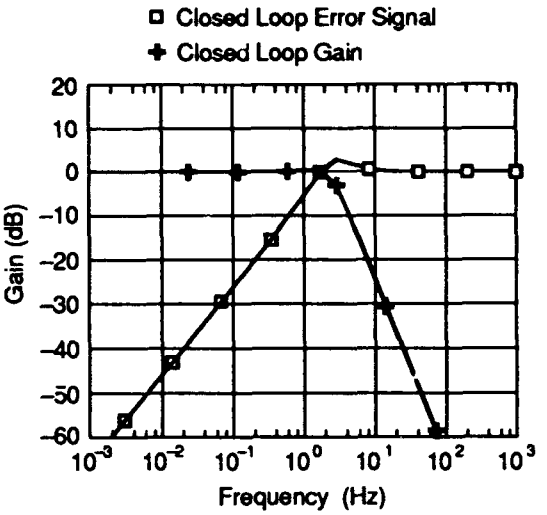


Figure 13. 19.056391 MHz Synthesizer Closed Loop Response.

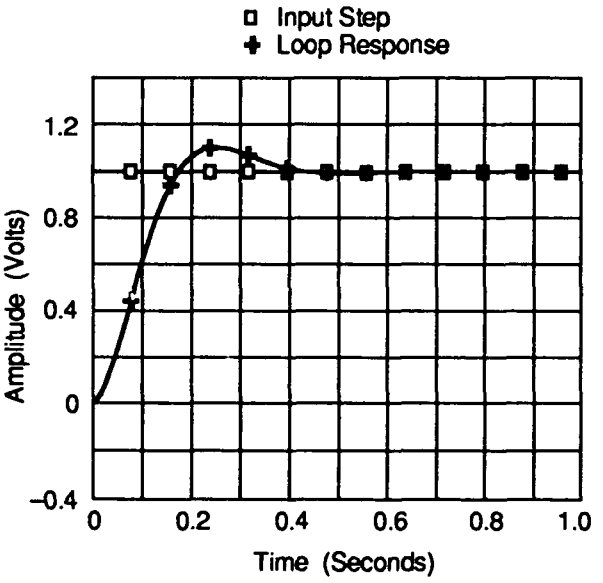


Figure 14. 19.056391 MHz Synthesizer Step Response.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH-STABILITY QUARTZ OSCILLATORS ON INTERNALLY-HEATED QUARTZ RESONATORS WITH AT AND SC CUTS

Ya. L. Vorokhovsky and B. G. Drakhlis

Morion Mfg. Co.
13a Kima Avenue, Leningrad 199155, USSR

Abstract: Precision frequency control requirements for future digital communication and position location systems require improved ovenized quartz oscillator performance in a number of aspects. An effective way to fulfil the requirements for frequency stability,power consumption, size,warm-up time,mechanical effect resistance etc. is to use internally-heated quartz resonators (IHQR).These devices in one construction functionally and technologically combine a quartz resonator itself and an oven.

A number of industrial types of quartz oscillators on IHQR with AT and SC cut piezoelements have been developed. They have frequency stability better then $\pm(2.5...5.0)\times10^{-8}$ in temperature range from -60 to +75°C, short - time stability for 1s up to 1×10^{-11},the level of phase noise on Fourier frequencies 1Hz and 10Hz -115 dB/Hz and -134 dB/Hz respectively, power consumtion less then 0.65W (at -60°C), warm-up time at -60°C less than 2...10 min to $\pm1\times10^{-7}$ of the final value, volume 55...100 cm^3.

According to the universal tendency in the development of electronics more stringent requirements for size,power consumption,and warm-up time of ovenized quartz oscillators are raised.The fulfilment of all the complex of these requirements is possible by means of decrease in the volume of the oven and improvement of its thermoinsulation,which lead to limitations on size of oscillator's oven.

The size of the oven will be minimal if minimal number of oscillator's elements are ovenized,the piezoelement only in limit. It is obvious that the best thermoinsulation is a vacuum gap.

Only two ways of keeping the piezoelement at a constant temperature are known. The first consists in putting the film heater directly on the piezoelement [1]. In this case temperature sensors are also placed (or evaporated) on the piezoelement. The second way is more traditional. The piezoelement is placed inside a metal chamber. A heater and (when it is necessary) a temperature sensor are placed on the base of this chamber. The metal chamber is placed in outer case (cabinet), and the standard package of a quartz resonator (glass or metal) is the most applicable in practice [2].Such devices functionally and technologically combining in the one construction quartz resonator and an oven are called "internally heated quartz resonators" (IHQR).

Further we shall discuss only internally heated quartz resonators (IHQR) made in the second way, using chip thermoresistors from materials with great positive temperature resistance coefficient (positive TC resistors) as heaters and temperature sensors [3]. The basic construction of the vacuum internally heated quartz resonator with positive TC thermoresistors used as heaters and temperature sensors is shown in Fig. 1. [4] In such IHQR in addition to the piezoelement one can easily ovenize the varicap diode of the frequency adjuster (it isn't shown in the figure).This doesn't enlarge the volume of the oven since it uses a chip varicap diode. At the same time placing of the varicap diode in the oven makes it possible to eliminate the effect of varicap temperature instability on frequency temperature stability of oscillator which can be significant if the varicap diode isn't ovenized.

Use of IHQR allows to decrease sharply power consumption, frequency transient time, size and weight of ovenized quartz oscillators, and to increase their mechanical resistance. But the development of IHQR leads to many problems,their potential possibilities can't be realised without solution of these problems.

On the one hand these problems are connected with the fact that in oscillators on IHQR the circuit isn't ovenized (except the varicap diode) and its frequency instability can be the same as IHQR instability or even more. This is an extremely serios problem in IHQR with SC-cut piezoelements, which have a strong additional undesirable resonance near the main resonance.

In this case it is necessary to use frequency selective circuits to provide stable operation on the main resonance.It makes circuits more complicated and can increase their instability.

On the other hand problems in development of oscillators on IHQR are connected with a strong thermal coupling between the heater and the piezoelement.It makes the reguirements to the oven control circuit sufficiently more strict than in case of traditional external ovens. Unsatisfactory design of the oven control circuit can result in sharp deterioration of short-time frequency stability, increase of oscillator's

CH2965-2/91/0000-447 $1.00 © 1991 IEEE

intrinsic noise,and in long transients due to voltage supply hops.

Problems mentioned above have been successfully solved [5,6]. Now a number of industrial types of quartz oscillators have been developed. Their characteristics are shown in table 1 and figures 2-4 [7]. All oscillators are made according to the basic structural diagram of a oscillator on IHQR shown in Fig.5. The basic construction of the IHQR is shown in Fig.1. The oscillators mentioned in the table differ in overtones used and piezoelement cut, circuitry of different units of the basic structural diagram, voltage supply, construction. Some oscillators include additional units, such as frequency multipliers and formers for shaping of pulse signals of TTL - type on the output of the oscillators, which are not included in the basic structural diagram.

From cited data it is obvious that quartz oscillators on IHQR allow to realize the whole complex of parameters including the short - time frequency stability and the spectral density of phase noise, which is in line with the best parameters of quartz oscillators with the same size,power consumption,temperature stability designed with help of other modern methods.

In the further analysis of ways of design of precision quartz oscillators on IHQR in connection with the basic tendency to integration of functionally completed units first of all is raised the possibility of ovenizing of all circuitry of oscillator, or only its part made as a specialized chip, in the volume of the IHQR together with the piezoelement and the varicap diode. The advantages of such a step are evident both from point of view of miniaturization and of further increase of frequency stability. The latter circumstance would be the result of elimination of the influence of the environment temperature on parameters of the circuits of oscillator by keeping them at constant temperature. These integral oscillators with circuits made as hybrid chips and ovenized in IHQR together with piezoelement are described in [8].

However besides evident advantages this approach has some disadvantages, the main one of which is the effect of the power scattered by the ovenized chip into IHQR.

Let $P_п$ be the power controlled by the oven control circuit of IHQR; $T_в$ - the temperature of the oven; $P_{сt}$ - power uncontrolled by the oven control circuit scattered by the chip ovenized in the IHQR; R_t -thermal resistance of the IHQR construction to heat irradiation in the environment. Then if the temperature of the environment is $T_о$ the thermal balance equation is:

$$P_п + P_{сt} = (T_в - T_о) / R_t \qquad (1)$$

or

$$P_п = [T_в - (T_о + P_{сt} \cdot R_t)] / R_t \qquad (2)$$

From (2) follows that an instabilizing effect of $P_{сt}$ is present. It is equivalent to the increase of temperature range of the environment by $P_{сt} \cdot R_t$ [9].

As it also follows from (2) it is possible to reduce the instabilizing effect of power scattered by the chip ovenized in the IHQR by the following ways: (i) to reduce R_t, (ii) to reduce $P_{сt}$; (iii) to increase oven temperature $T_в$. The first way is impossible because it increases power consumption of IHQR depriving it of one of its main advantages. The second way is admissible in principle but its possibilities are usually limited by the value of 15-30 mW. For modern IHQR with R_t about 600°C/W this is equivalent to increase of the environment temperature by 10 - 20°C approximately. The third way is admissible in principle too but in general it results in deterioration of long-time frequency stability, though it is possible to realize the stability of 1×10^{-9}/day at $T_в$ =90°C [10] and even at 84°C [11].

In fact for neutralization of inner overheating due to the heat irradiation of the ovenized chip it is necessary to provide the fulfilment of the condition $T_в - (T_{о\ max} + P_{сt} \cdot R_t) > 5°C$ at highest environment temperature $T_{о\ max}$. So in [11] due to the uncontrolled power of the ovenized chip they had to increase the oven temperature to 94°C though the upper temperature of the environment was $T_{о\ max}$ = +75°.

In addition to the mentioned problems ovenizing of the chips in the closed vacuum space of IHQR leads to some serious technological problems. They are connected with providing vacuum hygiene both of the chips themselves and their fastening to the thermodistributing elements of the construction. Lack of vacuum hygiene can result in deterioration of vacuum in the volume of the IHQR, that leads to increase of its power consumption and deterioration of the frequency stability.

Let us consider the possibilities of quartz oscillators on IHQR without ovenizing their circuits (except the varicap diode).

The analysis shows that if the varicap diode is ovenized with the piezoelement of IHQR as it is easily realized in the basic construction of IHQR (Fig.1) and vacuum inside the IHQR isn't deteriorated then corresponding circuitry design and choice of elements make it possible to bring temperature frequency stability of the unovenized rest of the circuit in a wide temperature range from -60 to +75°C up to 1×10^{-8}. The achievable temperature stability of IHQR itself in the same temperature range is also about 1×10^{-8}.

Conclusions

Thus oscillators on IHQR allow to cover the temperature frequency stability range from 2×10^{-8} to 2×10^{-7}. Starting with stabilites about 2×10^{-7} and worse it is possible to use thermocompensated oscillators. However, situations are possible when at such stabilites oscillators on IHQR can be useful due to their simplicity in spite of bigger power consumption than in case of thermocompensated oscillators.

Table 1.

Characteristics of quartz oscillators on internally heated quartz resonators (IHQR)

Parameters	1	2	3	4
Frequency, MHz	4.9 - 5.2	9.8-10.4	5.0	5.0
Output signal shape	sin	sin	sin	TTL
Temperature frequency instability, less then, 10^{-6}:				
temperature from -10 to +55°C	±2.5	±2.5	±1.0	±1.0
temperature from -60 to +70(75)°C	±5.0	±5.0	±2.5	±2.5
Frequency instability due to voltage supply hops ±10%, less then, 10^{-9}	±5.0	±5.0	±2.5	±5.0
Short time frequency stability for 1 sec, less then, 10^{-11}	5.0	5.0	1.0	1.0
Spectral density of phase noise dB/Hz, at frequency offset:	fig.4, curve 1	fig.4, curve 2	fig.4, curve 3	-
1 Hz	-90	-82	-115	-
10 Hz	-126	-120	-134	-
100 Hz	-138	-133	-141	-
Power consumtion at -60°C, less then, W	0.65	0.65	0.65	0.65
Warm-up time at -60°C, less then, min				
with accuracy $\pm 1 \cdot 10^{-7}$	10	10	3	2
with accuracy $\pm 5 \cdot 10^{-6}$	12	12	5	4
Size, mm	63x53x29.5	63x53x 29.5	63x53x 29.5	26.5x26.5 x81.5
The basic IHQR:				
cut	AT	AT	SC	SC
nominal frequency/operating overtone	5.0/5	5.0/5	5.0/3	5.0/3

References

[1] F.G. Tinta, A.S. Matistic, and G.A. Legasse, "The Direct Temperature Control of Quartz Crystals in Evacuated Enclosures," in Proc. of the 24th Annual Symposium on Frequency Control, USA, 1970, pp. 157-167.

[2] Я.Л.Вороховский, В.Б.Грузинченко, А.В.Золотов, Н.Н.Лепешкин, "Интегральные устройства пьезоэлектроники с внутренним термостатированием," Электронная техника. Сер. радиодетали и радиокомпоненты, 1987, вып.3 (68), с.52 - 56.

[3] Г.Н. Текстер-Проскурякова, И.Т. Шеетель "Автостабилизирующие позисторы," Электронная промышленность, 1975, N 7, с.64-65.

[4] А.А.Волков, Я.Л.Вороховский, И.Г.Петросян, И.С.Тровин,"Кварцевые резонаторы-термостаты," Электронная промышленность, 1984,вып.1 (129), с.64 - 66.

[5] Я.Л. Вороховский, Б.Г. Дрехлис, "Выбор и построение терморегулятора для резонатора-термостата прецизионного кварцевого генератора," Электронная техника. Сер. радиодетали и радиокомпоненты, 1987, вып.2 (67), с.67 - 70.

[6] А.А.Волков, Я.Л.Вороховский, Б.Г.Дрехлис, С.И.Липко, И.Г.Петросян, "Построение высокостабильных экономичных малогабаритных кварцевых генераторов с малым временем установления частоты на резонаторах-термостатах," Электронная техника. Сер. радиодетали и радиокомпоненты, 1987, вып.4 (69), с.59 - 64.

[7] Я.Л. Вороховский, Б.Г. Дрехлис , К.Г. Коженкин, "Прецизионные малогабаритные кварцевые генераторы с малым временем установления частоты на резонаторах-термостатах срезов АТ и SC," Электронная техника. Сер. радиодетали и радиокомпоненты, 1989, вып.3 (76), с.49 - 53.

[8] H.W. Jackson, "Tactical Miniature Crystal Oscillator," in Proc. of the 34th Annual Symposium on Frequency Control, USA, 1980, pp.449-456.

[9] А.А. Волков, Я.Л. Вороховский, В.Б. Грузиненко, "Интегральное пьезоэлектрическое устройство - генератор-термостат," Электронная техника. Сер. радиодетали и радиокомпоненты, 1978, вып.5 (30), с.58 - 61.

[10] В.Г.Андросова, В.Н.Банков,А.Н.Микидзи и др. Под ред. П.Г.Позднякова. Справочник по кварцевым резонаторам. Москва: Связь, 1978, -288 с.

[11] H.M. Greenhouse, R.L. McGill, and D.P. Clark, "A Fast Warm-up Quartz Crystal Oscillator," in Proc. of the 27th Annual Symposium on Frequency Control, USA, 1973, pp.199-217.

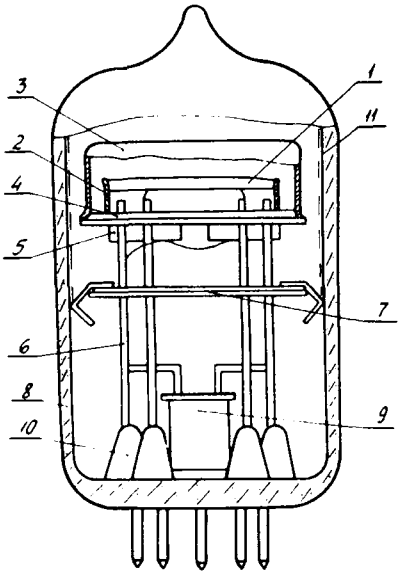


Fig. 1. The basic construction of vacuum IHQR with positive TC thermoresistors as the heaters and the temperature sensor: 1 - the piezoelement; 2 - the crystal holder; 3 - the cover - the thermal screen; 4 - the base of the crystal holder; 5 - the heaters and the temperature sensor on the positive TC thermoresistors; 6 - the holder; 7 - the intermediate fixing base with the centring springs; 8 - the tube; 9 - the solid-state getter; 10 - the base of the case with the terminals; 11 - the thermoreflective coating.

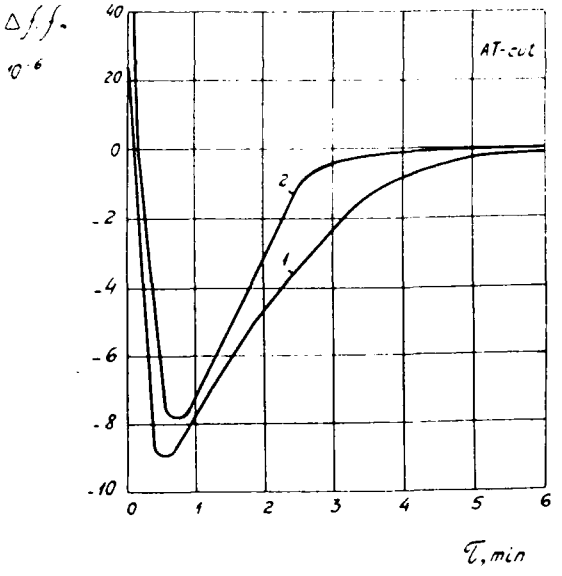


Fig. 2a. Typical characteristics of frequency shift of quartz oscillators on IHQR with AT- and SC -cut piezoelements during warm-up time after turn on at environment temperature -60°C (1) and at normal conditions (2):

AT-cut (5 MHz at the 5-th overtone).

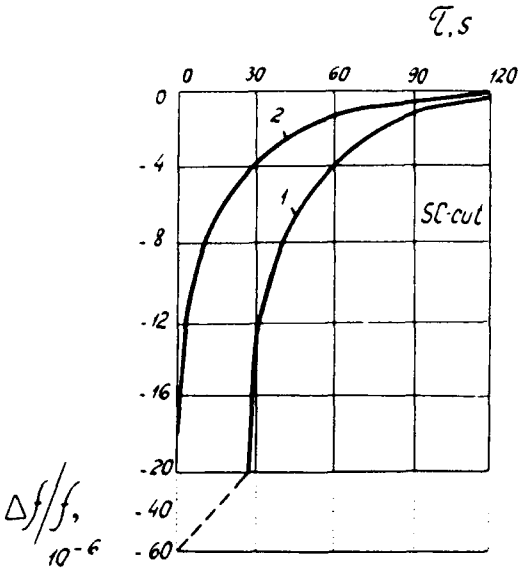


Fig. 2b. Typical characteristics of frequency shift of quartz oscillators on IHQR with AT- and SC -cut piezoelements during warm-up time after turn on at environment temperature -60°C (1) and at normal conditions (2):

SC-cut (5 MHz at the 3-d overtone).

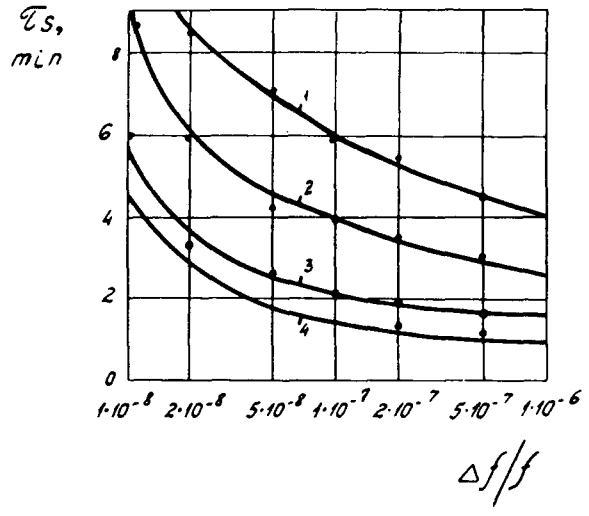


Fig. 3. Typical changes of warm-up time of oscillators on IHQR with AT- and SC -cuts at warm-up versus the required frequency accuracy (df/f):
1 - AT-cut, t = -60°C; 2 - AT-cut, t = +25°C;
3 - SC-cut, t = -60°C; 4 - SC-cut, t = +25°C.

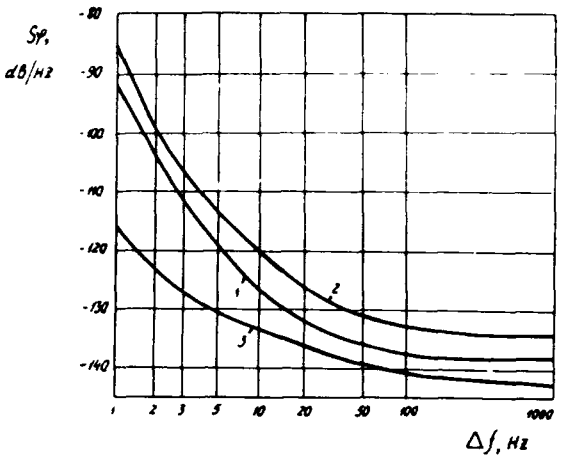


Fig. 4. Phase noise spectral density (S_φ) of quartz oscillators on IHQR versus frequency offset from the carrier frequency (Δf): 1,2 - IHQR on AT-cut, 5 MHz at the 5-th overtone: 1 - f_{out}= 5 MHz, 2 - f_{out}= 10 MHz; 3 - IHQR on SC-cut at the 3-d overtone (f_{out}=5 MHz).

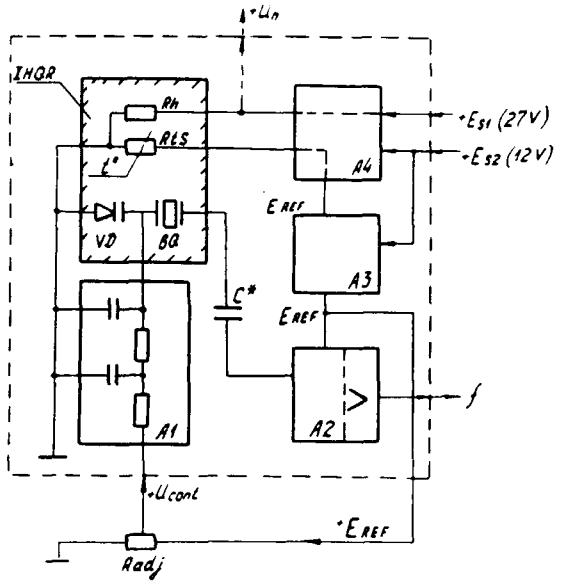


Fig. 5. The basic structural diagram of a quartz oscillator on IHQR: A1 - the elements of the electronic frequency adjuster; A2 - the active part of the oscillator (with the buffer amplifier); A3 - the reference voltage source; A4 - the proportional oven control circuit; E_{s1} - the heater voltage supply; E_{s2} - voltage supply of the rest of the oscillator.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH-TEMPERATURE SUPERCONDUCTING RESONATORS

J. N. Hollenhorst, R. C. Taber, and L. S. Cutler
Hewlett-Packard Laboratories, 3500 Deer Creek Road, Palo Alto, CA 94304

T. L. Bagwell
Hewlett-Packard, 1412 Fountaingrove Parkway, Santa Rosa, CA 95405

N. Newman
Conductus, Inc., 969 West Maude Avenue, Sunnyvale, CA 94086

Abstract

This paper will report preliminary measurements on high-T_c superconducting resonators and discuss why they are attractive candidates for incorporation in low-noise oscillators. We will review some of the important contributions to oscillator noise and show how they depend on the resonator parameters. A preliminary $YBa_2Cu_3O_7/LaAlO_3$ resonator with a Q of 9×10^4 at 6.9 GHz and 7×10^4 at 3.5 GHz has been fabricated. The temperature sensitivity, power dependence and residual phase-noise will be discussed. An upper-limit on the coefficient of the $1/f$ component of fractional-frequency fluctuations has been measured to be -204 dB at 60 K. This is, to our knowledge, the best value reported for any high-T_c superconducting resonator.

Introduction

As long ago as 1971, a superconducting Niobium microwave cavity with a Q of 5×10^{11} was reported.[1] Low-temperature superconducting-cavity-stabilized-oscillators (SCSO's) with outstanding short-term frequency stability have been constructed. Stein and Turneaure[2] demonstrated an X-band oscillator stabilized by a cavity with a Q of 10^{10} and obtained an Allan variance of $\sigma_y(\tau) = 6\times10^{-16}$. Unfortunately, these results were attained only at temperatures below 2 K. Recently, the advent of superconductors with transition temperatures above 80 K has raised the prospect of similar performance levels at more easily attained temperatures. It is now possible to conceive of a compact SCSO contained within a rack-mounted instrument incorporating a small and inexpensive closed-cycle cryocooler.

To understand the benefits of superconducting resonators, and more generally, of low-temperature operation, it is useful to review some of the important sources of oscillator phase-noise. If an oscillator has negligible amplitude modulation, the output may be written as $V(t) = V_o\sin(2\pi f_o t + \phi(t))$, where f_o is the nominal frequency and $\phi(t)$ is the instantaneous phase deviation. It is useful to introduce the fractional frequency deviation, $y(t) \equiv (d\phi/dt)/2\pi f_o$. The one-sided spectral densities $S_y(f)$ and $S_\phi(f)$ of $y(t)$ and $\phi(t)$ are related to each other and to the single-sideband noise power $L(f)$ by the equations:[3]

$$S_y(f) = \left[\frac{f}{f_o}\right]^2 S_\phi(f) = 2\left[\frac{f}{f_o}\right]^2 L(f). \qquad (1)$$

Here f is the frequency offset from the nominal carrier frequency f_o. Strictly-speaking, the second equality holds only in the limit of small phase deviations. While $L(f)$ is more commonly reported, $S_y(f)$ has the advantage that it is invariant under noiseless frequency multiplication and thus provides a useful figure-of-merit for comparison of oscillators operating at different frequencies.

An approximate expression for $S_y(f)$ that incorporates four of the most important noise sources is given by:[4]

$$S_y(f) = \frac{2kT_1}{P_s}\left[\frac{f}{f_o}\right]^2 + \frac{kT_2}{2P_sQ_l^2} + \frac{a}{4Q_l^2 f} + \frac{b}{f}. \qquad (2)$$

Here, k is Boltzmann's constant, Q_l is the loaded Q of the resonator, P_s is the total power delivered to the loaded resonator, and a and b are constants that specify the magnitude of the $1/f$ phase-noise in the amplifier and the $1/f$ frequency-noise in the resonator, respectively. T_1 and T_2 are effective-

CH2965-2/91/0000-452 $1.00 © 1991 IEEE

noise-temperatures that account for the amplifier and resonator thermal noise near the carrier and far from the carrier, respectively; their relation to the fundamental noise sources within the amplifier and resonator will depend on the details of the oscillator sustaining circuitry.

At frequencies far from the carrier, $S_y(f)$ will be dominated by the term proportional to f^2. To minimize this, it is desirable to lower the thermal noise kT_1 in the resonator and amplifier, to operate at a high power level P_s and to choose a high resonator frequency f_o. Operation at low temperature should allow lower values of T_1 and T_2 to be obtained. It may also be possible to lower the amplifier $1/f$ noise by operating at low temperatures, although this is not guaranteed. The maximum power level that a superconducting resonator can sustain will set a limit on P_s and thus is important in determining the far-from-carrier noise.

Within the loaded-bandwidth of the resonator, the last three terms in equation 2 become important. It is apparent that a high resonator Q_l is desired to minimize the two middle terms. As we will see, the low surface-loss in high-temperature superconducting films leads us to expect that very high values of Q_l will be possible. In addition, it is known that the loss-tangent in certain dielectric materials is strongly temperature dependent and can be remarkably small at low temperatures. For example, measurements have shown the loss-tangent of sapphire to be as low as 1.5×10^{-8} at 78 K.[5]

For sufficiently high Q_l, the close-to-carrier noise is expected to be dominated by frequency fluctuations of the resonator itself; these will be unaffected by the resonator Q_l. In other technologies it is often found that the resonator frequency-fluctuations obey a $1/f$ dependence. In the absence of experimental data, it is reasonable to hypothesize that the same will be true for superconducting cavity resonators; we have thus included a $1/f$ term as the last expression in equation 2. An important result of this work will be to set an upper limit on the size of this noise source.

An optimal resonator design would entail efforts to control each of the terms in equation 2 as well as consideration of tradeoffs among them. For example, going to higher resonant frequencies will reduce the far-from-carrier noise-floor but at the expense of the close-in performance since the Q is usually a decreasing function of resonant frequency. For the same Q value, electromagnetic resonators can typically be operated at much higher frequencies; thus they are expected to have better far-from-carrier behavior. As another example, efforts to increase Q_l often lead to reductions in P_s; if the relationship between these is known, equation 2 provides a simple basis for selecting an optimal compromise. The last term, the resonator $1/f$ noise, is almost invariably the least amenable to this type of analysis. $1/f$ behavior is often related to defects and other phenomena that are poorly understood and difficult to predict. This unpredictability makes it especially interesting to obtain an experimental value for the resonator $1/f$ coefficient in equation 2.

High-Temperature Superconducting Materials

Since their discovery in 1986, remarkable progress has been made in the development of high transition temperature, oxide-superconductors. For applications to cavity resonators, the most important property is the microwave surface loss and its dependence on magnetic field at the temperature of operation. The microwave loss is often characterized by the surface resistance R_s.[6] If J is the root-mean-square sheet current density, the power dissipation per unit area is given by R_sJ^2. Many laboratories are now routinely able to deposit thin-films of the 90 K superconductor $YBa_2Cu_3O_7$ with excellent performance at 10 GHz. Figure 1 shows some typical R_s measurements performed in our laboratory[7] on films grown by several different *in situ* processes.[8,9,10,11] For comparison, estimated values for the low-temperature surface resistance of copper are also shown. At all temperatures below 80 K, the surface resistance is significantly smaller than that of copper. The improvement is by more than one order of magnitude at 77 K and two orders of magnitude at 20 K. As one goes to lower frequencies, the improvement over conventional metals becomes even more dramatic. Near the transition temperature, R_s typically follows an ω^2 dependence while for normal metals a $\omega^{1/2}$ dependence is obeyed.[12] At 1 GHz, the R_s of $YBa_2Cu_3O_7$ films should be approximately 100 times lower than the 10 GHz values shown in figure 1. In the near vicinity of the transition temperature, the measurements are close to the fundamental limit that is expected due to the dissipative flow of normal currents in the films; below 70 K all samples show the presence of a residual resistance whose origin is not well understood but probably not fundamental. Thus, there is reason to expect continuing improvement in R_s in the future.

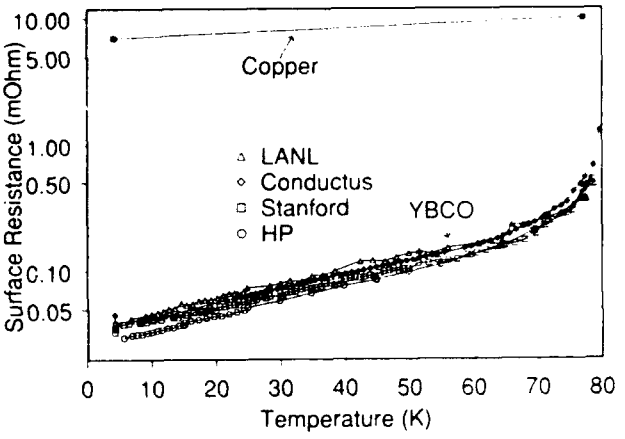


Figure 1: Surface resistance versus temperature. Measured values of the surface resistance R_s at 10 GHz are shown for four pairs of YBCO films. For comparison, estimated values are shown for Copper at 4.2 K and 77 K.

$YBa_2Cu_3O_7$, like most of the oxide-superconductors, is structurally asymmetric and highly anisotropic in its superconducting transport properties. To obtain the excellent performance shown in figure 1, it is essential to grow nearly epitaxial films on single crystal substrates. In fact, the surface resistance has been shown to correlate strongly with the percentage of misaligned grains in a highly-oriented film.[13] This requirement severely restricts the choice of substrate materials, as well as restricting the resonator designer (for the time being, at least) to planar superconducting films.

Cavity Design

An important resonator parameter appearing in equation 2 is the loaded quality factor Q_l. The best oscillator performance is usually obtained near critical coupling where Q_l is one-half of the unloaded quality factor Q given by:

$$Q^{-1} = Q_c^{-1} + Q_d^{-1} + Q_r^{-1}. \tag{3}$$

Here, Q_c^{-1} is the conductor loss due to finite R_s, Q_d^{-1} is the dielectric loss due to finite loss-tangent and Q_r^{-1} is the loss due to radiation that escapes the cavity. To gain some insight, we consider the case of a cavity composed of a uniform dielectric covered with a conductor. The conductor-limited Q is then given by:[6,14]

$$Q_c = 2\pi \frac{Z_m}{R_s} \frac{V}{S\lambda} G, \tag{4}$$

where $Z_m \equiv (\mu/\epsilon)^{1/2}$ is the impedance of the cavity medium, V/S is the volume-to-surface ratio of the cavity, λ is the free-space wavelength and G is a geometrical factor that is near unity and depends only weakly on the shape of the cavity and the mode of oscillation. For this simple example, Q_d is just the reciprocal of the loss-tangent of the dielectric. The geometrical factor G is unity for a cube, a square plate, and for a circular cylinder operating in the lowest mode. From the above expression it is clear that, when the conductor loss dominates, we should maximize the volume-to-surface ratio and minimize R_s. For a given oscillation frequency, the Q will be highest if we build a large cavity and operate in a high-order mode. For an empty cavity, $Z_m = 377\ \Omega$, and, for a cubical resonator in the lowest mode, the conductor loss gives a Q_c of:

$$Q_c = \frac{\sqrt{2}\pi}{6} \frac{377\ \Omega}{R_s} = \frac{279\ \Omega}{R_s}. \tag{5}$$

Referring to figure 1, we see that an R_s of 100 $\mu\Omega$ can be obtained at temperatures as high as 50 K. Substituting into equation 5, we get $Q_c \sim 3\times10^6$. If we were to operate at 1 GHz, we could expect a Q_c of about 3×10^8 because of the ω^2 dependence of R_s. Even higher values of Q should be achievable by operating at lower temperatures, using higher-order resonant modes and by further improvements in R_s.

Since Q's of 10^8 have been obtained in low-temperature sapphire, it is natural to consider a superconductor/dielectric hybrid resonator. In conventional dielectric resonators,[15] the fields must be kept away from any metal surfaces to minimize conductor loss. Often, "whispering gallery" or other high-order modes are used to confine the fields to the dielectric; in addition, the walls of the metal enclosure must be kept a substantial distance from the dielectric. This leads to resonators that are rather bulky even for operation near 10 GHz. By using superconducting films in conjunction with ultralow-loss dielectrics we can hope to achieve high performance resonators in a compact, mechanically stable package. As the technology matures and R_s is reduced further, it may be desirable to eliminate the dielectric or find materials with even lower loss.

Resonator Measurements

To elucidate some of the major issues in development of low-noise cryogenic oscillators, we have fabricated some superconductor/dielectric hybrid resonators. For these preliminary experiments, we have chosen a simple structure consisting of a rectangular dielectric slab coated on two sides with $YBa_2Cu_3O_7$ films. $LaAlO_3$ was chosen as the dielectric since a process had previously been developed for deposition of high quality films on this material. The relatively high dielectric constant ($\epsilon_r \sim 25$) is also convenient for these experiments. The $YBa_2Cu_3O_7$ films were deposited at Conductus, Inc. by an off-axis *in situ* sputtering technique, described in detail elsewhere.[8] The length of the slab was chosen as 1 cm to give a fundamental resonance at about 3 GHz. The width of the slab was chosen to be 0.5 cm. The thickness of the slab was chosen to be 0.05 cm due to availability of substrate material. Also, since four sides of the slab are not coated with superconducting material, one must minimize the dissipation due to fields that radiate beyond the dielectric slab. By using a thin slab with a high dielectric constant and enclosing the resonator in a metal box, the loading of the Q due to this effect is minimized. The high surface-to-volume ratio of the resulting 1cm × 0.5cm × 0.05cm cavity makes it more sensitive to the properties of the superconducting films than an optimized design would be.

Figure 2 shows the measured unloaded-Q as a function of temperature for the two lowest modes at 3.5 and 6.9 GHz. Over most of the temperature range shown, the loss is believed to be dominated by dielectric dissipation in the $LaAlO_3$. Only near the superconducting transition is the superconductor-loss believed to be important. At the lowest temperatures, some of the loss may be from dissipation in the metal enclosure surrounding the resonator. The highest Q values were obtained at the lowest temperature of ~5 K; they were 9×10^4 at 6.9 GHz and 7×10^4 at 3.5 GHz. While these values are well short of what should be possible, they are still among the best Q's yet reported with high-T_c material.

In order to perform sensitive measurements of the residual phase-noise in the resonator, it is important that the coupling to the resonator be near the critical value. Our measurements were typically made in transmission mode with coupling provided by shorted microstrip lines on opposite sides of the slab acting as magnetic loops. By changing the angle of the slab relative to the loops, the coupling could be adjusted

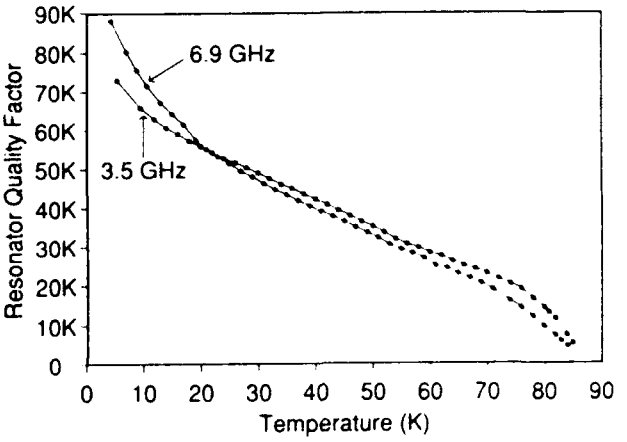


Figure 2: Unloaded resonator Q versus temperature. Data is shown for the two lowest modes at 3.5 GHz and 6.9 GHz.

prior to each cool-down cycle.

Recall from equation 2 that the far-from-carrier noise depends on the power level P_s in the resonator. The product of P_s and Q_l is proportional to the energy stored within the resonator; this, in turn, is proportional to the product of the resonator volume and the mean square magnetic field within the cavity. For superconducting devices, the P_sQ_l product will be limited by the maximum magnetic field that the superconductor can support before the dissipation becomes excessive. Figure 3 shows the loaded Q at 77 K as a function of the peak magnetic field in the center of the cavity. It is seen that a substantial degradation has occurred by the time the field reaches 4 gauss. A partial explanation for this is that the field near the edge of the resonator is much higher due to field crowding; this occurs as the magnetic flux lines expand into the region outside the dielectric slab. An alternative design approach, free from this problem, is to use one of the azimuthally symmetric modes of a disk resonator. These have the property that the magnetic field lines are parallel to the edge of the disk. Since there can be no radial current at the periphery, the magnetic field vanishes at the edge.

An important source of noise in practical resonators is the frequency shift due to temperature fluctuations. We have thus measured the temperature dependence of the lowest mode of our resonator; the results are shown in figure 4. As the temperature is raised, the magnetic penetration depth rises, leading to a decrease in resonant frequency as the transition

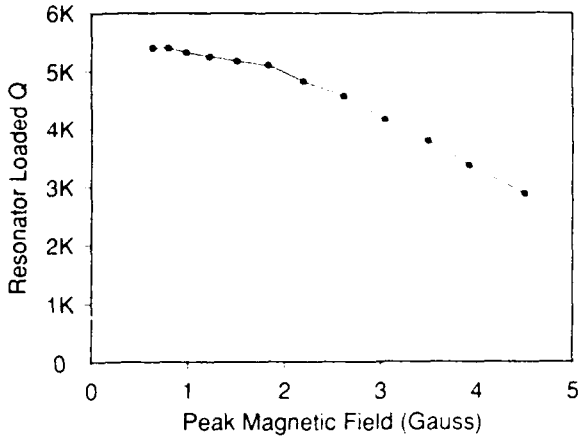


Figure 3: Loaded resonator Q versus peak magnetic field. Q_l is plotted versus the estimated peak magnetic field value at the center of the cavity. At the periphery of the resonator, the field is considerably higher because of field crowding.

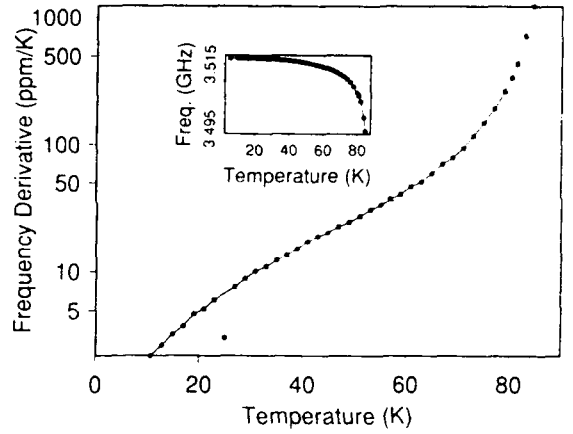


Figure 4: Resonator temperature coefficient. The inset shows the frequency of the fundamental mode as a function of temperature. The main plot shows the fractional frequency derivative $(1/f)(df/dT)$ in parts per million per Kelvin.

temperature is approached. The effect is exacerbated in our device by the thin (0.05 cm) spacing between the $YBa_2Cu_3O_7$ films. In spite of this, the temperature coefficient due to penetration depth effects is expected to be negligible below about 50 K. The residual coefficient of 2-30 ppm/K in the 10-50 K range is believed to be caused by differential thermal expansion and/or changes in the dielectric constant of $LaAlO_3$. In choosing a suitable dielectric material, these effects will be an important consideration.

The residual phase-noise was measured using an HP 3048A measurement system operated as a phase-bridge. The technique is similar to that described by Elliot and Bray.[16] The resonator is operated in transmission mode using a frequency synthesizer (HP 8340B) as the signal source. The device output is mixed with a 90° phase-shifted sample of the synthesizer output using a low-noise mixer. The resulting baseband signal is fourier analyzed in the 1-100 kHz range using a low-frequency signal analyzer (HP 3561). Figure 5 shows the mixer-limited measurement noise-floor determined by replacing the

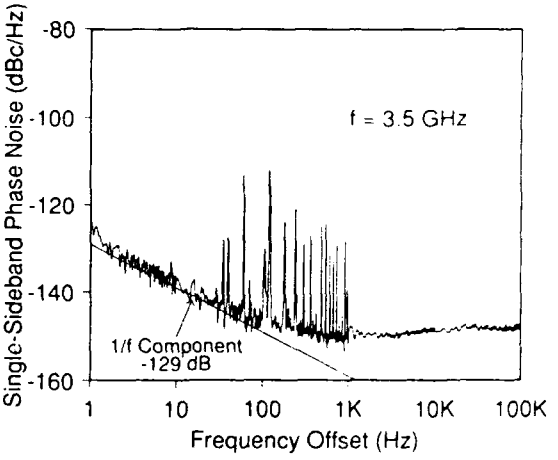


Figure 5: Phase noise measurement floor. The single-sideband noise power $L(f)$ is shown as a function of offset frequency for the phase-bridge operated without a resonator. The carrier frequency is 3.5 GHz.

resonator with a straight-through connection. The single-sideband noise power $L(f)$ is displayed versus offset frequency. A number of spurs are visible that are inherent in the measurement equipment. The $1/f$ component has an intercept of -129 dBc/Hz at 1 Hz offset. Thus, the coefficient of the $1/f$ component is -129 dBc.

Figure 6 shows the residual phase-noise with the resonator in the circuit and operating at 14 K with a Q_l of 7300. The output power from the resonator was 3.2 dBm. For comparison, a smoothed version of the noise-floor data is also shown. An approximately $1/f$ behavior is observed between 4 and 400 Hz. The coefficient of this $1/f$ component is -125 dBc, a value that is about 4 dB above the mixer noise-floor. Figure 7 shows the same measurement done at 60 K with a Q_l of 6400 and an output power of 2.2 dBm; in this

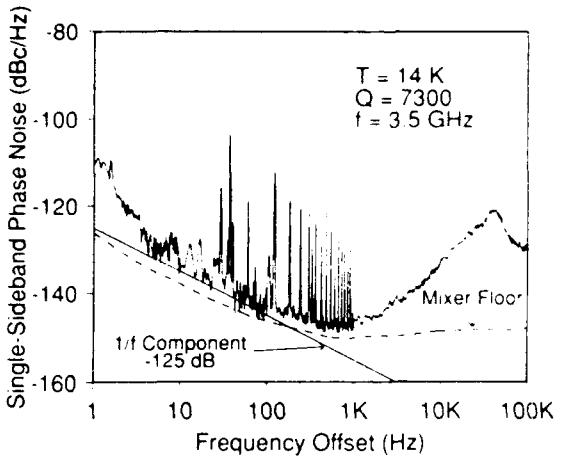


Figure 6: Residual phase noise at 14 K. The resonator center frequency is 3.5 GHz and the loaded-Q 7300. The power output from the resonator is 3.2 dBm. The broken curve is a smoothed representation of the mixer noise-floor. The resonator $1/f$ component is less than -125 dBc.

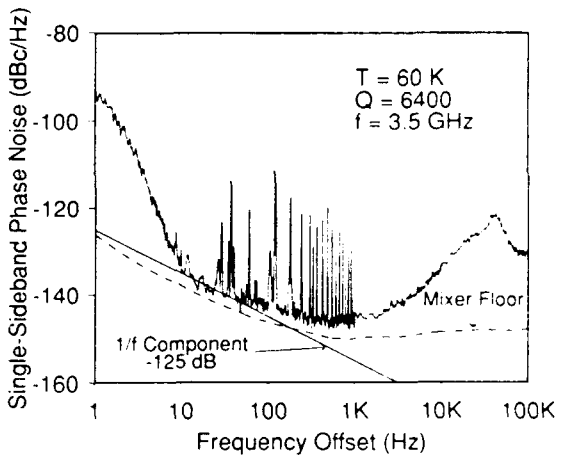


Figure 7: Residual phase noise at 60 K. The resonator center frequency is 3.5 GHz and the loaded-Q 6400. The power output from the resonator is 2.2 dBm. The broken curve is a smoothed representation of the mixer noise-floor. The resonator $1/f$ component is less than -125 dBc.

case, the $1/f$ term is also -125 dBc, very close to the result at 14 K. All but two of the spurs in figures 6 and 7 also appear in the noise-floor measurement and are, hence, instrument related. The origin of the other two spurs at 29 Hz and 37 Hz is not known and may also be an artifact of the measurement set-up. Vibration sensitivity was not measured but was small enough not to affect the phase-noise measurements.

Below 10 Hz there is excess noise apparent in both measurements. The excess at 1 Hz is about 15 dB at 14 K and 32 dB at 60 K, an increase of 17 dB. We believe that this additional noise is due to temperature fluctuations of the resonator. At 14 K the temperature coefficient is about 3 ppm/K, while at 60 K it is about 50 ppm/K. If the temperature fluctuations were comparable at the two temperatures, we would expect a difference of about 24 dB rather than the measured 17 dB. Since the thermal time constants are shorter at lower temperatures, it would not be surprising if the thermal fluctuations were somewhat greater at 14 K. The actual temperature fluctuations during the measurement were not known accurately, but did not exceed 10 mK. It is clear that, as in the case of acoustic devices, temperature control and minimization of temperature coefficient will be important design issues.

Since the results above 10 Hz are so close to the mixer noise-floor and very similar at 14 and 60 K, the question arises whether the measured $1/f$ noise is instrument related or truly representative of the resonator. In fact, the noise-floor evaluation shown if figure 5 is insensitive to phase fluctuations in the signal source. When the device under test is placed in the bridge, the measurement becomes more sensitive to the signal source noise. This effect gives rise to a large increase in the measured noise above 1 kHz that is clearly seen in figures 6 and 7. The effect below 1 kHz results from a tradeoff between the rapidly falling sensitivity to source noise and the rapidly rising level of this noise as the frequency is lowered. Our measurements of the source noise suggest that it can account for some, but not all, of the increased noise seen when the resonator is in the bridge. Given the uncertainties in determination of the true noise floor, we will take a conservative approach and interpret our results as upper-limits on the resonator $1/f$ noise; thus the coefficient of the $1/f$ component is less than -125 dBc for a 3.5 GHz resonator operating either at 60 K with a Q_l of 6400 or at 14 K with a Q_l of 7300.

To facilitate comparison with other results and to relate to equation 2, we now wish to compute the value of S_y that would be obtained in an oscillator. To do this, note that a deviation of Δf in the resonator center frequency would produce a phase deviation of $2Q_l\Delta f/f_o$. Thus,

$$S_y^r(f) = \frac{1}{2Q_l^2} L^r(f), \quad (5)$$

where the superscript (r) means that the quantity refers to the resonator. From the above, we obtain the result that the resonator $1/f$ coefficient (b in equation 2) is less than -204 dB. Table 1 compares this result to several other measurements that have been reported. The last column gives the $1/f$ coefficient (b) that we have calculated from the published results.

Table 1:Resonator Performance Comparison

Institution	Device	f_o GHz	Loaded Q	$1/f$ b
HP/Conductus (this work)	YBCO/$LaAlO_3$ Parallel Plate	3.5	7,300	-204
Avantek/STI	Thallium Microstrip	2.28	4,000	-174
Lincoln Labs	Nb/Si Stripline	1.0	80,000 250,000	-211 -206
Raytheon	SAW	0.5	8,000	-220

The Avantek/STI result[17] is for a TlBaCaCuO microstrip resonator operated as an oscillator using an FET amplifier. They also report results for a gold resonator, obtaining a value for S_y that scales roughly with Q_l^2. This is the expected behavior for amplifier-limited noise and suggests that their result should be interpreted as an upper-limit. The best Lincoln Labs result[18] is for a Nb/Si stripline device operated in an oscillator at 4.2 K. For Q's ranging between 23,000 and 250,000 they report S_y values ranging from -206 to -211 dBc/Hz at 1 Hz offset. To our knowledge, our own result is the best obtained to date with high-temperature superconducting films and gives an upper-limit within 7 dB of the Lincoln Labs low-temperature result. The Raytheon result[19] is for a state-of-the-art surface acoustic wave oscillator.

It is encouraging that, at this early stage, we have already approached measurement-limited performance.

Conclusions

We have fabricated $YBa_2Cu_3O_7/LaAlO_3$ parallel plate resonators with a small plate-spacing that should be sensitive to properties of the superconducting films. The residual frequency fluctuations were measured and an upper-limit of -204 dB placed on the resonator $1/f$ coefficient. Near the carrier, excess noise due to temperature fluctuations was observed; the residual temperature coefficient is believed to be due to properties of the $LaAlO_3$ dielectric. The resonator Q was seen to be strongly temperature dependent and is also believed to be dielectric-limited. The power dependence of the Q was seen to be a significant resonator design issue. We believe that the prospects are excellent for further improvements in superconducting/dielectric hybrid resonators particularly by exploiting higher quality dielectric materials.

The authors thank Brady Cole, Bob Bray and Bill Anklam for their help in the course of this work. In addition, we thank Paul Merchant for providing films for some of the R_s measurements.

References

[1] M. A. Allen, Z. D. Farkas, H. A. Hogg, E. W. Hoyt, and P. B. Wilson, IEEE Trans. Nucl. Sci., **NS-18**, 168 (1971).

[2] S. R. Stein and J. P. Turneaure, "Superconducting-Cavity Stabilized Oscillators with Improved Frequency Stability," Proc. IEEE, **63**, 1249 (1975).

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

[4] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators," 41st Annual Frequency Control Symposium, 99 (1987).

[5] V. B. Braginski, V. S. Ilchenko, and Kh. S. Bagdassarov, "Experimental Observation of Fundamental Microwave Absorption in High-Quality Dielectric Crystals," Phys. Lett. A, **120**, 300 (1987).

[6] S. Ramo, J. R. Whinnery and T. Van Duzer, *Fields and Waves in Communication Electronics*, p. 151-155, John Wiley and Sons, New York (1984).

[7] R. C. Taber, "A Parallel Plate Resonator

Technique for Microwave Loss Measurements on Superconductors," Rev. Sci. Instrum. **61**, 2200 (1990).

[8] N. Newman, K. Char, S. M. Garrison, R. W. Barton, R. C. Taber, C. B. Eom, T. H. Geballe, and B. Wilkens, "$YBa_2Cu_3O_7$ Superconducting Films with Low Microwave Surface Resistance over Large Areas," Appl. Phys. Lett. **57**, 520 (1990).

[9] C. B. Eom, J. Z. Sun, B. M. Lairson, S. K. Streiffer, A. F. Marshall, K. Yamamoto, S. M. Anlage, J. C. Bravman, T. H. Geballe, S. S. Laderman, R. C. Taber, and R. D. Jacowitz, "Synthesis and Properties of $YBa_2Cu_3O_7$ Thin Films Grown *in situ* by 90° off-axis Single Magnetron Sputtering," Physica C **171**, 354 (1990).

[10] D. W. Cooke, P. N. Arendt, E. R. Gray, R. E. Muenchausen, B. L. Bennet, S. R. Foltyn, R. C. Estler, X. D. Wu, G. A. Reeves, N. E. Elliot, D. R. Brown, A. M. Portis, R. C. Taber, and A. Mogro-Campero, "Microwave Properties of HTS Films," Proc. 2nd Int'l. Conf. on Elect. Mats., 93 (1990).

[11] Paul Merchant, Hewlett-Packard Laboratories, private communication.

[12] T. Van Duzer and C. W. Turner, *Principles of Superconductive Devices and Circuits*, p. 128-131, Elsevier North Holland, Inc., New York (1981).

[13] S. S. Laderman, R. C. Taber, R. D. Jacowitz, J. L. Moll, C. B. Eom, T. L. Hylton, A. F. Marshall, T. H. Geballe, and M. R. Beasley, "Resistive Loss at 10 GHz in c-axis-aligned *in situ*-grown $YBa_2Cu_3O_7$ films," Phys. Rev. B **43**, 2922 (1991).

[14] J. D. Jackson, *Classical Electrodynamics*, p. 255-259, John Wiley and Sons, New York (1962).

[15] G. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-low Noise," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., **UFFC-37**, 339 (1990).

[16] S. S. Elliot and R. C. Bray, "Direct Phase Noise Measurements of SAW Resonators," Proc. IEEE Ultrasonics Symposium, Nov. 14-16, p.180 (1984).

[17] A. P. S. Khanna, M. Schmidt, and R. B. Hammond, "A Superconducting Resonator Stabilized Low Phase Noise Oscillator," Microwave Journal, p. 127, February, 1991. The data cited in table 1 was actually taken from an STI application note that gave a result approximately 10 dB better than value in Khanna, et al.

[18] D. E. Oates, A. C. Anderson and B. S. Shih, "Superconducting Stripline Resonators and High-T_c Materials," IEEE MTT-S Digest, 627 (1989).

[19] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low-Phase-Noise SAW Resonators and Oscillators: Design and Performance," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., **UFFC-35**, 657 (1988).

Superconducting Thin-Film $YBa_2Cu_3O_{7-x}$ Resonators and Filters*

D. E. Oates, W. G. Lyons, and Alfredo C. Anderson,
Lincoln Laboratory, Massachusetts Institute of Technology,
Lexington, MA 02173-9108

Abstract

Measurements of surface resistance R_S in thin films of $YBa_2Cu_3O_{7-x}$ produced by magnetron sputtering have been carried out using a stripline resonator. The best values that we have obtained are $R_S = 2.6 \times 10^{-6}\ \Omega$ at 4 K and $R_S = 8.3 \times 10^{-6}\ \Omega$ at 77 K, both at 1.5 GHz. Four-pole microstrip bandpass filters have been fabricated using these thin films on 1.3 × 2.3-cm $LaAlO_3$ substrates. Such filters with a center frequency of 4.8 GHz and a bandwidth of 1% have exhibited a passband insertion loss as low as 0.5 dB at 77 K. The R_S of $YBa_2Cu_3O_{7-x}$ films has been found to increase with input power at power levels that depend on film quality, an effect that can be related to the rf magnetic field at the conductor surface. The power dependence of R_S leads to nonlinear effects, such as generation of harmonics and intermodulation products, which must be avoided in linear device applications. Measurements of intermodulation products in the filters are presented as well as measurements of phase noise in filters and resonators.

Introduction

Applications of the high-transition-temperature (high-T_c) superconducting materials at microwave frequencies are of interest because the losses in these materials at 10 GHz and at temperatures as high as 77 K have been shown to be at least an order of magnitude lower than those in copper at the same frequency and temperature. Furthermore, based on theoretical considerations, the surface resistance R_S might be lowered by another factor of 10 to 100 by improvements in the material fabrication methods. In addition to lower R_S, superconducting transmission lines have much lower dispersion than normal metal lines. High-Q-factor resonators, low-loss narrowband filters, and long delay lines are therefore the most promising applications, because they require very low conductor losses and (in the case of delay lines) low dispersion. Numerous practical devices have already been demonstrated.[1,2]

*This work was conducted under the auspices of the Consortium for Superconducting Electronics with full support by the Defense Advanced Research Projects Agency.

It has been known for some time that values of R_S of the high-T_c materials are dependent on the level of microwave power used in the measurement. This effect can be directly related to the peak rf magnetic field H_{rf} at the surface of the superconductor. Recently completed calculations of current distributions in stripline transmission line structures[3] enable quantitative measurements of R_S vs H_{rf} at the conductor's surface. By using the results of the calculations and the measured values of Q and resonant frequency of a stripline resonator, microwave-frequency R_S has been measured as a function of frequency from 1.5 to 20 GHz, temperature from 4 K to the transition temperature ($T_c \approx 90$ K), and H_{rf} from 0 to 300 Oe.

Measurements of $R_S(H_{rf})$ have shown that the best epitaxial films exhibit weak dependence of R_S on H_{rf}, while lower-quality films, believed to be composed of grains connected by weak links, show an R_S increasing in linear proportion to H_{rf}. The magnetic field dependence of R_S leads to the generation of nonlinear effects at power levels that in lower-quality films can be well below those corresponding to the critical field H_{c1}. In this paper, measurements of the intermodulation (IM) products of low-loss filters fabricated from $YBa_2Cu_3O_7$ are presented as well as measurements of the microwave-frequency phase noise in the filters and in high-Q resonators. Projections of the phase noise for oscillators employing stripline resonators as the stabilizing element are also reported.

Measurements of R_S

Measurements of R_S in thin films of $YBa_2Cu_3O_{7-x}$ have been carried out using the stripline resonator shown schematically in Fig. 1. The films were deposited on $LaAlO_3$ substrates by an in-situ off-axis sputtering technique described previously.[4] The stripline resonator method of determining R_S has been reported in detail elsewhere,[5] and only a brief summary is given here. To determine R_S, we first determine the magnetic penetration depth of the films, λ, by measuring the center frequency f of the fundamental resonance as a function of temperature from low temperatures to as close to T_c as possible. We then fit f(T) to a two-fluid-model expression of λ (T) by using the calculated current distributions[3] to model the dependence of the stripline inductance on λ.

CH2965-2/91/0000-460 $1.00 © 1991 IEEE

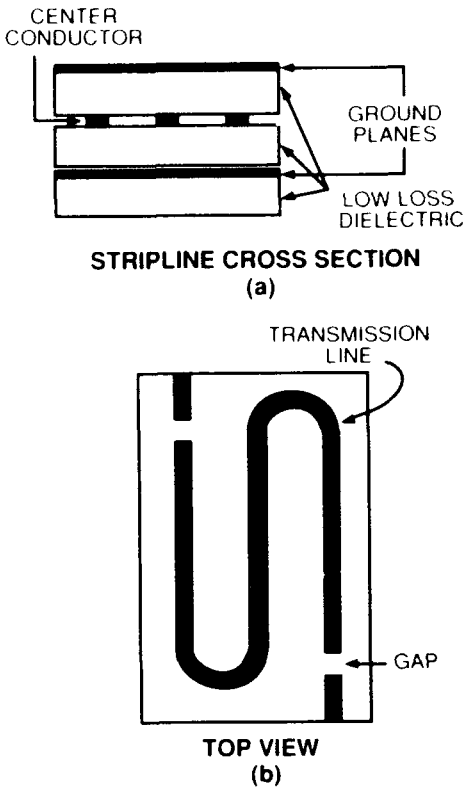


Figure 1. Schematic (a) cross-sectional and (b) top views of the stripline resonator used in the measurements.

Then we measure the Q of the resonator as a function of temperature at the fundamental of 1.5 GHz and all of the overtone frequencies up to 20 GHz. By knowing λ (T), the calculated current distribution is used to extract the surface resistance from the Q. The results of these measurements are shown in Fig. 2 for R_S vs T and in Fig. 3 for R_S vs f for a resonator fabricated from some of the best films produced at Lincoln Laboratory. These values of R_S represent results for patterned films and are comparable to the best results reported by other laboratories.[6] Surface resistance in the films at 10 GHz and 77 K is more than 10 times smaller than that in copper at the same temperature.

The Q of the resonator and therefore the surface resistance of the films of $YBa_2Cu_3O_{7-x}$ are dependent on the power input to, and H_{rf} in, the resonator. The power dependence of the surface resistance has been observed in other high-quality films by a number of workers[7] in $YBa_2Cu_3O_{7-x}$ and other high-T_c superconductors. It seems to be a property of all such films fabricated to date, irrespective of film quality. The degree of power dependence, however, does seem to be related to film quality because films with high quality, determined by means other than measurement of the microwave properties, appear to exhibit lower dependence on H_{rf}. Yet, the origins of the power dependence are not well understood, and it is not known, for instance, if the power dependence is an intrinsic property of the high-T_c materials of if better fabrication methods will eliminate it. Figure 4 shows the results of our measurements of R_S as a function of H_{rf}.

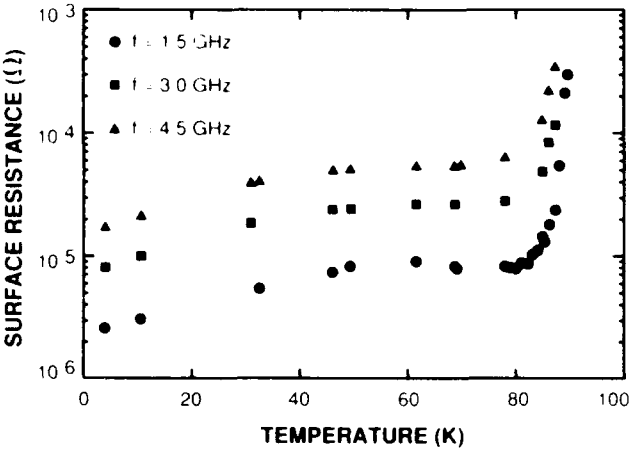


Figure 2. Surface resistance vs temperature for the first three modes of the best sputtered film. The frequency of each mode is given.

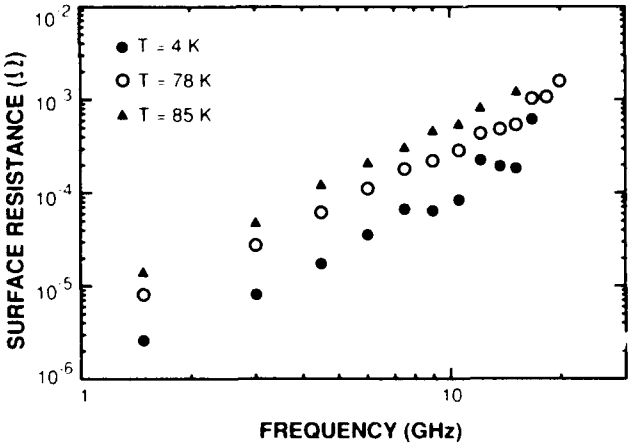


Figure 3. Surface resistance vs frequency for three different temperatures as indicated. These results are for the best sputtered films.

We note that conventional, low-T_c superconductors also exhibit power-dependent R_S.[8] It is encouraging that the power dependence measured at low temperatures in $YBa_2Cu_3O_{7-x}$ resonators is similar to that measured in resonators fabricated from the conventional superconductor Nb. In the Nb resonators, R_S is essentially independent of H_{rf} up to peak fields of 300 Oe.

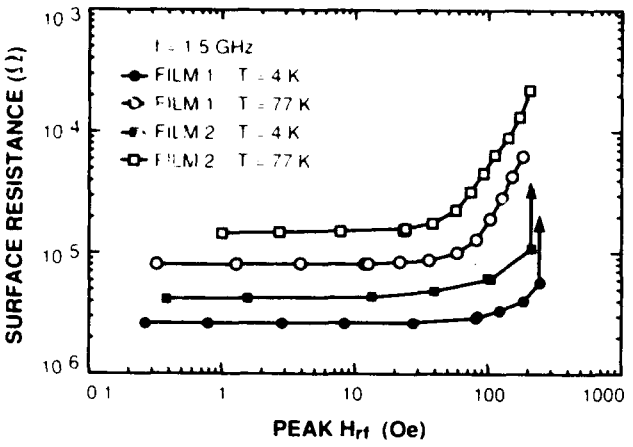


Figure 4. Surface resistance vs peak H_{rf}, the peak field at the edges of the center strip, for two films at 4 and 77 K. The frequency in all cases is 1.5 GHz.

Devices

Resonators for Oscillator Stabilization

Superconductors have been proposed for the stabilization of low-noise, microwave-frequency oscillators because of the high Q that can be achieved at a high fundamental frequency in the resonators.[9] It is well known that increased Q results in decreased noise. The resonator used to obtain the surface resistance values shown in Fig. 2 have a Q of 120,000 at 1.5 GHz at low temperature and power. This resonator was designed for material evaluation, and the Q value can be raised by as much as a factor of 10 by optimizing the design for high Q. Such high Q values immediately suggest application to oscillator stabilization, and since the high Q is attainable at high frequencies, direct X-band operation is feasible without the need for frequency multiplication.

In order to project the performance of an oscillator stabilized by a resonator fabricated with a high-T_c superconducting thin film, we have calculated the phase noise for typical performance parameters. The analysis of the phase noise of feedback oscillators given by Leeson[10] leads to the following well-known expression for the single-sideband phase noise L(ω):

$$L(\omega) = 10\log\left[N^2\left(1+\frac{\omega_0^2}{4Q^2\omega^2}\right)\left(\frac{\alpha}{\omega}+\frac{GFkT}{P}\right)\right]$$

$$= 10\log N^2\left[\frac{\alpha\omega_0^2}{4Q^2\omega^3}+\frac{GFkT\omega_0^2}{4Q^2P\omega^2}+\frac{\alpha}{\omega}+\frac{GFkT}{P}\right]$$

where L(ω) is the noise power relative to the carrier in a 1-Hz bandwidth given in dBc/Hz, N is the frequency-multiplication factor, ω_0 is the oscillator fundamental frequency, Q is the loaded resonator quality factor, ω is the offset frequency in rad/s, α is the empirically determined flicker-noise constant, G is the loop gain, F is the amplifier noise figure, k is Boltzmann's constant, T is the absolute temperature of the amplifier, and P is the power at the output of the amplifier. Since the Q enters in the denominator of the first two terms, it is obvious that high-Q resonators are desirable for low phase noise.

The flicker (1/f) noise, which is characterized by α, can originate in either the amplifier or the resonator or both. Amplifier flicker noise is suppressed by the $1/Q^2$ term in the Leeson expression, but flicker noise that originates in the resonator (typically variations in resonant frequency) usually increases as Q^2. Therefore, the resonator flicker-noise contribution is independent of the Q. Since the flicker noise is not well understood theoretically, it must be determined empirically. We can make reasonable estimates for the amplifier, but for the resonator we must rely on experimentally determined values. Our projections of phase noise that follow assume that the contribution to the flicker noise from the resonator is small compared to that from the amplifier. The phase noise at offsets of less than about 5 kHz is dominated by the flicker noise. If the flicker noise contribution of the superconducting resonator is larger than that of the amplifier, then the actual phase noise will be greater than the following predictions.

To project the phase noise of an oscillator operating at 10 GHz, we have made a set of assumptions, presented in the following. We assume that the resonator and amplifier are at 77 K and room temperature, respectively, and that the R_S of the $YBa_2Cu_3O_{7-x}$ is 10^{-4} Ω at 10 GHz and 77 K. We also assume a modified stripline resonator design that is optimized for high-Q operation by using a line of 2-mm width rather than the 150-μm width used in the resonators for film characterization. This choice of linewidth yields a calculated Q of 10^5 for our assumed R_S. (To calculate the Q, we used the incremental inductance technique discussed in Ref. 5.) We assume an insertion loss of 15 dB and an amplifier noise figure of 3 dB. For this calculation we have assumed a loop power level of +20 dBm. This is a high power level for high-Q operation of the high-T_c materials, and in reality we might need to reduce the power to approximately +10 dBm, which would imply an internal rf current of about 3 A. On the 150-μm-wide line used for the measurements shown in Fig. 4, a current of 0.3 A at low temperature produces a small increase in R_S. However, as will be seen, the exact value of power chosen does not strongly influence the conclusions drawn from the projections. We also assume that losses in the dielectric substrate are negligible.

The value of α assumed for the projections is 5×10^{-13}, which corresponds to a 1-Hz intercept of −130 dBc/Hz in an open-loop configuration. While this is smaller by 10 dB than the value normally achieved in the best GaAs FET amplifiers, there are indications that Si permeable base transistors should be able to provide this level of performance.[11] The use of this value of α in the phase-noise projections assumes that the resonator will not contribute to the oscillator flicker noise. This assumption may be wrong as we will discuss later. At present, outside of our own measurements (presented in Ref. 7 and here), there are no published values of superconductor flicker noise. Figure 5 shows the results of the projections using Leeson's model with the various assumptions made here. Also shown in Fig. 5 are values of phase noise for state-of-the-art quartz oscillators and for the best reported oscillators stabilized by surface acoustic wave (SAW) resonators.[12] Clearly, with our assumptions, the superconducting oscillator provides lower noise than the other low-noise oscillators over the range of offset frequencies important to Doppler radar applications. An added important advantage of the superconducting oscillator over the others is direct operation at X-band without multiplier stages, resulting in considerable savings in size, weight, and power.

The power in the oscillator loop only affects the phase noise at offset frequencies greater than 100 kHz, where the superconductor shows a large advantage over the other technologies. Reduction of the operating power by 10 dB would not appreciably lessen the advantages of the superconductor-stabilized oscillator.

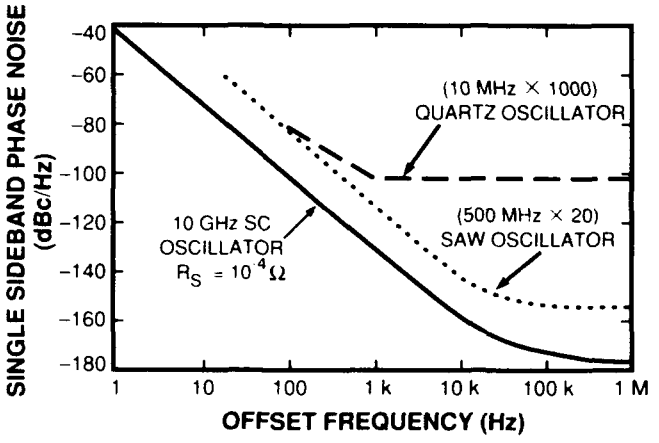


Figure 5. Results of phase noise projections for a superconducting-resonator-stabilized oscillator operating at 10 GHz, neglecting flicker noise in the superconducting resonator. Parameters used in the projection are given in the text. Also shown for comparison are results for a 10-MHz quartz-crystal oscillator and a 500-MHz SAW-resonator-stabilized oscillator after multiplication to 10 GHz.

Filters

To demonstrate the potential performance benefits of high-T_c materials in filter technology, we chose for implementation in $YBa_2Cu_3O_{7-x}$ a four-pole Chebyshev bandpass design, using microstrip geometry, with 1.3% bandwidth, 0.05-dB passband ripple, and 4.8-GHz center frequency. Figure 6 shows the layout of the filter. It was designed using an iterative technique[13] and was initially fabricated using gold signal lines on a 500-μm-thick $LaAlO_3$ substrate. $LaAlO_3$ is the most frequently used substrate material for $YBa_2Cu_3O_{7-x}$, because it has low losses at microwave frequencies and because high-quality thin films can be grown epitaxially on its polished surface. However, the dielectric constant ($\varepsilon_r \approx 24$) is larger than that which typical microwave circuit design software can treat with accuracy, necessitating the iterative design technique using gold metallization. The filter was fabricated in a microstrip configuration consisting of a $YBa_2Cu_3O_{7-x}$ signal line and a silver ground plane on opposite sides of a single 1.3 × 2.3-cm substrate. The $YBa_2Cu_3O_{7-x}$ film in the filter was 0.3 μm thick. A silver ground plane does not appreciably affect the performance of the filter because the losses of the patterned stripline dominate those of the ground plane. The current is much more strongly concentrated in thin patterned lines than in the unpatterned ground plane.

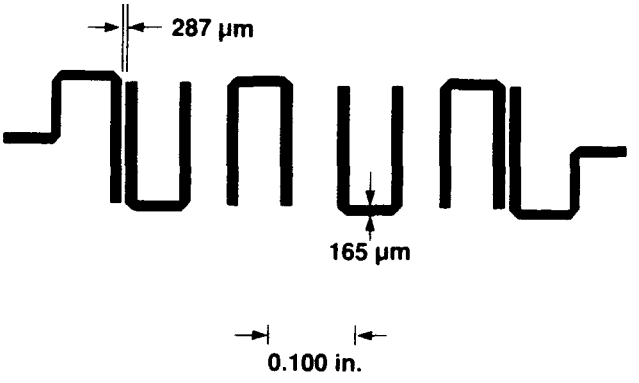


Figure 6. Four-pole microstrip filter design for operation at 4.8 GHz with a 1.3% bandwidth.

The filter was patterned by standard photolithographic techniques using positive photoresist and by wet etching using dilute phosphoric acid. To make electrical contact to the $YBa_2Cu_3O_{7-x}$, silver bonding pads were deposited by electron-beam evaporation and annealed at 400 °C for 1 h. Ultrasonically bonded aluminum ribbons were used to connect the bonding pads to Wiltron-K connectors. The filter was mounted in a hermetically sealed package. The filter and package have been qualified for use in space in the Naval Research Laboratory High Temperature Superconductivity Space Experiment.[14]

Filter Performance

The frequency response of the filter measured at 77 K is shown in Fig. 7 along with the measured responses of identical designs, also on $LaAlO_3$ substrates, fabricated from silver measured at 77 K and gold measured at 300 K. The thickness of the metals of both of the normal-metal filters was several skin depths. The $YBa_2Cu_3O_{7-x}$ filter clearly shows superior performance because of the lower R_S of the superconductor film. Passband insertion loss of the $YBa_2Cu_3O_{7-x}$ filter is 0.5 dB, which agrees well with the calculated response and is due entirely to the R_S of the material. The filter fabricated from the silver film has a loss of 6 dB at 77 K, which can also be attributed to the losses in the metal. The filter made with gold film shows a 12-dB loss. The advantage of the superconducting filters over normal-metal filters becomes substantially greater with narrower passbands. For example, calculations indicate that a ten-pole filter with a center frequency of 4 GHz would exhibit a 1-dB insertion loss when implemented with $YBa_2Cu_3O_{7-x}$ of quality identical to these films, while a silver filter at 77 K would have more than 10 dB of loss.

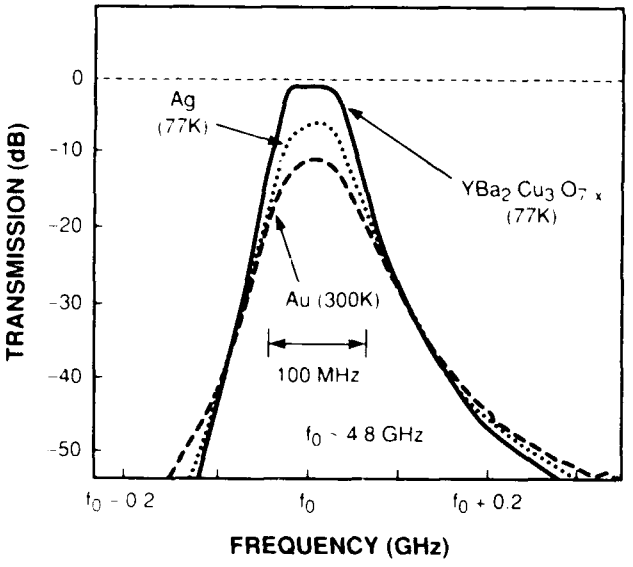


Figure 7. Measured filter response of the design shown in Fig. 6 fabricated from gold (measured at 300 K), silver (measured at 77 K), and $YBa_2Cu_3O_{7-x}$ (measured at 77 K).

Measurement of Intermodulation Distortion and Noise

As we have demonstrated by our measurements, the R_S of the $YBa_2Cu_3O_{7-x}$ films is nonlinear so that effects such as IM distortion must be considered. Figure 8 shows the results of measurement of the IM products in one of the filters of a design identical to that shown in Fig. 6 but having a 2.7-dB insertion loss due to an impedance mismatch at one of the input ports. The measurements shown were not corrected for approximately 1 dB of probe loss. These results were obtained by a standard two-tone IM measurement, in which two frequencies f_1 and f_2 are applied to the input of the filter and third-order products such as $2f_1-f_2$ are observed in a spectrum analyzer at the output. As expected, the third-order products are proportional to the cube of the input power and show a slope of 3 on the log-log plot while the fundamental shows a slope of 1. The intersection of the slope 3 and slope 1 lines is the third-order intercept (TOI) and is the parameter normally used to characterize the IM distortion. The TOI is approximately +25 dBm at 77 K. This implies that at normal operating power levels of -10 to 0 dBm the IM products will be more than 50 dB below the main signal. This is sufficient suppression for most system applications. Operation at higher power levels, however, requires that IM distortion be considered.

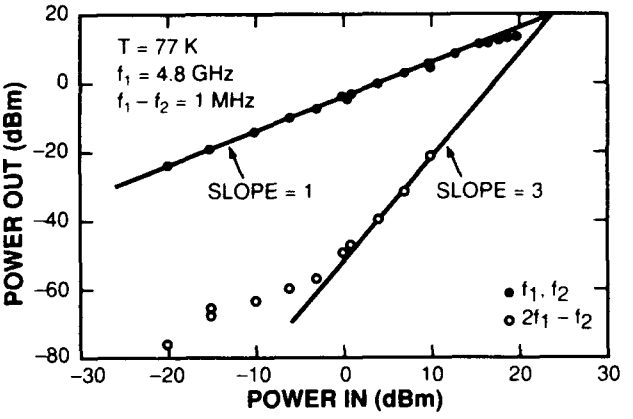


Figure 8. Results of two-tone intermodulation-product measurements for the filter design in Fig. 6. Shown are the fundamental and the third-order intermodulation product. Measurements were done at 77 K and 4.8 GHz with tone separation of 1 MHz.

Also evident in Fig. 8 is the onset of saturation of the output of the fundamental at input power between +15 and +20 dBm. This is a result of the increase in R_S at high levels of rf current. We have observed substantially higher IM products, saturation at much lower input power, and much lower power TOI in filters fabricated with films that showed larger levels of rf magnetic field dependence in the R_S. The TOI of +25 dBm observed here must be understood as that of a high-quality film and will not necessarily hold for films of lower quality. IM measurements in resonators have been reported previously by us.[15]

Figure 9 shows results of phase noise measurements in the bandpass filter, which were made using the standard bridge technique of splitting a synthesizer output signal, passing one arm through the device under test, and combining the two signals with a $\pi/2$ phase shift in a mixer.[16] The measurements were done at 77 K and 4.8 GHz. Plotted is the single-sideband phase noise (in dBc/Hz) vs frequency offset from the carrier (in Hz). The lower curve shows the system noise floor, which is dominated by the amplitude modulation noise of the low-noise synthesizer. The upper curve is the filter phase noise. It shows the usual 1/f dependence. The rise in noise at offset frequencies greater than 1 MHz is due to the decorrelation of the source phase noise resulting from unequal delays in the two arms of the bridge setup used for the measurements. The slight bump in phase noise at about 400 kHz is not understood. The 1-Hz intercept of -110 dBc/Hz is somewhat higher than expected but still small enough that there should be no impact in a real communication system, which is the intended application of this filter. No measurable white noise is generated in the filter as would be noted by a rise in the noise floor at large offset frequencies.

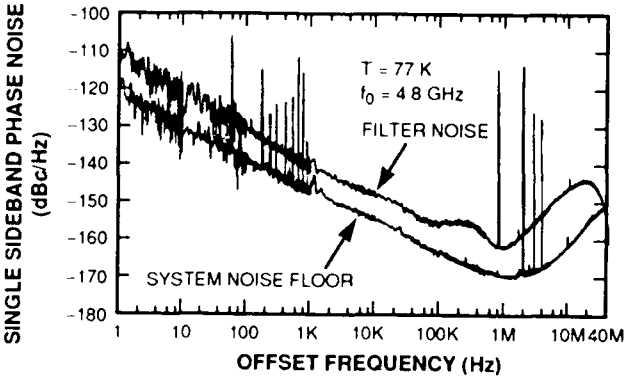


Figure 9. Measurements of single-sideband phase noise of the filter in Fig. 8. Shown are the system noise floor (lower curve) and the filter noise (upper curve). Measurements were done at 77 K and 4.8 GHz.

Figure 10 shows the measurements of phase noise in the $YBa_2Cu_3O_{7-x}$ resonator used to carry out the R_S measurements of Fig. 2. The measurements were done at 4 K and 1.5 GHz. Input power was approximately -20 dBm and the resonator Q was 10^5. Because of the low input power and the 30-dB insertion loss of the resonator, the system noise floor is -130 dBc/Hz as shown at large offset frequencies in Fig. 10. The system 1/f noise is not observable at this frequency with this noise floor. The measurements appear to contain effects due to vibration. We know that the present design of the resonator package is sensitive to vibration. However, one can discern a 1/f trend in the data at offset frequencies lower than 1 kHz. We believe that the peak at 1 kHz is due to a collection of unresolved spurious peaks resulting from power-line interference and is not associated with the resonator. The 1/f line drawn on the figure indicates a 1-Hz intercept of -100 dBc/Hz. This is 30 dB greater than the value assumed for the phase-noise projections made above, and if such a value were used it would seriously degrade the results of those projections for small offset frequencies.

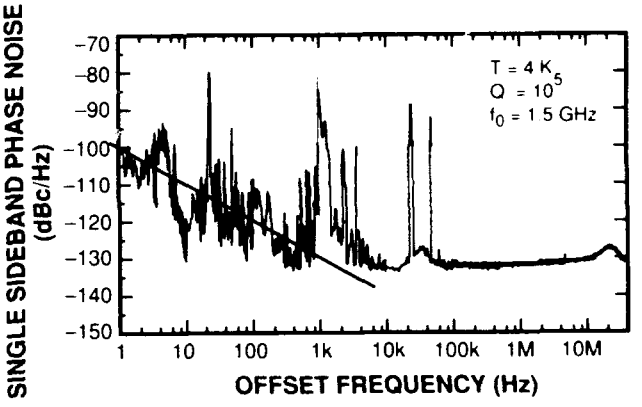


Figure 10. Measurements of phase noise of the $YBa_2Cu_3O_{7-x}$ resonator used to make the R_S measurements. The resonator Q = 10^5. Measurements were done at 4.2 K and 1.5 GHz with -20-dBm input power. The 1/f line drawn on the figure indicates a 1-Hz intercept of -100 dBc/Hz.

These results of resonator phase noise should be considered preliminary. More films must be measured to determine if this 1/f noise performance is typical of the high-T_c materials. Investigations of the dependence of 1/f noise on film fabrication and photolithography must also be undertaken.

Summary

Measurements of surface resistance of high-quality $YBa_2Cu_3O_{7-x}$ thin films have been carried out using a stripline resonator. The film quality is comparable to the best that has been reported to date in patterned thin films, with an R_S more than 10 times better than that in copper at 10 GHz and 77 K. The films are capable of handling the amounts of rf magnetic field and power encountered in many practical devices.

Narrow-bandpass filters fabricated from the $YBa_2Cu_3O_{7-x}$ films show nearly ideal performance with losses significantly lower than those of filters fabricated with normal metals. The IM distortion and noise levels measured in the filters show that both are low enough that there would be no degradation of system performance using the superconducting filters. These filters were designed to replace conventional dielectrically loaded cavity filters currently used in satellite communication systems. Use of the superconducting filters would provide very significant savings of size and weight in such a satellite. For a system with 50 superconducting filters the weight reduction, even when the refrigeration unit for the filters is included, would be approximately 50%.

Projections of the phase noise of oscillators stabilized with high-Q superconducting resonators indicate that improvement of up to 20 dB is possible. Noise measurements in the resonators indicate that the flicker noise might be large enough that oscillator performance would not reach that of the projections. Modifications in the resonator packaging and design will be investigated in order to understand the origins of the flicker noise in resonators.

Acknowledgment

We thank P.M. Mankiewich of AT & T Bell Laboratories for some of the films used for development of the filters. We thank J.M. Hamm, R.P. Konieczka, R.S. Tompkins, and S.A. Ladd for device fabrication. We also thank G.L. Fitch for computer processing.

References

[1] W.G. Lyons, R.R. Bonetti, A.E. Williams, P.M. Mankiewich, M.L. O'Malley, Alfredo C. Anderson, R.S. Withers, and R.E. Howard, "High-T_c Superconducting Microwave Filters," IEEE Trans. Magn., vol. 27, pp. 2537-2539, March 1991.

[2] W.G. Lyons, R.S. Withers, J.M. Hamm, Alfredo C. Anderson, P.M. Mankiewich, M.L. O'Malley, and R.E. Howard, "High-T_c Superconductive Delay Line Structures and Signal Conditioning Networks," IEEE Trans. Magn., vol. 27, pp. 2932-2935, March 1991.

[3] D.M. Sheen, S.M. Ali, D.E. Oates, R.S. Withers, and J.A. Kong, "Current Distribution for Superconducting Strip Transmission Lines," IEEE Trans. Appl. Superconduct., vol. 1, pp. 108-115, June 1991.

[4] A.C. Westerheim, L.S. Yu-Jahnes, and Alfredo C. Anderson, "Off-Axis Magnetron Sputtering on YBCO Films: The Influence of Atomic Oxygen," IEEE Trans. Magn., vol. 27, pp. 1001-1005, March 1991.

[5] D.E. Oates, P.M. Mankiewich, and Alfredo C. Anderson, "Measurements of the Surface Impedance of $YBa_2Cu_3O_{7-x}$ Thin Films Using Stripline Resonators," J. Superconduct., vol. 3, pp. 251-259, 1990.

[6] H. Piel and G. Müller, "The Microwave Surface Impedance of High-T_c Superconductors," IEEE Trans. Magn., vol. 27, pp. 854-862, March 1991.

[7] D.E. Oates and Alfredo C. Anderson, "Surface Impedance Measurements of $YBa_2Cu_3O_{7-x}$ Thin Films in Stripline Resonators," IEEE Trans. Magn., vol. 27, pp. 867-871, March 1991, and references therein.

[8] D.E. Oates, C.C. Chin, G. Dresselhaus, and M.S. Dresselhaus (unpublished).

[9] D.E. Oates, Alfredo C. Anderson, and B.H. Shih, "Superconducting Resonators and High-T_c Materials," in 1989 MTT-S Digest, pp. 627-630, 1989.

[10] D.B. Leeson, "Short-Term Stable Microwave Sources," Microwave J., vol. 13, p. 59, 1990.

[11] D.E. Oates and D. Rathman (unpublished).

[12] G.K. Montress, T.E. Parker, and M.J. Laboda, "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," in Proceedings of the 1987 Ultrasonics Symposium, pp. 47-52, 1987.

[13] R.R. Bonetti and A.E. Williams, "Preliminary Design Steps for Thin-Film Superconducting Filters," in 1990 MTT-S Digest, pp. 273-277, 1990.

[14] J.C. Ritter, M. Nisenoff, G. Price, and S.A. Wolf, "High Temperature Superconductivity Space Experiment (HTSSE)," IEEE Trans. Magn., vol. 27, pp. 2533-2536, March 1991.

[15] D.E. Oates, Alfredo C. Anderson, D.M. Sheen, and S.M. Ali, "Stripline Resonator Measurements of Z_s vs H_{rf} in $YBa_2Cu_3O_{7-x}$ Thin Films," to be published in IEEE Trans. MTT, Special Issue on Applications of High-T_c Superconductors, Sept. 1991.

[16] T.E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators," in Proceedings of the 41st Annual Frequency Control Symposium, pp. 99-110, 1987.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A High T_c Superconducting Resonator for a Compact Hydrogen Maser

D. B. Opie, H. E. Schone
College of William and Mary, Williamsburg, VA 23185

M. Hein, G. Müller, H. Piel, H.-P. Schneider
University of Wuppertal, Wuppertal, Germany

V. Folen, A. Frank, W. M. Golding, S. Wolf
Naval Research Laboratory, Washington D.C., 20375

Abstract

The advent of high T_c superconductors (HTSC) has made feasible the application of superconductivity to practical microwave devices. The measured surface resistance, R_s, of the new HTSC materials is lower than that of copper measured at the same temperature, $77K$, and frequency, 1.42GHz. An interesting application of these new materials is the miniaturization of microwave cavity resonators. In this report we describe the development, testing and evaluation of a superconducting compact hydrogen maser resonator made from electrophoretic $Y_1Ba_2Cu_3O_{7-\delta}$ (YBCO). This compact loop-gap resonator, based on a previously suggested maser resonator [1], is made superconducting using an electrophoretic process developed for the deposition of thick film polycrystalline HTSC on large non-planar metallic substrates. At $77K$ we obtain cavity quality factors comparable to those of standard size, room temperature TE_{011} maser resonators. The fields of the resonator have been studied using numerical techniques to determine the dependence of the filling factor, η', and the cavity quality factor, Q_c, on the geometric parameters. This information is used to optimize the cavity design with respect to the effects of thermal radiation on the maser performance at $77K$.

Introduction

The development of hydrogen masers as atomic clocks for space applications requires that the size and weight of the maser physics package be held to a minimum. For these applications the standard maser resonator, a cylindrical cavity operated in the TE_{011} mode at 1.42GHz, is prohibitively large. A miniaturized resonator made of non-superconducting material, but otherwise similar to the one described in this paper, is used in a compact hydrogen maser developed under NRL contract by Hughes [2]. In that maser the technique of Q enhancement or active feedback is necessary to attain sustained oscillations in the room temperature physics package. The superconducting resonator described here was designed to achieve sustained maser oscillations at $77K$ in a small package without the need for active feedback. The use of YBCO coated electrodes has made it possible to obtain the high Q_c values necessary for self oscillation within the miniaturized maser cavity.

Briefly, the operation of the maser is as follows: A beam of spin polarized atomic hydrogen is allowed to enter the interaction region of a microwave resonator which is tuned to the hyperfine transition at 1420.405MHz. The atoms are trapped in the interaction region by a Teflon coated storage bulb. The storage bulb increases the effective observation time of the atoms, thus narrowing the observed atomic linewidth. Collisions with the wall of the storage bulb and spin exchange collisions between hydrogen atoms limit the minimum achievable atomic linewidth. Provided that the atomic resonance line is narrow enough and the cavity Q is sufficient, maser oscillation is obtained. The hydrogen maser is described in detail elsewhere [3, 4].

In figure 1 we show the design of the superconducting loop-gap resonator. It consists of two parts: the loop-gap electrodes, and the outer rf shield-can. The two half cylinders are the electrodes, which behave as inductive elements. The gaps between the electrodes act as capacitive elements. This combined resonant

CH2965-2/91/0000-467 $1.00 © 1991 IEEE

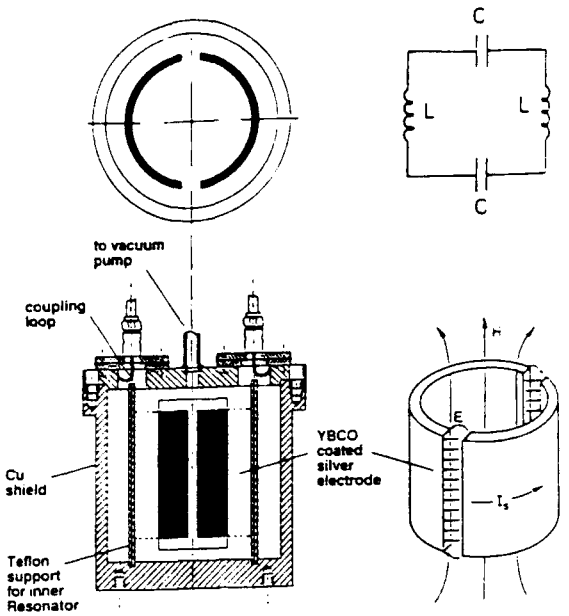


Figure 1: Schematic of a compact loop-gap maser resonator with the field configuration near the inner electrodes and the equivalent circuit.

structure largely determines the resonant frequency of the cavity. The electrodes are supported by two interlocking teflon cylinders and are held fixed, in a concentric position inside the rf shield-can. The rf currents flow azimuthally, as in a solenoid, and produce a nearly uniform axial rf magnetic field within the atomic interaction region, which is the cylindrical volume between the electrodes. The electrode surface currents account for 75% of the power losses in the maser resonator. The remaining losses occur on the surface of the rf shield-can and in the teflon support material. To reduce the dielectric losses, the teflon material has been removed from the high electric field regions near the $4mm$ wide electrode gaps.

The rf shield-can is constructed from OFHC copper and also serves as the vacuum container. Both the inner diameter and the length of the electrodes are $5.0cm$. The rf shield-can diameter and length are both $7.5cm$ producing a cavity volume of $331cm^3$. The loop-gap resonator has only 1.5% of the volume of the standard maser resonator.

The stability of an atomic frequency standard is measured by the square root of the Allan variance of fractional frequency fluctuations, $\sigma_y(\tau)$, where τ is the sampling time of the measurement [5]. The frequency fluctuations may be divided into two classes: systematic and random [6]. Systematic fluctuations are dependent on the quality of the isolation between the oscillator and the environment. Temperature changes, mechanical vibrations, and changes in the dc magnetic field are disturbances which cause systematic fluctuations. Random fluctuations are due to fundamental noise processes such as the thermal radiation within the microwave cavity and the thermal noise added by amplifiers.

It is difficult to predict the stability limitations due to systematic effects. However, the portion of $\sigma_y(\tau)$ that is limited by the cavity thermal noise within the atomic linewidth, which we call $\sigma_l(\tau)$, is predictable. The cavity design is optimized for minimum $\sigma_l(\tau)$.

Electrophoresis

Electrophoresis is the migration of charged particles under the influence of an applied electric field. It is used to coat the electrode structure of the loop-gap resonator with YBCO. The electrophoretic process allows for greater flexibility in the size and shape of the substrate than thin film deposition techniques. In this process grains of YBCO powder are colloidally suspended in a polar medium (for example, acetone or butanol) producing a positive surface charge on each grain. Under the influence of an applied electric field, these grains will move towards the negatively charged cathode, depositing themselves on its surface. This deposition process has been described in detail elsewhere [7].

The first step towards developing a fully superconducting loop-gap resonator is covering silver electrode half cylinders electrophoretically with a $20\mu m$ polycrystalline layer of YBCO. The deposition structure consisting of the silver substrate and the anode are immersed in the YBCO solution. A voltage is applied between the electrodes (figure 2) for 30 to 60 seconds to achieve a 5 to 10 μm thick layer. It is preferable to have the c axis of the YBCO crystal oriented perpendicular to the direction of the current flow because of the anisotropic current carrying properties of the YBCO. To achieve this orientation, the deposition is done in a 7 Tesla applied magnetic field. Due to the magnetic anisotropy in the room temperature YBCO

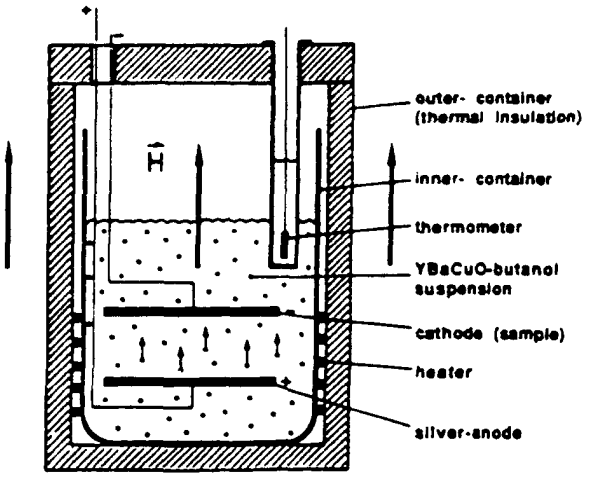


Figure 2: Diagram of an electrophoretic deposition setup for the fabrication of textured thick films of the YBCO material.

grains, the YBCO is deposited with the c axis parallel to the direction of the applied magnetic field. The substrate and film are post annealed at $930°C$ in flowing oxygen for approximately 120 hours. This magnetic texturing technique yields lower R_s values than randomly oriented YBCO thick film samples.

Several pairs of superconducting electrodes have been made by this procedure. These electrodes have nearly $200cm^2$ of non-planar surface area. The surface resistance dependence on temperature is plotted in figure 3. At 1.42GHz and $77K$, the surface resistance is about $1m\Omega$, while the residual resistance measured at $4K$ is less than $100\mu\Omega$. The measured Q_c of the superconducting resonator at $77K$ is 30,000 which is sufficient for maser oscillation in a compact maser without the need for Q-enhancement. An identical resonator built using copper electrodes and operated at $77K$ would have a Q_c of 12,000 and require external feedback to provide Q-enhancement for oscillation.

The Loop-gap Resonator

The loop-gap resonator is a variation of the magnetron resonator which originated as early as 1935 [8]. The applications and characteristics of the loop-gap design have been discussed in more than 15 articles over the past decade. Applications for this design are found in nuclear magnetic resonance, electron spin resonance, radar, and maser spectroscopy at frequencies from 200MHzto 11GHz. The compact hydrogen maser resonator using the lumped element loop-gap design was originally suggested by H. E. Peters [1].

Empirical relations for the frequency and quality factor can be obtained as a function of resonator geometry. No analytic solution for this resonant mode exists. Recently, a technique [9] was developed to approximate the fields inside and outside the electrodes. It yields considerable agreement between measured and calculated values of the fields away from the electrodes. However this technique ignores the greatly enhanced field density at the edges of the electrodes. Without accounting for these enhanced fields incorrect values of R_s are obtained for the YBCO electrodes. Moreover, the filling factor of the maser resonator cannot be calculated properly without first understanding this effect.

It should be noted that, although the current distribution on the surface of the electrodes is solenoidal, the classical picture of a solenoid is an inadequate approximation. Since a solenoid has flux lines that can close *through* its body, the high frequency boundary conditions in the loop-gap will force all flux lines to close around the ends of the conductor and not through the electrode surfaces resulting in a substantially different distribution of currents.

The loop-gap resonator is analogous to the shielded stripline which is a coaxial transmission line of rectangular cross section. It is also similar to the microstrip transmission line. The relations between these three models are depicted in figure 4. The currents and the fields for the three similar structures can be described as TEM. The geometry of the loop- gap can be obtained from the shielded stripline geometry by rotating the cross section of the stripline about the horizontal axis and letting the radius of the ground plane go to zero. The axis of rotation becomes the $z-axis$ of the loop-gap resonator and the center conductor of the stripline becomes the electrode structure of resonator.

The boundary conditions for the loop-gap resonator are the same as those of the shielded stripline. The rf magnetic field distribution of the loop-gap cavity is equivalent to the resonant magnetic field of the shielded stripline. The magnetic field encircles the electrodes as the equivalent field encircles the center conductor of the stripline. The electrode gaps may be neglected due to the displacement currents in the region of the gap. For the range of frequencies considered the fields are TEM and propagating in the azimuthal direction between the electrodes and the rf shield.

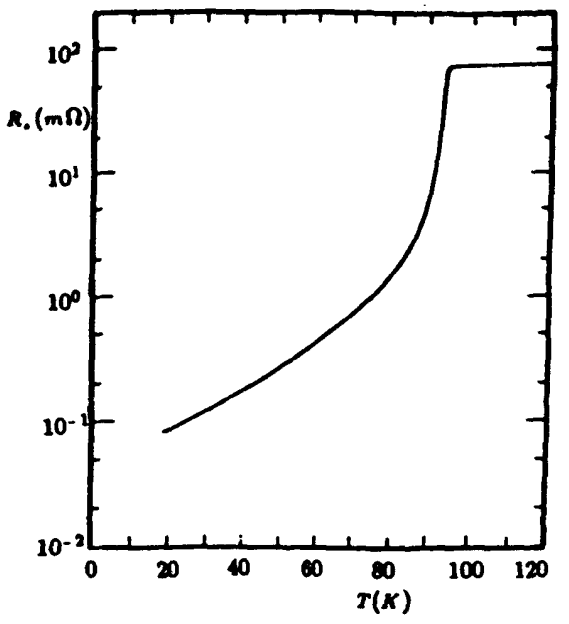


Figure 3: The temperature dependence of the surface resistance of the inner electrodes in the maser cavity measured at 1.42GHz.

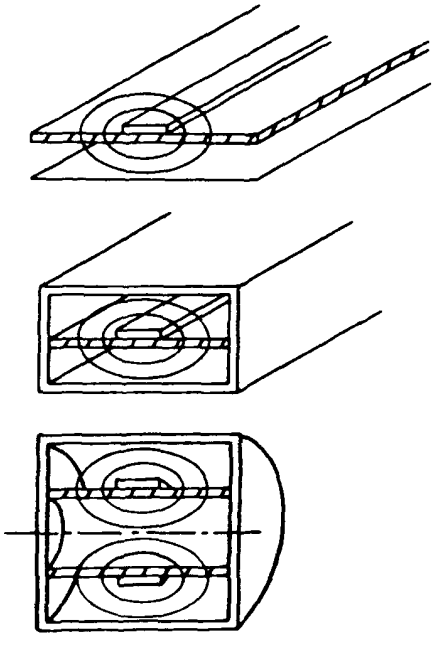


Figure 4: Diagram of the microstrip, shielded stripline, and the maser resonator to illustrate the analogy between the three structures. TEM magnetic field flux lines are shown to be similar for each case

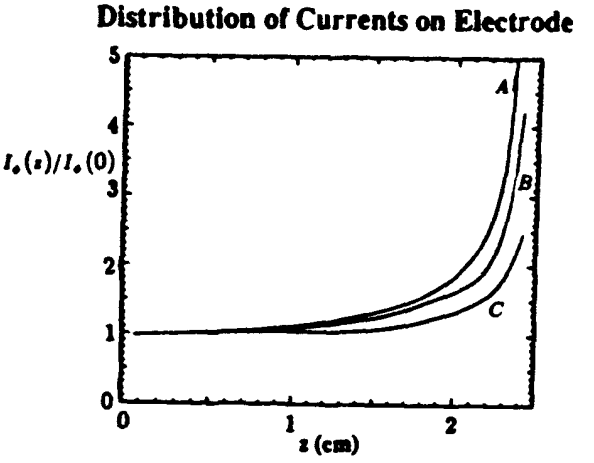


Figure 5: Distribution of the currents calculated along the $\hat{z}$ axis using three different methods: (A) conformal mapping, (B) Green's function, (C) finite element calculation.

Many attempts have been made to solve for the fields around a strip conductor. The difficulty in finding a solution to this family of problems is due to the singularities at the sharp metal corners of the finite inner conductor. However, the assumption that the conductor is infinitely thin has led to solutions that give qualitative descriptions of the fields and currents in limited cases.

As mentioned above, all of the geometries in this class of problems provide current distributions that are not constant along the strip (loop-gap electrodes) width. Specifically, the current density increases sharply at the conductor edges [10, 11, 12, 13]. This is true so long as the fields are considered to be in the "quasi-static" limit; the fields are TEM and the dimensions of the strip conductor are small compared to the wavelength. In this quasi-static limit, the current distribution is related to the distribution of excess static charge placed on the isolated strip conductor.

The relation for the currents on the microstrip is

$$I_\phi(z) = v\sigma(z) \qquad (1)$$

where v is the phase velocity and $\sigma(z)$ is the static charge distribution. The solution for an infinitely thin isolated conducting strip above a ground plane was derived by Maxwell [13]. He found that by using a conformal mapping technique, the static charge distribution on an infinitely thin conducting sheet above a

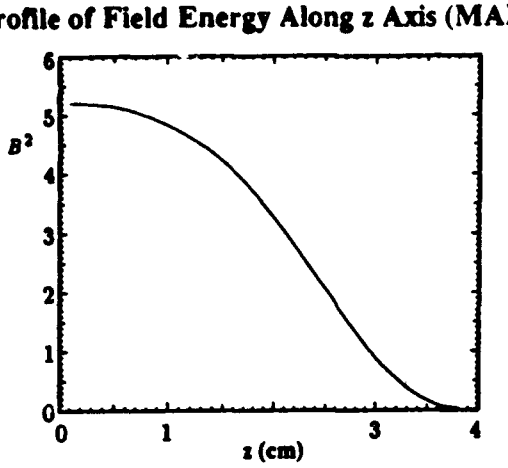


Figure 6: Profile of the magnetic field energy along the $\hat{z}$ axis where $r = 0$ showing the solenoidal field profile within the interaction region of the cavity.

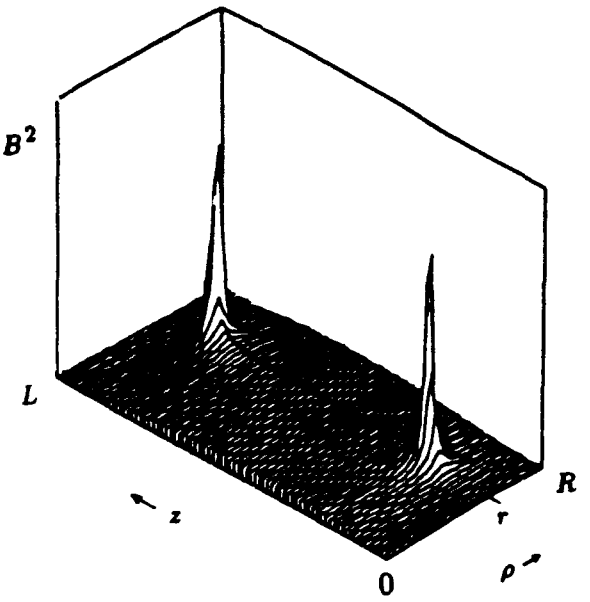


Figure 7: Plot of the field energy in the $r - z$ plane. The peaks in the energy distribution are located at the upper and lower edges of the electrodes. The relative magnitude of the fields in figure 6 are nearly too small to be seen on the scale appropriate to the peak field values.

ground plane is given by

$$\sigma(z) = \frac{\sigma_o}{\pi\sqrt{1-(2z/l)^2}}, \qquad (2)$$

where l is the width of the strip conductor and z is the point along the strip where the charge is evaluated. In this solution the current distribution is infinite but integrable at the conductor edges. A more realistic solution is found by using a Green function method [11]. The Green function and conformal mapping solutions are plotted in figure 5. Both of these solutions have the same general behavior; uniform current distribution in the center of the strip, and a sharp increase at the edges. For the shielded stripline, the edge currents on the strip are further increased by the existence of images in the shield wall [10]. This is similar to what occurs in electrostatic image problems.

A three dimensional finite difference code, called MAFIA, was used to investigate the fields in the loop-gap resonator. Plots of the magnetic and electric fields from MAFIA confirm the qualitative field descriptions reported in other loop-gap papers. As expected, the electric field is concentrated in the gap regions, and the magnetic fields are generally in the axial direction within the interaction region of the maser cavity. Figure 6 is a plot of the rf magnetic field in the center of the cavity ($r = 0$) from $z = 0$ up to the shield-can endplate, where $z = L/2$. This axial profile is very similar to the static fields of a dc solenoid and its image which is required to satisfy the boundary conditions at the endplates. However, in the regions near the electrodes, differences in the solenoid loop-gap comparison arise from the high frequency boundary conditions on the electrode surfaces. The current is distributed to satisfy these boundary conditions as shown in figure 5. Figure 7 shows the profile of magnetic field energy across the r, z plane, where $0 < r \leq R$. The enhancement of the magnetic field energy at the edges of the electrodes is as expected, and is in agreement with the stripline analogy.

The filling factor and the distribution of power losses in the resonator cannot be accurately calculated without knowing the fields. The filling factor is defined as

$$\eta' = \frac{< H_z >^2_{bulb}}{< H^2 >_{cavity}} . \qquad (3)$$

This is the ratio of the energy that stimulates the hydrogen transition to the total energy in the cavity. The approximation that the fields are uniform in the region of the electrode edges produces an optimistically large value of η'. We obtained a value of 0.39 for η' using the uniform field approximation and a value of 0.30 using the results of the finite difference code.

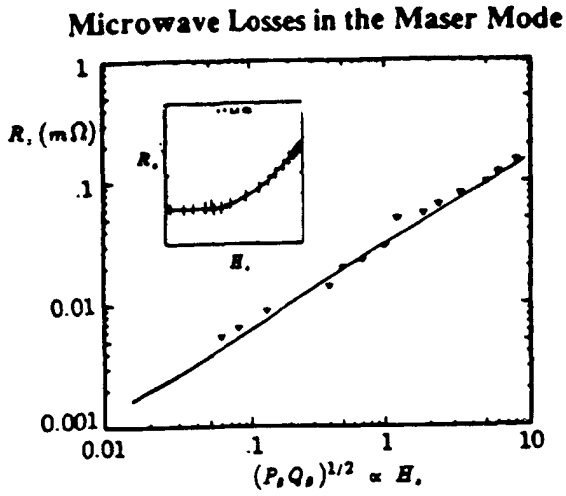


Figure 8: The surface resistance measured for the superconducting electrodes as a function of the maximum rf magnetic field on the surface of the electrodes. The expected behavior is shown in the inset indicating that at low input powers the YBCO behaves as if there is a large surface field.

The current distribution on the electrodes and on the surfaces of the shield-can vary in the same qualitative manner as in the shielded stripline. As the edges of the electrode conductor are moved closer to the shield surfaces, the currents increase due to the image currents on the shielding surfaces [11]. This change in the current distribution increases the local field energy and thereby reduces the fraction of total magnetic energy in the interaction region. The effect that this has on Q_c and η' has been studied numerically. The product $\eta' Q_c$ has a maximum where the electrode radius is about half of the shield-can radius and the electrode length is about 80% of the shield-can length.

Measuring the surface resistance in the YBCO superconducting resonator is complicated by the electrode current distribution. It has been shown that the R_s of polycrystalline YBCO is dependent on the magnitude of the rf magnetic fields at the surface [14].

The expected behavior for a cavity that does not have the unique field and current distribution shown in figure 5 is plotted as an inset in figure 8. R_s remains fairly constant with increasing surface field up to a critical value, then R_s increases sharply with further increases in the field. Figure 8 indicates that for a broad range of rf power levels, some part of the superconducting surface in the loop-gap resonator is experiencing field values beyond this critical value, even though the input power to the cavity is as low as $-30dbm$. This effect makes the characterization of the resonator surface difficult but should not cause a problem at the $-90dbm$ levels expected in an oscillating maser.

In the preceding arguments, the frequency has been assumed low enough that the wavelength is much larger than the dimensions of the strip. In this quasi-static approximation the fields are TEM and the only component of current is the azimuthal component, I_ϕ. The effect of operating at a frequency where the wavelength is on the order of the strip dimensions is that the longitudinal component of current, I_z becomes significant. It has been suggested that the form of this component is [10]

$$I_z(z) = \begin{cases} I_{z0} \sin\left(\frac{\pi z}{0.7l}\right), & |z| \leq 0.8\frac{l}{2}; \\ I_{z0} \cos\left(\frac{\pi z}{0.2l}\right), & 0.8\frac{l}{2} < |z| \leq \frac{l}{2} \end{cases} \tag{4}$$

where l is the full strip width. The magnitude of I_{z0} is proportional to the normalized strip width l/λ, where λ is the wavelength corresponding to the resonant frequency. When $l/\lambda < 0.1$, the average current amplitude I_{z0} across one half of the strip width is less than ten percent of the average $I_\phi(z = 0)$ current amplitude. For the cavity built at the University of Wuppertal, the condition that $l/\lambda < 0.1$ is not satisfied. Specifically, for an operating frequency of 1.42GHz (corresponding to a $21cm$ wavelength) and an electrode length of $5cm$, this cavity is not operating within the quasi-static limit. This would imply that in the maser resonator, the currents no longer flow only azimuthally around the electrodes but that there is a component of the current in the longitudinal (z) direction.

Shown in figure 9 is a MAFIA plot of $B_\phi(z)$ at the surface of the conductor compared with the function in equation (4). The shape of this curve is similar to that expected from a stripline where $l > 0.1\lambda$. The numerical solution confirms that the transverse magnetic fields are much smaller ($< 0.01\%$) than the longitudinal magnetic field component over most of the interaction region; therefore, this will not be a serious problem in our design. However, if the overall size of the loop-gap cavity is increased the transverse fields would become significant.

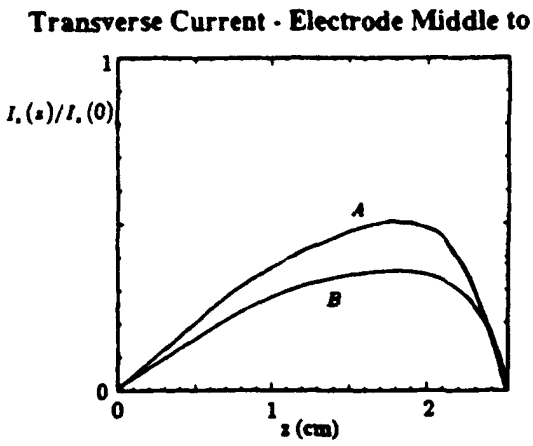


Figure 9: The component of the rf currents on the electrodes that are perpendicular to the direction of the TEM currents. The upper curve, A, is from the MAFIA calculation and the lower curve, B, is from the relationship given in reference [10].

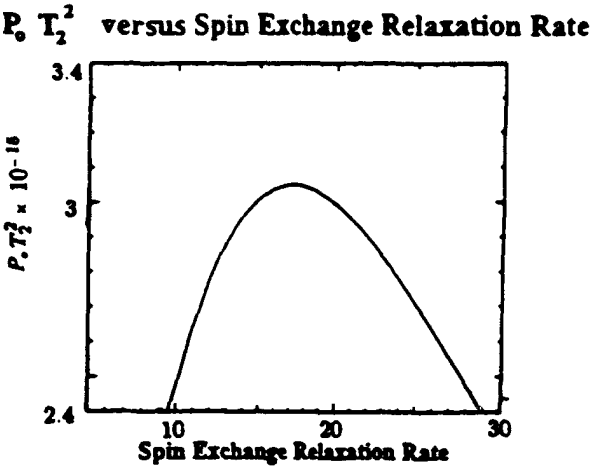


Figure 10: The product of PT_2^2 as a function of $\gamma_{1s.e.}$ which is proportional to the density of atoms in the storage bulb.

Loop-gap Design Optimization

To find the loop-gap maser resonator design that minimizes the thermal noise contributions the Allan deviation for integration times between *100* and *10000* seconds, $\sigma_l(\tau)$, must be minimized. We must express σ_l in terms of geometric variables like the resonator length L and radius R as well as the cavity quality factor, Q_c. The bulb relaxation rate, γ_b, is treated as a parameter. The spin exchange relaxation rate, $\gamma_{s.e.}$, is removed from the expressions for σ_l by minimizing σ_l with respect to $\gamma_{s.e.}$. Using these simplifications and the relationships for the filling factor and the quality factor as functions of the cavity geometry, σ_l is minimized in terms of the cavity dimensions.

The component of the Allan deviation due to the thermal noise within the cavity is given by

$$\sigma_l(\tau) = \frac{1}{Q_l}\sqrt{\frac{kT}{2P\tau}} \,. \tag{5}$$

Where Q_l is the atomic line quality factor, k is the Boltzmann constant, T is the temperature inside the cavity, and P is the power delivered by the atomic beam to the cavity. The Allan deviation will be evaluated at $\tau = 1 second$.

Equation (5) can be rewritten as

$$\sigma_l(\tau) = \sqrt{\frac{kT}{2\pi^2 f_o^2}}\sqrt{\frac{1}{PT_2^2}} \tag{6}$$

where $Q_l = \pi f_o T_2$, and T_2 is the transverse decay time of the atoms. The first term under the radical is not geometry dependent. Therefore it is only necessary to maximize the product PT_2^2 for the thermal noise component to be minimized. Separating out terms that are constant for changes in cavity geometry and defining these terms as the constants A and B, we have

$$PT_2^2 = BV_b\left(\frac{\gamma_1}{\gamma_2}\right)\left(A\frac{\gamma_b\gamma_{1s.e.}}{\gamma_1\gamma_2} - \frac{1}{\eta' Q}\right) \tag{7}$$

$$A = \frac{\mu_o\mu_B^2 C_b}{\hbar\sigma v_r}$$

$$B = \frac{\hbar^2\pi f_o}{\mu_o\mu_B^2}$$

The total relaxation rates are defined as

$$\begin{aligned}\gamma_1 = \frac{1}{T_1} &= \gamma_b + \gamma_{1w} + \gamma_{1s.e.} \\ \gamma_2 = \frac{1}{T_2} &= \gamma_b + \gamma_{2w} + \gamma_{2s.e.}.\end{aligned} \tag{8}$$

In the above expressions, γ_1 and γ_2 are the total longitudinal and transverse relaxation rates, respectively. $\gamma_{1s.e.}$ is the spin exchange relaxation rate. As described in [4]:

$$\gamma_{1s.e.} = 2\gamma_{2s.e.} \qquad (9)$$
$$= n\sigma\bar{v}_r = n\sigma 4\left(\frac{kT}{\pi m}\right)^{1/2}.$$

Here, σ is the spin exchange cross section, $\bar{v}_r$ is the average relative atomic velocity, and n is hydrogen density given by

$$n = \frac{I_{tot}T_b}{V_b}. \qquad (10)$$

γ_{1w} and γ_{2w} are the longitudinal and transverse relaxation rates due to wall collisions. p_{1w} is defined as the transverse relaxation probability per collision and τ_c is the average time between atomic wall collisions[15]. With A_b defined as the surface area of the storage container, we have

$$\gamma_{1w} = \frac{p_{1w}}{\tau_c} \approx \frac{4}{3}\gamma_{2w}$$

$$p_{1w} = 7.3 \times 10^{-6}\exp\left(\frac{230}{T}\right) = 1.45 \times 10^{-4} \qquad (11)$$

$$\frac{1}{\tau_c} = \frac{\text{mean atomic velocity}}{\text{mean free path}} = \sqrt{\frac{8kT}{\pi m}}\frac{A_b}{4V_b}.$$

Since the spin exchange relaxation rate is proportional to the hydrogen density in the bulb, we can solve for the condition of optimal density. A plot of PT_2^2 as a function of $\gamma_{1s.e.}$ is shown in figure 10. By setting the derivative of PT_2^2 with respect to $\gamma_{1s.e.}$ equal to zero and solving for $\gamma_{1s.e.}$, we find the relation between PT_2^2 and the optimal value of $\gamma_{s.e.}$ or equivalently the optimal density. This condition on $\gamma_{1s.e.}$ is substituted into equation (7). This ensures that the product of PT_2^2 is always evaluated at the optimal hydrogen density within the bulb. It is found that $\gamma_{1s.e.\ opt}$ for optimal stability is given by

$$\gamma_{1s.e.\ opt} = \frac{2(\gamma_{2w}+\gamma_b)(2\gamma_{2w}-\gamma_{1w}+\gamma_b-2A\gamma_b\eta'Q_c)}{\gamma_{1w}-2\gamma_{2w}-\gamma_b-2A\gamma_b\eta'Q_c} \qquad (12)$$

We assume that the bulb escape rate γ_b can be adjusted by changing the collimator tube length. All other variables, such as the filling factor, the bulb volume, the atomic collision rate with the walls, and the quality factor, Q_c, are dependent on the particular cavity geometry. Since no analytical solution exists for the fields in the cavity there are no analytical solutions for Q_c and η'. Therefore a numerical optimization has been performed by choosing a shield-can geometry

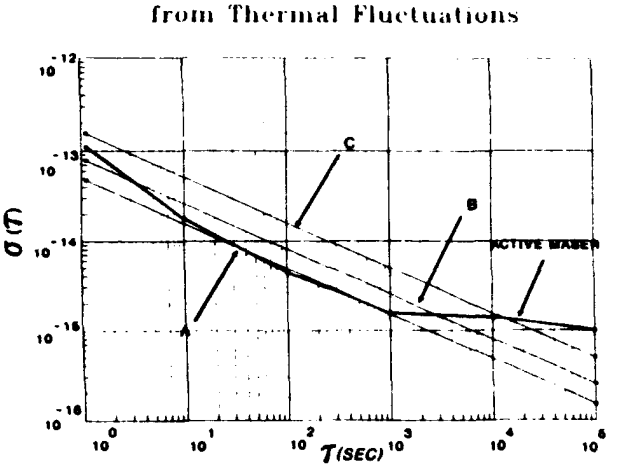


Figure 11: The Allan deviation as a function of the integration time for three different cavity geometries: (A) optimal design according to the electrode geometry for maximum PT_2^2, (B) the cavity shown in figure 1, (C) cavity designed for maximum $\eta'Q_c$. Also shown is the stability of a room temperature active maser[16]

and finding the electrode geometry that would give the best stability in that case. In order to find the optimal resonator design, many iterations of this process are needed. When PT_2^2 is considered as a function of the electrode radius and length for a given shield can geometry, there is a maximum that implies an optimal design for the electrodes. The location of this maximum is dominated by the collision rate of the atoms with the storage bulb walls. It is clear that the wall collision rate is minimized for a large bulb volume to surface area ratio.

The Allan deviation can now be calculated for any given loop-gap cavity design. The results of this calculation are shown, for three cases, in figure 11. A cavity geometry designed using the maximum $\eta'Q_c$ criteria will have a narrow storage bulb and therefore a large wall relaxation rate. This is represented by curve A in figure 11. Curve B describes the predicted level of thermal fluctuations for the cavity shown in figure 1. Curve C represents the minimum fluctuations for a cavity with a shield-can diameter and length equal to $7.5cm$ and electrodes $6.8cm$ in length and $3.2cm$ in radius. Shown for comparison is actual data from a standard active maser[16]. The level of fluctuations are reduced by a factor of three in going from the $\eta'Q_c$ geometry to the PT_2^2 maximum design where $\sigma_l(1\ second) = 6 \times 10^{-14}$.

Conclusion

A technique developed for the deposition of textured thick films of YBCO onto silver substrates of arbitrary shape has been applied to the electrodes of the loop-gap resonator. The surface resistance of the electrophoretic YBCO electrodes in this maser resonator is shown to be less than $1m\Omega$ at $77K$ and 1420MHz. Maser oscillation can now be achieved in the compact resonator without the need for Q-enhancement techniques. The design of the resonator cavity has been optimized with respect to the contribution of thermal noise to the maser oscillation using the field solutions from a 3 dimensional finite element code (MAFIA). The thermal noise within the cavity limits the stability to $\sigma_y(10000\ seconds) = 6 \times 10^{-16}$. Figure 11 shows the calculated deviations compared with fluctuations measured in a room temperature active maser. The frequency fluctuations of the optimized compact superconducting resonator coincide with these measured values. This compact superconducting maser design has frequency stability competitive with full size room temperature masers at a substantial weight and size reduction making it attractive for space applications.

Acknowledgements

We would like to thank the Research Computation Division of the Naval Research Laboratory including Kay Howell, Rob Rosenberg and Upul Obeysekare for their assistance in the finding and use of 3D Electromagnetic codes for this study. The computer department at CEBAF was also invaluable in providing computing resources for this project. This work was supported by the Naval Research Laboratory within the framework of the HTSSE project and Interatom of Germany.

References

[1] H. E. Peters, "Small, Very Small and Extremely Small Hydrogen Masers", *Proc. of the* 32^{nd} *Annual Symp. on Freq. Control*, 469, 1978.

[2] Harry T. M. Wang, "Subcompact Hydrogen Maser Atomic Clocks",*Proc. of the IEEE*, Vol. 77, No. 7, July 1989, pp. 982-992.

[3] D. Kleppner, H. M. Goldenberg and N. F. Ramsey, "Theory of the Hydrogen Maser," *Phys. Rev.*,Vol. 126, No. 2, pp. 603-615 1962.

[4] D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters, and J. Vanier, "Hydrogen Maser Principles and Techniques," *Phys. Rev.* **A 138**, 972, 1965.

[5] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of Signal Sources and Measurement Methods", *Proc. of the* 35^{th} *Annual Symp. on Freq. Control*, pp. 1-47, 1981.

[6] E. M. Mattison and R. F. C. Vessot, "Physics of Systematic Frequency Variations in Hydrogen Masersy", *Proc.* 22^{nd} *Annual Precision Time and Time Interval Applications and Planning Meeting*, pp. 453-464, 1990 .

[7] M. Hein, S. Kraut, E. Mahner, G. Mueller, D. B. Opie, H. Piel, L. Ponto, D. Wehler, M. Becks, U. Klein, M. Peiniger, "Electromagnetic Properties of Electrophoretic $Y_1Ba_2Cu_3)_{7-\delta}$ films", J. Supercond. Vol. 3, No. 3, p. 323, 1990

[8] G. B. Collins, *Microwave Magnetrons*, York, PA., Maple Press Co., 1948.

[9] M. Mehdizadeh and T. Koryu Ishii, "Electromagnetic Field Analysis and Calculation of the Resonant Characteristics of the Loop-Gap Resonator", *IEEE Trans. Micro. Theory and Tech.*, Vol. 37, No. 7, 1989.

[10] E. J. Denlinger, "A Frequency Dependent Solution for Microstrip Transmission Lines", *IEEE Trans. Micro. Theory and Tech.*, Vol. 19, No. 1, 1971.

[11] T. G. Bryant and J. A. Weiss, "Parameters of Microstrip Transmission Lines and of Coupled Pairs of Microstrip Lines", *IEEE Trans. Micro. Theory and Tech.*", Vol. 16, No. 12, 1968.

[12] H. A. Wheeler, "Transmission-Line Properties of Parallel Strips Separated by a Dielectric Sheet", *IEEE Trans. Micro. Theory and Tech.*, Vol. MTT-13, No.2, pp. 172-185, March 1965.

[13] J. C. Maxwell, *A Treatise on Electricity and Magnetsim*, 3^{rd} ed., vol. 1, New York: Dover, 1954, pp. 296-297.

[14] G. Mueller, et al., IEEE Trans. Mag., MAG-25,p.2402, 1989.

[15] M. Desaintfuscien, J. Viennet and C. Audoin, "Discussion of the Temperature Dependence of Wall and Spin Exchange Effects in the Hydrogen Maser", *Metrologica*, vol. 13, p.125, 1977.

[16] H. E. Peters, H. B. Owings, and P. A. Koppang, "Hydrogen Masers with Cavity Switching Servos", *Proc.* 22^{nd} *Annual Precision Time and Time Interval Applications and Planning Meeting*, pp. 283-292, 1990 .

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH-TC SUPERCONDUCTING MICROSTRIP RESONATORS AND FILTERS ON $LaAlO_3$

June H. Takemoto, Charles M. Jackson, Roger Hu, Harold M. Manasevit, Claire Petite-Hall, John F. Burch, Dennis C. St. John, Alfred Lee, Ken P. Daly

TRW Space and Technology Group, One Space Park, Redondo Beach, CA 90278

Abstract

High-Tc superconducting (HTS) thin films increase the Q of microstrip by more than a factor of 50 at 1 GHz, and at 10 GHz by more than a factor of 20. This increase in Q enables extremely high Q resonators to be developed for low noise oscillators. This paper will discuss recent developments in high Q microstrip resonators and filters using various deposition processes.

Introduction

The extremely low surface resistance in HTS thin films facilitates the development of microwave devices with better performance than conventional devices. In radar applications, HTS filters and resonators provide much lower attenuation at X-band microwave frequencies than microstrip circuits and SAWs fabricated with normal metal [1]. HTS microstrip resonators have been used to stabilize oscillators [2].

Delay lines are an important component for many radar systems. The combination of low surface resistance and fixed skin depth allow very compact, low velocity delay lines to be manufactured from superconducting microstrip. Figure 1 compares the insertion loss for a few transmission delay lines with a 10 ns delay. For comparison, transmission lines of copper and HTS are compared on the same substrate. Superconducting MMIC refers to microstrip lines formed on a substrate with a layer of HTS followed by 300 nm layer of insulator, followed by an HTS transmission line.

Coaxial lines have too high a loss for many applications. SAW devices work over a wide range of delay times and frequencies, so the region marked for the SAW devices is only for comparison. At low frequencies the HTS delay lines have negligible loss, and even at 10 GHz, the losses will be very low.

Many techniques have been developed for the deposition of HTS films. Early work concentrated on sputtering techniques which have resulted in high quality films [3]. More recently, pulsed laser deposition has provided high quality films. A new and interesting technique

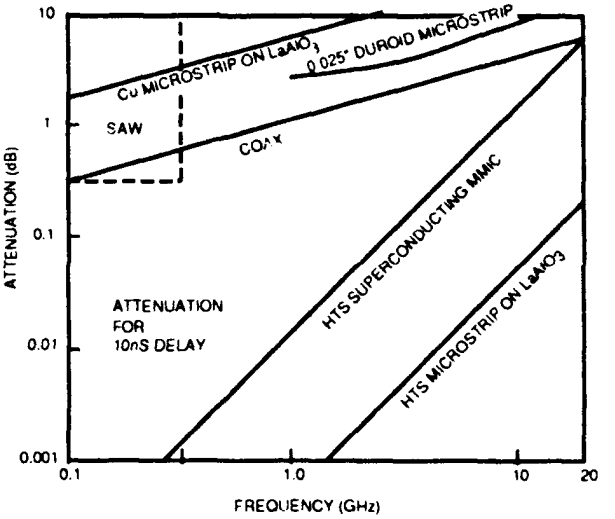


Figure 1. Insertion loss for candidate delay lines at different frequencies.

metal-organic vapor deposition (MOCVD) has been developed at TRW to produce high quality films on both sides of the substrate [4].

HTS Film Growth

Sputtered Thin Films

Early microstrip samples were fabricated by sputtering YBa_2Cu_3O on $LaAlO_3$ substrates [3]. HTS films were sputtered onto whole 1.5-inch diameter $LaAlO_3$ wafers using a single target of YBCO with a stoichiometric 1-2-3 compound. The substrate was thermally contacted to the platform via silver paste and heated to 700 - 750 C in an atmosphere of 30 - 50 mT of oxygen and 40 - 200 mT of argon. Off-axis sputtering minimized damage caused by negative ion bombardment. RF power of 125 W provided a film growth rate of 700 Å/hr. Immediately after deposition, the samples were cooled in 600 T of pure oxygen. The high pressure of oxygen eliminates the need for additional thermal annealing; thus, smooth films suitable for device fabrication can be deposited on substrates.

After removing the baked silver paste, Ag was deposited on the backside of $LaAlO_3$ to provide a ground plane for the microstrip structures. The HTS was patterned with standard photo-

CH2965-2/91/0000-477 $1.00 © 1991 IEEE

lithographic techniques and etched with dilute phosphoric acid. Ag or Au contacts were deposited on some of the circuits to improve contacts between K-connector pins and microstrip lines. Completed wafers were diced into 0.2 in. x 0.5 in. pieces by a diamond saw. Film deterioration was significantly reduced by shortening the processing time, minimizing the exposure of the HTS solvents and other chemicals, and avoiding ion milling directly on the surface of HTS.

The resultant films yield dc critical temperature of 86 - 91 K with transition widths of 0.5 K. The critical current of the HTS films is 2×10^7 A/cm^2 at 4.2 K and 1×10^6 A/cm^2 at 77 K. Microwave response was superior to copper below 81 - 85 K, indicating that the films were of high quality.

Laser Deposited Films

High quality YBCO films are routinely deposited by laser ablation. A KrF excimer laser (Lambda Physik 315i), operating at 4 Hz, is focused to roughly 2 J/cm^2 onto the rotating YBCO target. The chamber pressure is held at 200 mTorr of O_2, and the substrate heater is kept between 770-780 C. Large are depositions are achieved by rastering the substrate heater over a 2-inch square area as the laser is firing. Upon completion of a deposition, the films are slowly cooled to room temperature in a 600 Torr atmosphere of O_2.

The resulting films exhibit sharp dc resistance transitions (Tc > 90 K) and sharp magnetization transitions. High current densities of approximately 3×10^6 A/cm^2 are achieved at 77 K and these increase to 10^7 A/cm^2 below 65 K. The films typically exhibit low microwave resistance which starts at approximately 160 μΩ at 4 K and rises to 500 μΩ around 77 K.

Step edge HTS microbridges are fabricated over sharp steps etched in the $LaAlO_3$ substrate. A steep (90 degree edge) 0.3 μm trough is ion-milled into the substrate, and HTS film (0.2 μm) is deposited over the entire area. Large area depositions are achieved by rastering the substrate heater over a 2-inch square area during the laser deposition. The film is then patterned and ion-milled to produce a 2 μm wide microbridge over the step edge, resulting in an engineered weak link Josephson element. Although there are two step edges in series, only the one with smaller critical current dominates. Over 90 % of the films yield Tc values greater than 87 K at microwave frequencies.

MOCVD Thin Films

A more interesting process of fabricating very low microwave losses in microstrip resonators has been developed using MOCVD which deposits high-temperature EBCO superconducting films on both sides of a $LaAlO_3$ substrate. For the thermal contact required for growth in low pressure sputtering and laser depostion systems, the backside of the substrates are coated with silver paste. However, the silver paste is subsequently difficult to remove without affecting the YBCO film on the reverse side. Silver paste is not needed in MOCVD. This process provides thermal contact of the substrate to the susceptor by the high pressure gaseous atmosphere used for film growth.

EBCO films were grown in-situ by MOCVD on both sides of polished $LaAlO_3$ using the appropriate metal β-diketonate complexes of $Er(C_{11}H_{19}O_2)_3$, $Ba(C_{11}H_{19}O_2)_2$, and $Cu(C_{11}H_{19}O_2)_2$. A carrier gas was used to transport the metalorganic precursors and mix them at about 250 C with O_2. The precursors were then introduced into the chamber where the $LaAlO_3$ substrates sit. The films used for the resonators were grown at atmospheric pressure (in He) at 800 C, as measured with an optical pyrometer focused on the side of the susceptor. The EBCO film thickness of 5500 Å on the 20-mil substrate exceeds the penetration depth of HTS. A second film was then grown on the exposed polished backside of the $LaAlO_3$ wafer as described above. The annealing time of the second film at 400 C in He/O_2 was extended to 30 minutes in an attempt to replace the O_2 presumably lost from the first film while being heated at 800 C.

These films perform well at microwave frequencies even though scanning electron microscope (SEM) photographs show inhomogeneity at 10K magnification. Resistance-temperature, R(T), measurements showed sharp transition widths of 1-2 degrees, $T_c(0) = 92$ K, and an extrapolated intercept for the slope of R(T) above T_c nearly equal to T = 0 K. R(T) measurements of both films produced essentially overlapping curves.

Microstrip Design

Numerous structures exist for microstrip resonators and filters. Our $LaAlO_3$ studies include straight 50 Ω linear resonators, closed-loop ring resonators, meanderline resonators, and parallel-coupled, series-gap bandpass filters. Figure 2 shows most of the microstrip circuits reported in this paper. The dielectric constant ε_r of $LaAlO_3$ was determined from these resonator measurements to be 26. $LaAlO_3$ was chosen as a substrate because the lattice match to YBCO

provides high quality films, and its ε_r is suitable to microwave applications while the ε_r of $SrTiO_3$ is not.

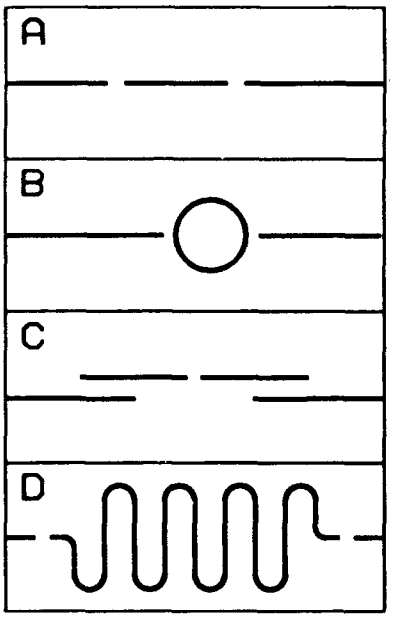


Figure 2. Microstrip circuits: a) 10 GHz linear resonator, b) ring resonator, c) 2-pole filter, and d)1.3 GHz meanderline resonator

The 50 Ω linear resonators were designed for 5 and 10 GHz. The low impedance resonator was designed for 10 GHz with an impedance transformer from 50 to 25 Ω. The coupling gap of the 50 Ω linear resonators was varied to investigate the external loading on the resonance and determine the minimum R_S of the HTS films. Resonators with 8, 14, and 20 mil gaps were calculated to have total loaded Q's of about 340, 600, and 1025, respectively.

Ring resonators were designed at 15 GHz with a ring impedance of 40 Ω and coupling gap size of 8. The ring circumference for the 40 Ω ring was increased to the second harmonic wavelength to decrease the ratio of the width to the circumference of the ring. Larger rings minimize field interactions across the ring width and distortion of the field around the coupling gap. The drawback of the larger rings is the increased possibility of distorting the uniformity of the ring during fabrication, leading to resonance splitting. The size of the rings was limited to an area of 0.2 in. x 0.5 in. because of the test fixture design. A 10 GHz, 50 Ω ring resonator with a 14 mil gap was also tested. These designs were optimized from the previously reported ring resonator design [5].

We tested our superconducting microstrip in a practical circuit design using a 2-pole, 0.1 dB ripple, 1 percent bandwidth Chebyshev filter [6]. Commercial software (TOUCHSTONE) optimized the design. The ε_r of 26 for $LaAlO_3$ is outside the range of ε_r that this model can accurately simulate. So some feedback from experimental measurements of actual filters was needed to optimize the design. The variety of bandpass filter configurations typically involve either end-coupled or parallel-coupled structures. Since microstrip gaps are not as well modelled as coupled lines, a compromise was made in the filter design. The filter design is parallel coupled in the outer sections and end coupled in the inner sections.

We calculated the effects of kinetic inductance on the design of the resonators and filters. The kinetic energy of the superconducting electron pairs generates inductance, which slows the propagation velocity in the microwave circuit. The larger penetration depth of HTS compared to LTS and the high ε_r of $LaAlO_3$ increases the kinetic inductance effect, producing larger frequency shifts. At T/Tc = 0.5, the kinetic inductance of Nb exhibits a change of 0.7 MHz, while the HTS produces a change of 9.4 MHz in the same structure. Frequency shifts in the response can also be caused by the change in dielectric constant of the substrate material due to thermal contraction. However, the estimated frequency shift due to this effect is only 0.15 MHz for a 10 K change in temperature. Deviations in the substrate thickness from the nominal 20 mil value contribute significantly to frequency shifts. A 10 μm (0.4 mils) change in the microstrip linewidth, which is proportional to a 36 μm (1.4 mil) change in the substrate thickness, causes a 44 MHz shift in frequency. Therefore, substrates with the same thickness should be selected to accurately incorporate kinetic inductance effects into the design of HTS filters and resonators.

Measurements and Results

Sputtered Single-sided Devices

HTS linear microstrip resonators demonstrated Q's greater than 1200 at 10 GHz, corresponding to surface resistances less than 300 μΩ. Identical silver resonators showed Q's of 60 and surface resistance of 15 mΩ. The high frequency transition temperature for these HTS films was greater than 83 K. We measured Q's of 240 in ring resonators at 15 GHz.

We also fabricated and tested a 2-pole, Chebyshev narrow-bandwidth bandpass filter. The reduction in surface resistance offered by HTS allows a reduction in insertion loss for filter applications (Figure 3). The HTS filter was designed to have a 1 percent bandwidth

Chebyshev response with 0.1 dB ripple. Insertion loss was 2 dB at 4 K, rising to 3 dB at 77 K. In comparison, the insertion loss of a similar silver filter was 24 dB. This is 20 dB better than a simialr filter fabricated in Ag.

We observed temperature dependence in the filter center frequency, which we attribute to kinetic inductance effects. Our HTS filters outperform similar cryogenic silver filters, indicating that practical levels of HTS performance have been achieved.

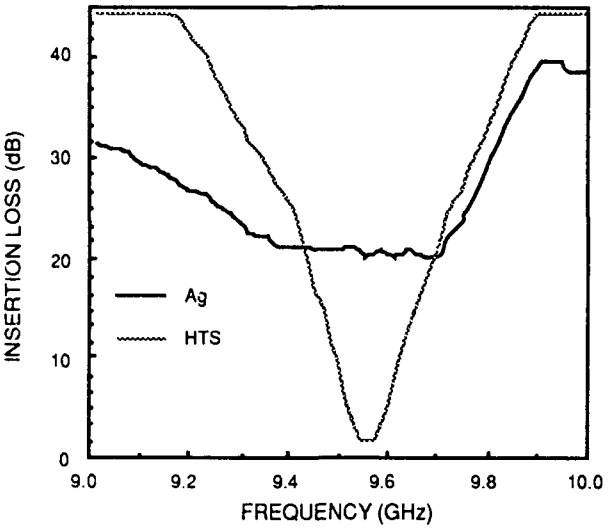


Figure 3. Filter response of HTS versus silver microstrip circuit.

Two-Sided MOCVD Devices

We have demonstrated Q-values of MOCVD films that are twice those produced by single-sided processes. For higher frequency microstrip circuits, thinner substrates are necessary, but this increases the ohmic loss. This implies that MOCVD can give an order of magnitude improvement due to its double sided film growth capabilities. These *in-situ* films have critical temperatures of 92 K. HTS linear microstrip resonators demonstrated Q's of 3500 at 10 GHz and 4.2 K, corresponding to surface resistances less than 330 μΩ, which is 45 times better than measured silver resonators (Figure 4). HTS meanderline resonators with a fundamental frequency of 1.3 GHz have Q's of 10,000 at 4.2 K. Identical silver meanderline circuits have 300 times more surface resistance than HTS circuits which are typically 16 μΩ.

The Q values obtained for samples with MOCVD HTS on both sides of $LaAlO_3$ were compared to those with sputtered HTS on one side and Ag as the ground plane. The Q's of two-sided HTS resonators (3600) on 20-mil thick substrates were twice that of the one-sided resonators (1500). For 5-mil thick substrates, the two-sided resonators will be 8 times better than the one-sided resonators since the silver ground plane will contribute more losses and inhibit the microwave performance of the circuit.

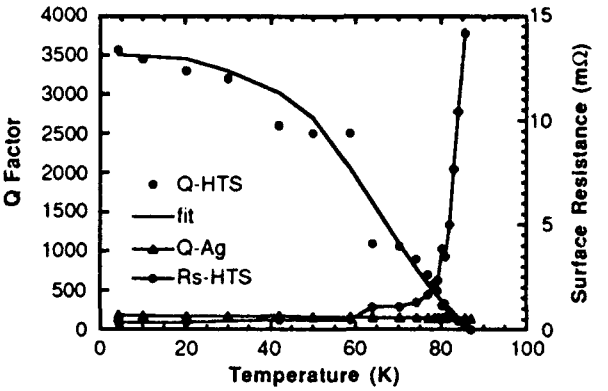


Figure 4. Q and Rs as a function of temperature for the two-sided 10 GHz linear resonator.

Values of surface resistance, R_S, were calculated [5,7] from the measured Q. Note that unlike the single-sided samples, the R_S of HTS films can be found directly from the measured Q of HTS. Surface resistance of the linear resonators at 4.2 K was 328 μΩ, and at 77 K was 1.66 mΩ. The lowest R_S that we measured at 10.4 GHz is shown in Figure 4 for a 20-mil linear resonator. The effective surface resistances of our patterned single-sided HTS films are less than 500 μΩ, while double-sided resonators had R_S less than 330 μΩ. The latter value is more than 10 times better than Cu.

The meanderline resonator fabricated from double-sided HTS samples had the highest Q of the measured microstrip circuits, equal to 9600 at 1.3 GHz, compared to about 200 for a Ag meanderline resonator of identical dimensions (Figure 5). Therefore, HTS had at least a factor of 45 better microwave performance than Ag. Currently the coupling gap is 14 mils; hence, larger coupling gaps may show even larger Q's. In terms of surface resistance, the R_S of HTS and Ag was 16 μΩ and 5 mΩ, respectively at 1.3 GHz and 4.2 K, which corresponds to the HTS being 300 times better than Ag.

Since the fundamental frequency is 1.3 GHz, we were able to measure several harmonic resonances. The Q was 5700 at 4.9 GHz, and 2900 at 11.3 GHz. After calculating the surface resistance on all the measured resonances, we plotted R_S against the square of the frequency (f^2)

and found that these HTS films followed the expected f^2 dependence.

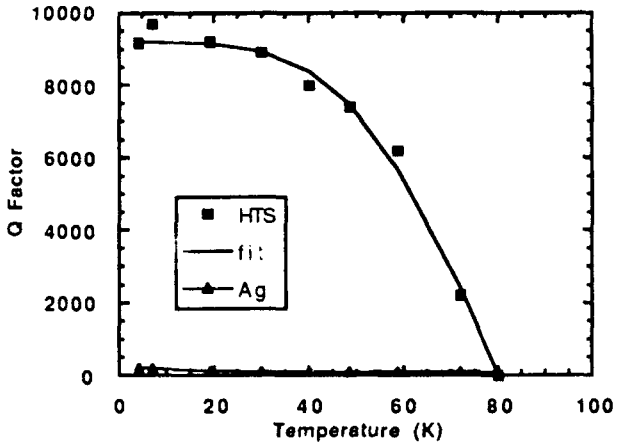


Figure 5. Q of the MOCVD 1.3 GHz meanderline resonator with HTS ground plane.

Conclusions

We have successfully developed sputtering, MOCVD, and laser deposition processes that produces high quality HTS films on $LaAlO_3$. For the double-sided MOCVD process, replacing the Ag ground plane with an HTS ground plane in silver ground plane eliminates extra surface losses due to normal metal ground conductors. These two-sided microwave circuits yield higher microwave performances than one-sided HTS circuits and much higher performance than identical Ag circuits.

We achieved HTS surface resistances less than 330 $\mu\Omega$ at 10 GHz. Q's of 9600 at 1.3 GHz with R_S of 16 $\mu\Omega$ were achieved in meanderline resonators. For thinner substrates, MOCVD can provide an order of magnitude improvement compared to a one-sided process with an HTS transmission line with a silver ground plane. The extremely low microwave surface resistance of HTS will greatly improve the performance of oscillators in radar applications.

References

[1] C. M. Jackson and L. J. Cantafio, "High Temperature Superconductors for Radar Applications," presented at the 1991 IEEE National Radar Conference, Los Angeles, CA, March 12-13, 1991, pp. 122-126.

[2] R. Babbitt, E. Lenzing, W. Wilbur, and A. Rachlin, "Fabrication and Evaluation of Superconducting Devices," *Microwave Journal*, pp. 40-48, April 1991.

[3] A. E. Lee, C. E. Platt, J. F. Burch, and R. W. Simon, "Epitaxially Grown Sputtered LaAlO3 films," *Appl. Phys. Lett.*, vol. 57, no. 19, pp. 2019-2021.

[4] J. H. Takemoto, C. M. Jackson, H. M. Manasevit, D. C. St. John, J. F. Burch, K. P. Daly, R. W. Simon, "Microstrip Resonators Using Two-sided MOCVD EBCO Thin Films," *Appl. Phys. Lett.*, vol. 58, no. 10, pp. 1109-1111.

[5] J. H. Takemoto, F. K. Oshita, H. R. Fetterman, P. Kobrin, E. Sovero, "Microstrip Ring Resonator Technique for Measuring Microwave Attenuation in High Tc Superconducting Thin Films," *IEEE Trans. Microwave Theory and Tech.* , vol. MTT-37, pp. 1650-1652, Oct. 1989.

[6] G. L. Matthaei, L. Young, and E.M.T. Jones, Microwave Filters, Impedance-Matching Networks, and Coupling Structures, Aertech House, Dedham MA, 1964.

[7] R. A. Pucel, D. S. Masse, and C. P. Hartwig, "Losses in Microstrip," *IEEE Trans. Microwave Theory and Tech.* , vol. MTT-16, pp. 342-350, June 1986; Correction, vol. MTT-16 pp. 1064, Dec. 1968.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SUPERCONDUCTING STABILIZED X-BAND OSCILLATORS

Roland Cadotte, Adam Rachlin, Richard Babbit, and William Wilber

U.S. Army Electronics Technology and Devices Laboratory
LABCOM, Ft. Monmouth, NJ 07703-5601

ABSTRACT

Two X - Band oscillators have been designed, built and tested using planar HTSC superconducting resonators for stabilization. The reaction type oscillator uses a YBaCuO HTSC ring resonator and the parallel feedback oscillator uses a thalium based HTSC half wavelength resonator with an unloaded Q = 21000. The design and development of these oscillators and charactersistics of HTSC resonators will be discussed.

INTRODUCTION

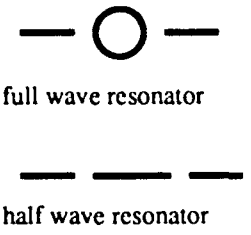
Superconducting films of both $Y_1Ba_2Cu_3O_x$ (YBCO) and $Tl_2Ba_2Ca_1Cu_2O_y$ (TBCCO) were used for this work. The 123 films were grown at ETDL by laser ablation. The laser ablation system is typical [1] in that 248 nm light from a pulsed excimer laser is focused onto a bulk target of YBCO so that the ablated material is deposited onto a substrate positioned 7 cm from the target. In this particular case, the substrate is a single crystal of MgO heated during deposition to a temperature of 720 C. The ambient atmosphere during deposition is 200 mTorr of oxygen and is raised to approximately one-half an atmosphere during the post-deposition cooling phase. The depostion rate is typically 550 nm/minute and the YBCO film used here was 800 nm thick.

The TBCCO films used for these experiments were also deposited by laser ablation but were obtained commercially from Superconductor Technologies, Inc. The deposition process for these films is somewhat similar to that previously described, although a post-deposition anneal in a thallium atmosphere is required. The substrate material for the Tl-based film was single-crystal $LaAlO_3$. The two different substrate materials have considerably different dielectric constants. [1] [2] At 77 K, the dielectric constant for MgO is 9.6 at 10 GHz, but is 23.8 at 6.2 GHz for $LaAlO_3$. The loss tangent is about than 6×10^{-5} for both materials. Thalium based films have transition temperatures of approximately 110K, while YBaCuO films transition at

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

approximately 90K.

A resonator was patterned in each of the HTSC films previously described to stabilize two X -band oscillators. A ring resonator was patterned in a YBaCuO HTSC film and a half wavelength resonator was patterned in the thalium based HTSC film. The signal is capacitively coupled in and out of the resonators, (see figure 1).



full wave resonator

half wave resonator

Figure 1.

The unloaded Q of the thalium based resonator and the YBaCuO resonator are 21,000 and 1,200 respectively, measured with approximately -25dBm at the resonator. The unloaded Q's of the resonators are calculated from their loaded Q's and their insertion losses at resonance. [3], [4] The high Q of the thalium HTSC ring is attributable to a very high quality HTSC film and also having an HTSC ground plane, whereas the YBaCuO resonator has a silver ground plane. It is believed that a double sided HTSC substrate can increase the Q of a resonator by an order of magnitude. Both of these HTSC resonators exhibit similar characteristics versus temperature. As temperature decreases the surface resistivity of the HTSC decreases, resulting in a higher Q and thus a lower insertion loss for a constant amount of coupling in and out of the resonator. Actual meaurements of insertion loss and Q versus temperature for the double sided Thalium resonator show that the insertion loss of the resonator decreases from 27dB at 93.6K to 15.6dB at 46.0K, (see figures 2, 3).

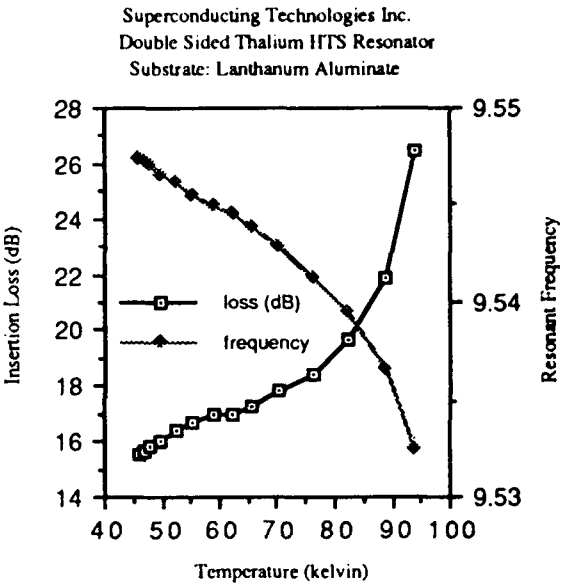


Figure 2.

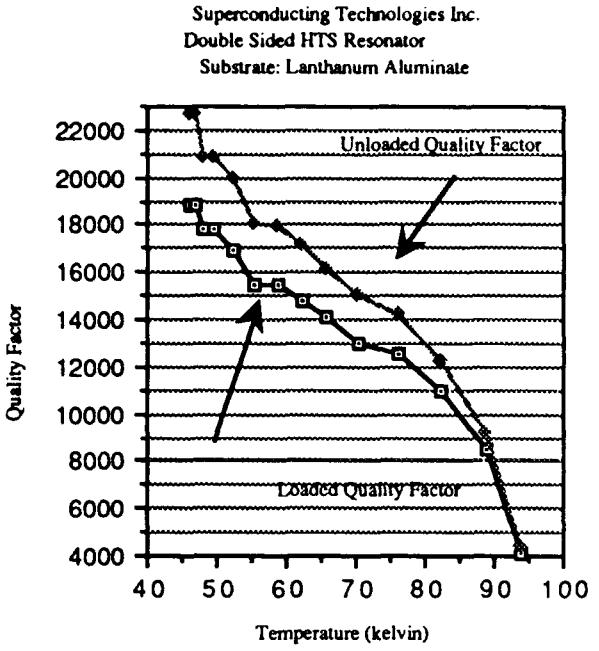


Figure 3.

A second characteristic which changes as a function of temperature is the ring's resonant frequency. This is explained by the following relation. [5]

1) $f = v_{ph}/\lambda_c$

2) $v_{ph} = 1(\mu_o\varepsilon_{ff})^{1/2}[1+(2\lambda_{(T)}/d)\coth(b/\lambda_{(T)})]^{-1/2}$

3) $\lambda_{(T)} = \lambda_{(0)}[1-(T/T_c)^4]^{-1/2}$

$\lambda(T)$ = Penetration Depth
T = Resonator Temperature
T_c = Transition Temperature of Superconducting Film
f = resonant frequency
v_{ph} = phase velocity
λ_c = wavelength
ε_{ff} = effective dielectric
b = film thickness
d = substrate thickness

Measurements show that the thalium based HTSC resonator's frequency changes 20MHz as the ring's temperature decreases from 94K to 46 K, (see figure 4). The design and results of two X band oscillators using these superconducting resonators for stabilization are represented.

HTSC THALIUM BASED RESONATOR'S RESONANT FREQUENCY VS. TEMPERATURE

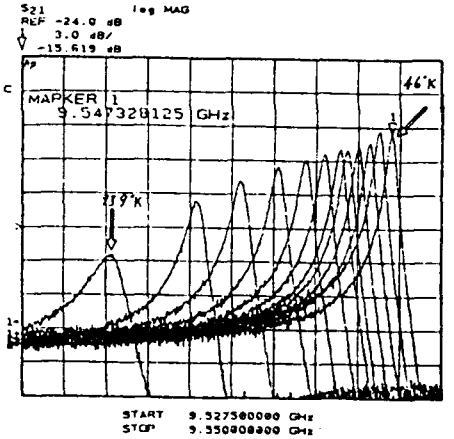


Figure 4.

SUPERCONDUCTING OSCILLATORS

REACTION TYPE OSCILLATOR

The reaction type oscillator is composed of two separate and different substrates, an

alumina substrate and a magnesium oxide (MgO) substrate (see Figure 5).

REACTION TYPE OSCILLATOR

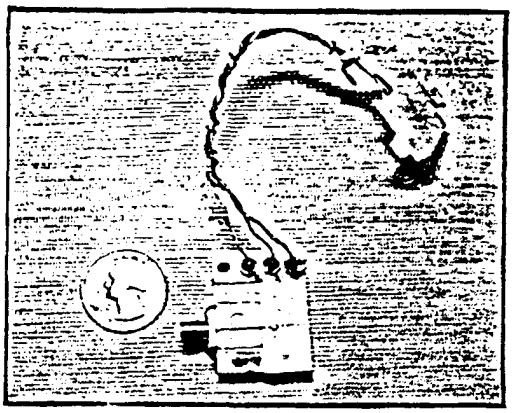


Figure 5.

The alumina substrate is metallized with gold on both top and ground planes, and the MgO substrate is metallized with silver on the ground plane and with a YBaCuO superconducting film deposited on the other plane. Both substrates are soldered to a brass test fixture with a low temperature solder . The low temperature solder is precautionary and meant to reduce stress on the superconductor film. The active part of the oscillator is located on the alumina subtrate. A NEC 73183 HEMT is attached to the gold microstrip pattern that includes DC bias circuitry and RF stubs designed to induce oscillation. The MgO substrate has a ring resonator and a 50 ohm line etched into its superconducting film. The 50 ohm superconducting line is soldered to the gold 50 ohm impedance line leading from the HEMT's drain. The oscillator's output is soldered at the opposite end of the superconducting line. All superconductor soldering is done with an indium based solder manufactured by Supersolder Technologies Inc. The oscillator, when cooled below liquid nitrogen temperatures, locks to the superconducting ring resonator and oscillates at 9.33GHz.

This oscillator was designed on a linear microwave simulator using measured small signal s-parameters. [6], [7] A reaction type oscillator has its stabilizing resonator located at its output, in this case the drain. The simulator was used to design the required networks to induce a free running oscillation, (see figure 6).

SCHEMATIC OF REACTION TYPE OSCILLATOR

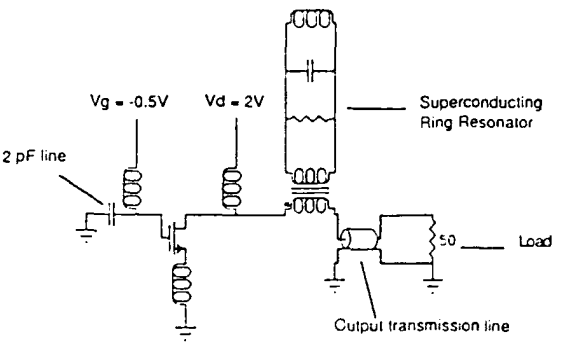


Figure 6.

[8] This small signal analysis is a rough approximation, since oscillators actually operate in a large signal, nonlinear mode. Also room temperature s- parameter data was used instead of cryogenic temperature data,

contributing to the inaccuracies in the predicted operating frequency. Provided the ring resonator has a high quality factor, predicting the exact free running oscillation frequency is unnecessary, the ring will lock the oscillator to its resonant frequency.

The ring resonator, designed to resonate at 9.3GHz, is separated from the 50 ohm line by five mils. In this configuation the ring resonator can be modelled as a parallel RLC circuit connected to the 50 ohm line through a transformer. At resonance, the real impedance at the ring's plane is very large and the rate of reactance change versus frequency near resonance is also very large. The ring resonator was placed 3/4 λ away from the drain of the transistor, so the high impedance is transformed to a low impedance at the transistor, while retaining a large reactance change versus frequency. Optimal placement for the ring is 1/4λ away from the drain, but the ring size made this placement physically unattainable. This transformation insures that the resistance seen by the drain of the transistor is smaller than the negative resistance seen looking into the drain of the transistor, a condition for oscillation. The simulation predicts a negative resistance approximately equal to -80 ohms. Assuming the superconducting ring has a loaded Q = 600, the drain sees approximately 20 ohms, therefore this condition for oscillation is met. The main purpose of the ring is to keep the oscillator locked at a specific frequency as the HEMT's characteristics change, due to a shift in temperature, change in operating voltage, transistor aging, etc. Since the ring's reactance change versus frequency is large compared to the HEMT's reactance change, the oscillator locks to the ring's resonant frequency.

The oscillator was tested using both a spectrum analyzer and a network analyzer. The oscillator was placed in a cryostat and its output connected to the spectrum analyzer. The cryostat was cooled to 39K, well below the superconductor's transition temperature. The HEMT was biased at a drain voltage of 2.0V (Vds) and the gate was biased at -.5V (Vgs), (Id=10mA). The spectrum analyzer showed an oscillation at 9.33GHz, (see figure 7). From this measurement alone, it is indeterminate whether or not the oscillator is locked to the superconducting ring. To test for lock, with the oscillator still in the cryostat, its output was connected to a network analyzer. [9]

OSCILLATION FROM REACTION TYPE OSCILLATOR

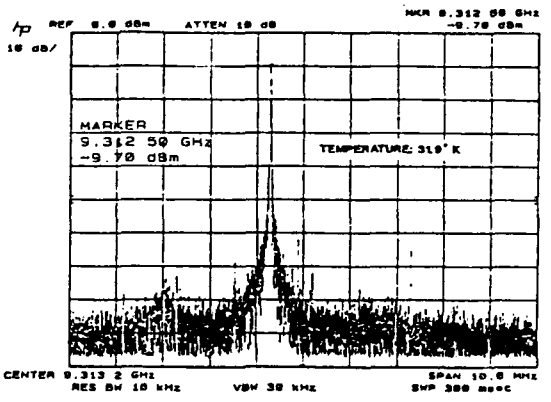


Figure 7.

A one port measurement with the oscillator turned off, (DC voltages set to zero) showed a large return loss at 9.33GHz. This large return loss, "suck out", at the ring's resonant frequency, indicates a high probability that the oscillator is locked to the superconducting ring. The "suck out" disappears when the resonator 's temperature is brought above its transition temperature. An additional test, turning the oscillator on and continuing to measure return loss, should show a reflection coefficient much greater than unity at the ring's resonant frequency. Setting Vgs and Vds to -0.5V and 2.0V, respectively produced a large reflection coefficient much greater than unity at 9.33GHz and therefore provides additional verification that the oscillator is locked to the superconducting ring.

PARALLEL FEEDBACK OSCILLATOR

The parallel feedback oscillator consists of three distinct parts, a superconducting resonator, a high gain amplifier, and a 3dB splitter. This oscillator works by feeding part of the output signal back to the input of the amplifier, through the superconducting resonator. The resonator acts as a very narrow band pass filter whose pass band decreases as the Q - quality factor of the ring increases. If the feedback signal is in phase with the input signal, a phase shift of 2π around the feedback loop, the amplifier will oscillate. [10]

The parallel feedback oscillator was built with the double sided thalium HTSC half wavelength resonator. For this test, the resonator was cooled to cryogenic temperatures, while the amplifier and the 3dB power splitter were kept at room temperature external to the cryogenic chamber. For oscillation to occur the loop gain has to be greater than one. An amplifier with gain greater than 22dB is required, since insertion loss around the loop totals approximately 22dB, 17dBof loss from the ring, 3dB splitter from the splitter, and approximately 2dB from the cables and connectors . An amplifier with 26dB of gain was used. Proper phase shift to provide for positive feedback around the loop, was provided by an appropriate length of semirigid coaxial cable. [10] A schematic of the oscillator configuration is shown in figure 8.

PARALLEL FEEDBACK OSCILLATOR

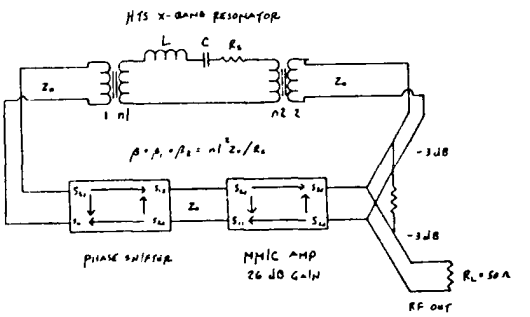


Figure 8.

Cooling the resonator to 46.7 K and turning on the amplifier produced an oscillation at 9.547GHz, which coincides with the resonant frequency of the ring at 46K, (see figure 9). The signal is extremely stable mainly due to the high quality factor of the double sided HTSC resonator.

PARALLEL FEEDBACK OSCILLATION

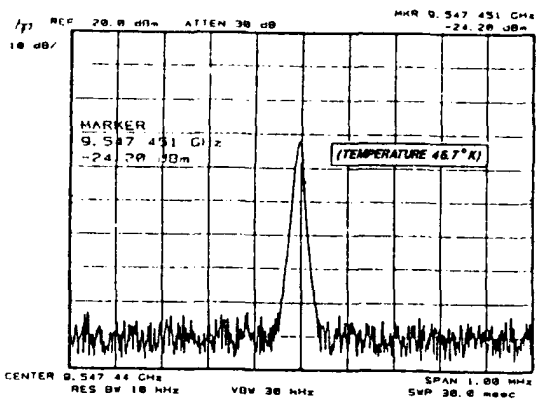


Figure 9.

A theoretical analysis of oscillator phase noise characteristics was made so that general trends and the effects of HTSC on oscillators could be understood. The relative numbers should only be viewed since phase noise is dependent on a large number of variables. For this analysis, the standard Leeson model was used. [11] The single sideband phase noise of a feedback oscillator is:

$$4)\ L(\omega)=10\log_{10}[N^2(1+\omega_o^2/4Q^2\omega^2)(GFkT/P + \alpha/\omega)]$$

$L(\omega)$ =noise power relative to carrier

N =frequency multiplication factor

ω_o =unmultiplied oscillator frequency

ω =offset frequency in radian/s

P =power output of amplifier

G =loop gain of oscillator

k =Boltzmann's constant

T =absolute temperature

α =emperically determined flicker-noise constant

F =amplifier noise figure

Q =unloaded quality factor of superconducting resonator

Present day superconducting resonators have Q's comparable to those of dielectric resonators. Commercially available dielectric resonators at X - band have Q's in the range of about 10,000 to 20,000. [12] A theoretical comparison, using the Leeson model, between two feedback oscillators, shows the oscillator which uses a resonator with a loaded Q = 19,000 has single sideband phasenoise approximately 5db lower than the oscillator which uses a resonator with a Q = 10,000, given all other factors are the same, (see figure 10). Also shown is superconducting oscillator's phase noise as a function of temperature (see figure 11). The phase noise decreases as temperature decreases, since the Q of the resonator increase. This shows the great influence a resonator's quality factor can have on oscillator phase noise.

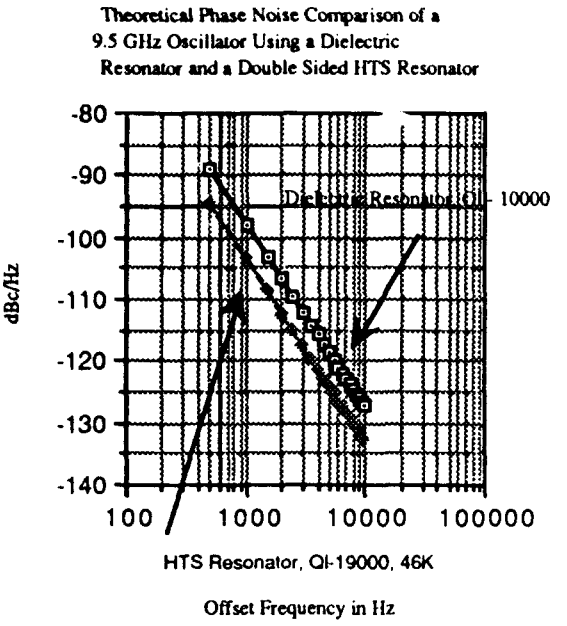


Figure 10.

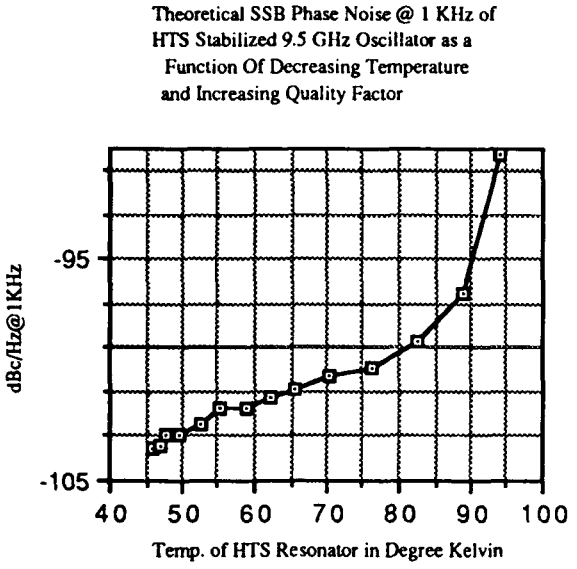


Figure 11.

CONCLUSION

Implementation of HTSC oscillators is heavily dependent on the targeted systems. A distinct advantage of superconducting resonators over dielectric resonators is that superconducting resonators are planar and can be etched into the exact position each time to control coupling. A dielectric resonator oscillator requires the labor intensive task of positioning and securing the dielectric resonator for each oscillator built. For some systems the HTSC's need for cooling can overcome the HTSC advantages. Other systems such as certain ground systems and infrared systems which contain cryogenic cooling can take advantage of the low phase noise and other advantages that HTSC stabilized oscillators offer.

REFERENCES

1. A. Inam, M.S. Hegde, X.D. Wu,. T. Venkatesan, P. England, P.F. Miceli, E.W. Chase, C.C. Chang, J.M. Tarascon and J.B. Watchman, Appl. Phys. Lett. 53, 908 (1988).

2. S. Kaur, A. Fathy, J. Matey and R. Brown, presented at the International Conference on Electronic Materials, Newark, NJ September, 1990, to be published in the conference proceedings.

3. C.G. Montgomery, Principles of Microwave Circuits, Radiation Laboratory Series, McGraw-Hill, New York 1948, pp 237-239.

4. A. Rachlin, Richard Babbit, Erik Lenzing, Roland Cadotte, "Superconducting Devices and System Insertion', Proceedings of SPIE Conference, April 1991.

5. P. A. Polakos, C.E. Rice, M.V. Schneider, and R. Trambarulo, "Electrical Characteristics of Thin-Film BaYCuO Superconducting Ring Resonators," IEEE Microwave and Guided Wave Letters, Vol. 1, No. 3, March 1991.

6. David A. Warren, Michael J. Golio, and Warren l. Seely, "Large and Small Signal Oscillator Analysis," Microwave Journal, pp. 229-244, May 1989.

7. A.P.S. Khanna, and E. Topacio, "A Highly Stable 36 GHz GaAs FET DRO with Phase-Lock Capabilities," Microwaves Journal, pp. 117-122 July 1989.

8. Patrick Champagne, "CAD Software Optimized DRO Embedding Network," Microwaves and RF, pp 79-85 October 1986.

9. Robert Soares, GaAs MESFET Circuit Design, Artech House, Boston 1899 pp 347-411.

10. A.P.S Khanna, Mike Schmidt, "Low Phase Noise Superconducting Oscillators," 1991 IEEE MTT-S Digest, 1239-1242.

11. Steven T. Ruggiero, David A. Rudman, Superconducting Devices, Academic Press, New York, 1990 pp. 253-255.

12. E.C. Niehenke andP.A. Green, "A Low Noise L-Band Dielectric Resonator Stabilized Microstrip Oscillator," MTT-S International Microwave Symposium, 1987.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

COMBINED STABILITY OF HYDROGEN MASER AND SUPERCONDUCTING CAVITY MASER OSCILLATOR*

R. T. Wang and G. J. Dick
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

We have combined a Superconducting Cavity Maser Oscillator (SCMO) and a Hydrogen Maser to form a standard which shows the medium-term performance of the hydrogen maser together with improved short-term performance only available from cryogenic oscillators. A double loop receiver combines the two sources with a 100 MHz quartz crystal oscillator. Fractional frequency stability of $2 \cdot 10^{-14}$ was obtained for a measuring time of $\tau = 1$ second, and $1 \cdot 10^{-15}$ at $\tau = 1000$ seconds. The 1 second value is almost as low as the floor ($1 \cdot 10^{-14}$) for our measurement system while the 1000 second value is identical to hydrogen maser performance. The design is based on models of hydrogen maser and SCMO phase noise for offset frequencies down to $1 \cdot 10^{-6}$ Hz.

Introduction

The superconducting cavity maser oscillator (SCMO) is a helium cooled, all-cryogenic oscillator with superior stability at short measuring times[1,2,3,4]. It differs from other superconducting cavity stabilized oscillator (SCSO) designs[5,6,7] in its use of a very rigid ($Q \approx 10^9$) sapphire-filled stabilizing cavity, and in its all-cryogenic design; excitation being provided by an ultra-low noise cryogenic ruby maser.

The three cavity oscillator, consisting of a ruby maser, coupling cavity, and a high-Q lead-on-sapphire cavity, have been discussed previously[2]. Oscillation at a frequency of 2.69 GHz results from ruby maser operation with a 13.1 GHz pump frequency to create a population inversion. In order to combine the short term stability of the SCMO with the long term stability of the hydrogen maser we have modified the SCMO to allow its frequency to be tuned. A coil has been installed on the ruby housing to allow the bias field to be slightly modified. This coil gives a sensitivity of $7 \cdot 10^{-12}$ per mA with a range of approximately 10^{-10}. This range is sufficient to accommodate the typical SCMO drift of $4 \cdot 10^{-13}$/day in long term operation.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

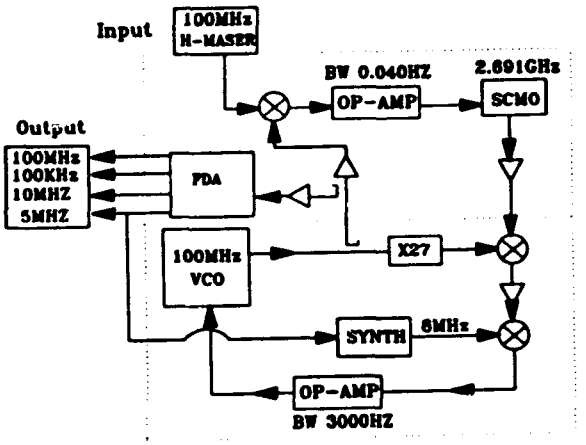


Figure 1: Schematic of SCMO double loop receiver at 100MHz. The two loops are VCO/SCMO and SCMO/Hydrogen Maser. The required input is 100MHz Hydrogen Maser signal and the output has four different frequencies: 100, 10, 5, and 0.1MHz.

Design and Model

A double phase lock loop (PLL) was designed to optimally combine the SCMO and hydrogen maser stabilities. Figure 1 shows the schematic of this de-

CH2965-2/91/0000-491 $1.00 © 1991 IEEE

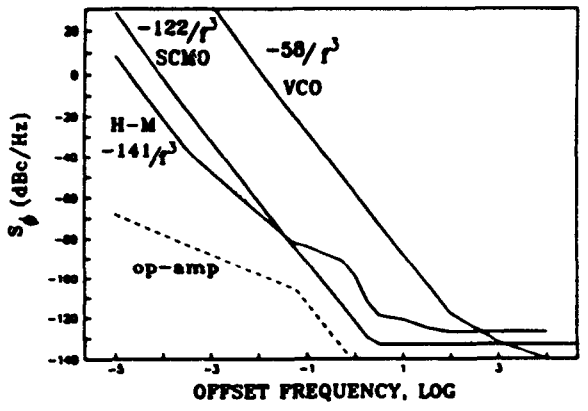


Figure 2: Phase noise plot of the three oscillators at 100MHz. The combined signal is expected to follow closely the lowest noise of the three oscillator in the Fourier frequency window of $1 \cdot 10^{-5}$ to 10^4 Hz. Op-amp noise also shown to be much lower than oscillator noise.

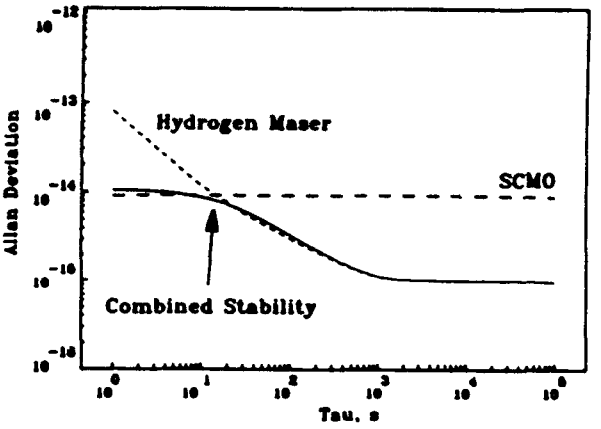


Figure 3: Allan Deviation derived from phase noise data (Fig. 2). The expected performance of the combined signal is $1 \cdot 10^{-14}$ at one second and $1 \cdot 10^{-15}$ at 1000 seconds and beyond.

sign. The first loop is a quartz crystal voltage control oscillator (VCO) phase locked to the SCMO and the second loop locks the SCMO-VCO combination to a hydrogen maser through a tuning coil in the SCMO. The design goals are to preserve SCMO short term stability through first PLL and to optimize the second PLL so that long term performance of the hydrogen maser is preserved without significantly degrading SCMO performance at $\tau = 1$ second measuring time. The VCO/SCMO loop is a standard PLL with bandwidth of 3000 Hz. The SCMO/hydrogen maser loop deserves a little more attention.

The bandwidth of the SCMO/hydrogen maser PLL is determined by the phase noise crossover of SCMO and hydrogen maser. Since the hydrogen maser close-in noise can't be easily measured a model was created with assumption that conversion between frequency and time domain falls within reasonable margins. Necessary information contains in Table 1. It includes noise type, Fourier frequency window, Allan deviation and calculated phase noise.

Figure 2 shows phase noise plot of the three oscillators and op-amp noise contribution is also calculated. In a preliminary test the difference between phase noise model and the measured data is within 3 dB for $1 \cdot 10^{-4}$ to $1 \cdot 10^{-1}$ Hz region[8]. A combined phase noise data were calculated. Through a sophisticated integration the combined phase noise was converted to Allan deviation. It shows agreement between the measured Allan deviation and the calculated one. Based on this model the phase noise crossover between SCMO and hydrogen maser is determined to be 0.040 Hz.

In order to check the calculation method a precautionary step has been taken. Not only was phase noise calculated for the SCMO, hydrogen maser and the combined source but also converted back again to Allan deviation. Figure 3 shows the result of our calculations. The calculated Allan deviations reflect the measured data on SCMO and hydrogen maser. Therefore it confirms our method of calculation. The calculated combined source shows a stability of $1 \cdot 10^{-14}$ at one second and $1 \cdot 10^{-15}$ from 1000 seconds onward. Allan deviation of the combined source follows very closely to the best source in the time window of 1 to 10,000 seconds.

Measurements

To confirm the stability of the combined source, we will need a second identical system. At present we can only test with available oscillators. Figure 4 shows the best short term stability of the combined signal at test frequency of 100MHz. Using another cryogenic oscillator (superconducting cavity stabilized oscillator[5]) as reference, these results show a stability of $2 \cdot 10^{-14}$ at one second. Figure 5 shows the Allan Deviation of fractional frequency fluctuations for the SCMO locked to a hydrogen maser,

Noise type	Fourier Frequency	Allan Deviation	Phase Noise
Flicker frequency	$1 \cdot 10^{-6}$ - $4 \cdot 10^{-4}$	$1 \cdot 10^{-15}$	$7.5 \cdot 10^{-15}/f^3$
White frequency	$4 \cdot 10^{-4}$ - $4.5 \cdot 10^{-2}$	$3 \cdot 10^{-14}/\tau^{1/2}$	$1.8 \cdot 10^{-11}/f^2$
Flicker phase	$4.5 \cdot 10^{-2}$ - $1 \cdot 10^{-1}$	$2 \cdot 10^{-13}/\tau$	$4 \cdot 10^{-10}$/f

Table 1: Contributions to the total noise of the DSN hydrogen maser used for loop design

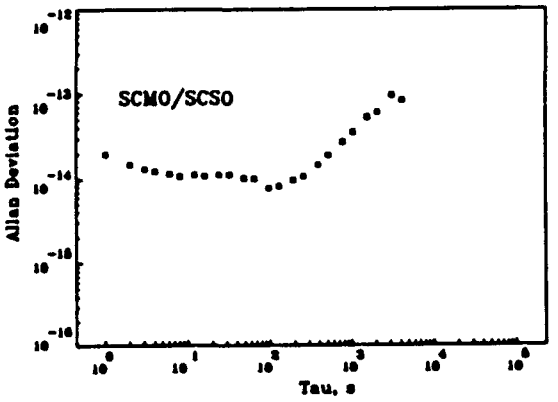


Figure 4: Two sample Allan Deviation of the SCMO tested with SCSO at 100 MHz. The measured frequency stability is $2 \cdot 10^{-14}$ at one second.

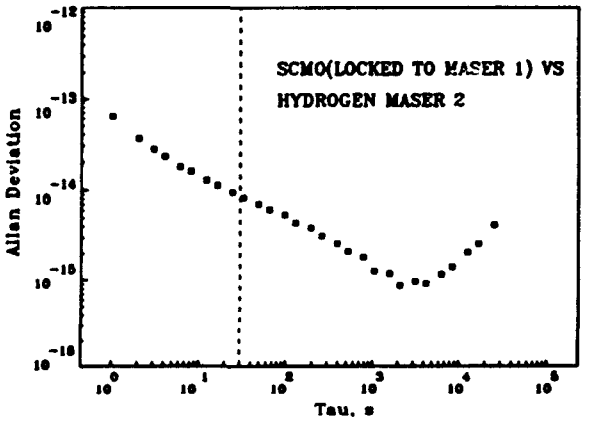


Figure 5: Two sample Allan Deviation of the SCMO locked to one Hydrogen maser and second maser is used as a reference. The dotted line shows the 30 second loop time constant. Below 30 seconds plot shows one Hydrogen maser stability and beyond 30 seconds it shows two maser's stability.

using a second hydrogen maser as reference. A special feature is the stability of $6 \cdot 10^{-14}$ at one second, a value which is one maser's stability. Another view of this data is from 1 to 30 seconds the stability is from one maser and beyond 30 seconds it represents two maser performance.

Conclusions

We have demonstrated a new frequency standard. It combines two stable frequency sources which show complementary stabilities over the range of measuring times from 1 to 10,000 seconds. The hydrogen maser gives the highest possible stability for any frequency source for measuring times from 30 to 10,000 seconds while the cryogenic SCMO supercedes even this performance for times shorter than 30 seconds. Measured performance for the combined standard is $2 \cdot 10^{-14}$ at 1 second measuring time and $1 \cdot 10^{-15}$ at 1000 seconds. Significant aspects of this experiment are a new time window for scientific experiments and the first successful demonstration combining two different types of ultra-stable microwave oscillators.

At present, the combined SCMO-hydrogen maser source is performing a field test at Goldstone, NASA/Deep Space Tracking station. This exercise is to prepare SCMO for Gravitational Wave Detection experiment in connection with Galileo in 1992.

Acknowledgements

Special thanks to W. A. Diener for his assistance in ultrastable measurements, R. E. Taylor for easy access to the capabilities of the Test Facility and assistance with the SCMO field tests, Dr. C. A. Greenhall for software enhancements meeting our special requirements, and T. K Tucker, for assistance and support with hydrogen maser measurements. The authors would like to thank Dr. L. Maleki for many helpful suggestions.

References

[1] S. Thakoor, D. M. Strayer, G. J. Dick and J. E. Mercereau, "A Lead-on-Sapphire Superconducting Cavity of Superior Quality," *J. Appl. Phys.*, **59**, 854-858 (1986).

[2] G. J. Dick and D. M. Strayer, "Development of the Superconducting Cavity Maser as a Stable Frequency Source," in *Proceedings of the 38th Annual Frequency Control Symposium*, 435-446 (1984).

[3] R. T. Wang, G. J. Dick and D. M. Strayer, "Operational Parameters for the Superconducting Cavity Maser," in *Proceedings of the 20th Annual Precise Time and Time interval (PTTI) Planning and Applications Meeting*, 345-354 (1988).

[4] R. T. Wang and G. J. Dick, "Improved Performance of the Superconducting Cavity Maser at Short Measuring Time," in *Proceedings of the 44th Annual Frequency Control Symposium*, 89-93 (1990).

[5] S. R. Stein and J. P. Turneaure,"The Development of The Superconducting Cavity Stabilized Oscillator," in *Proceedings of the 27th Annual Frequency Control Symposium*, 414-420 (1973).

[6] S. R. Stein, "Space Application of Superconductivity: Resonators for High Stability Oscillator and Other Applications," *Cryogenics*, **20**, 363-371 (1980).

[7] A. J. Giles, S. K. Jones, D. G. Blair, and M. J. Buckingham, "A High Stability Microwave Oscillator based on a Sapphire Loaded Superconducting Cavity," *Proceedings of the 43rd Annual Frequency Control Symposium*, 89-93 (1989).

[8] Private communication Dr. C. A. Greenhall.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ANALYSIS OF A LOW NOISE TUNABLE OSCILLATOR BASED ON A TUNABLE SAPPHIRE LOADED SUPERCONDUCTING CAVITY

Michael E. Tobar and David G. Blair

Department of Physics, University of Western Australia, Nedlands 6009, Australia.

Abstract: A Tunable Sapphire Loaded Superconducting Cavity (T-SLOSC) has been constructed and is used as a high Q filter in a X-Band loop oscillator. It is configured with two cryogenically operational GaAsFET amplifiers. The oscillator is situated in a vacuum can with a two mode vibration isolation system designed to filter vibrations by 50 dB at 700 Hz from the carrier. By matching theory with measurement we identify the field patterns of some mode families at room temperature. Calculating frequency shift with temperature enables identification of modes when cooled. Tuning ranges can also be estimated. To measure phase noise, we make the T-SLOSC a VCO, and lock it to a fixed frequency SLOSC. The phase noise is measured at room temperature and a range of cryogenic temperatures. The major limitation is the excessive flicker noise exhibited by GaAsFET amplifiers in the feedback loop.

Introduction

The ultra low loss tangent and substantial dielectric constant of sapphire make it an ideal material for the construction of ultra low phase noise microwave resonators. This has been demonstrated in the development of the sapphire clock at the University of Western Australia, which has state of the art Allan Variance of $< 10^{-14}$ for integration times from 1 to 30 seconds [1] [2].

The loss tangent decreases rapidly as the temperature is reduced, giving quality factors rising from $3 \cdot 10^5$ at 300K, to $5 \cdot 10^7$ at 77K, and exceeding 10^9 at 4.2K. The low losses of sapphire may be exploited in several different ways. Large high mode number resonators may be used in which radiative losses are negligible. Alternatively, fundamental mode resonators may be shielded using a superconductor. Neither of these configurations is convenient. The first is large and bulky, while the second restricts operation to temperatures for which the shield is superconducting, which also contributes a significant loss. The third, and in our opinion optimum configuration, trades of wall interactions with physical size. The sapphire is shaped as a spindle, and mounted with a spacing of several evanescent field scale lengths from the walls. Typically these resonators are used with azimuthal mode numbers between 5 and 10.

This work extends the sapphire resonator technology to a tunable frequency source. The sapphire spindle is replaced by a mushroom, and a small tuning disk is adjusted to tune the resonant frequency. Due to the high mode number and hence large number of modes present, a reasonable fraction of the X-band spectrum may be assessed by such a resonator.

Construction

Construction of the T-SLOSC under investigation has been illustrated previously [3]. The heart of this resonator is a cylindrical sapphire mushroom and a tuning disk inside a cylindrical niobium cavity. The sapphire disk affects the evanescent field outside the sapphire, and perturbs its resonant frequency. Tuning is achieved by a stepper motor that is capable of cryogenic operation. Three niobium loop probes are used to couple to the cavity, stepper motors attached to the probes are used to vary this coupling.

The feedback loop for the T-SLOSC must be designed so that mode selection is possible over the required range of tuning. Two cryogenic amplifiers and a cryogenic varactor phase shifter are situated in the vacuum can with the T-SLOSC. The feedback loop comprises of these cryogenic components, together with a tunable room temperature filter for mode selectivity. Mode selection was shown to be possible with just phase shift for a compact loop at room temperature. Thus an all cryogenic loop may be possible, however the versatility required for the initial experiments made this not possible.

A two mode vibration isolation system was designed to eliminate excessive low frequency vibrations, and still give a significant attenuation at 700 Hz. It comprises of an intermediate mass attached via leaf springs to the top plate of the vacuum can, and the T-SLOSC in the same way attached to the intermediate mass.

Mode Analysis

Previous work has enabled us to identify the mode families in a cylindrical piece of sapphire [4], at 4.2 K two modes at 10.22 GHz and 10.44 GHz were identified to be quasi $TE_{6\,1\,1+\delta}$ and quasi $TM_{8\,1\,1+\delta}$ respectively. Exploring the performance of the T-SLOSC oscillator constructed, we have concentrated on the 10.22 GHz mode which has an unloaded Q of about 10^9 [3].

Tuning ranges may be predicted by calculating the resonant frequency of the cylindrical resonator with height h and h+t. For the 10.22 GHz mode, 94 MHz is predicted which is consistent with measurement. From table 1 we can see that modes with greater axial mode number p have larger tuning ranges.

Frequency shifts with temperature are predicted by the change in dielectric constant, which is the dominant effect. Calculations agree favorably with experiment. TM modes have a larger frequency shift with temperature than TE modes, because the permittivity parallel to the c-axis changes more than the permittivity perpendicular to the c-axis (see table 1).

CH2965-2/91/0000-495 $1.00 © 1991 IEEE

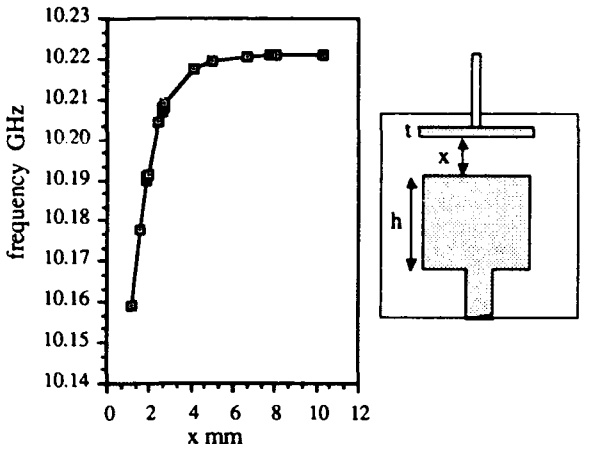


Fig. 1. Tuning curve for the quasi $TE_{6\,1\,1+\delta}$ mode, with a schematic of the T-SLOSC.

TE Modes m,n,p+ δ	Predicted Frequency @ 290 K (GHz)	Experimental Frequency @ 290 K (GHz)	Predicted Frequency Shift @ 4K (MHz)	Predicted Tuning Range (MHz)
8, 1, 0	11.921	11.911	95.9	20.6
7, 1, 0	10.828	10.809	74.3	22.4
6, 1, 0	9.712	9.688	66.5	25.1
5, 1, 0	8.586	8.553	58.9	28.5
7, 1, 1	11.223	11.220	79.0	85.6
6, 1, 1	10.154	10.151	71.8	94.3
5, 1, 1	9.087	9.090	64.5	104.8
TM Modes m,n,p+ δ				
10, 1, 0	11.890	11.888	131.9	24.8
9, 1, 0	10.909	10.882	114.2	21.1
8, 1, 0	9.917	9.888	107.7	28.0
7, 1, 0	8.920	8.887	96.9	30.0
9, 1, 1	11.289	11.320	121.8	90.6
8, 1, 1	10.336	10.368	110.3	98.6
7, 1, 1	9.384	9.415	99.9	108.2
6, 1, 1	8.436	8.463	88.9	120.7

Table 1. Theoretical and experimental comparisons of the four lowest order in p, mode families.

Vibration Isolation

Vibrations cause movement of the tuning plunger, probes and walls with respect to the sapphire resonator. These vibrations will modulate the loop oscillator signal and manifest as phase noise. A one dimensional model of the isolation system designed is shown in fig. 2.

The transfer function X_2 / X_0, of this system has been modelled and is shown in fig. 3. We measure the intermediate mass and T-SLOSC mass to be; $m_1 = 2.9$ kg and $m_2 = 5.95$ kg respectively. Also we calculate from the leaf spring properties that; $f_1 = 117$ Hz and $f_2 = 167$ Hz. The transfer function in fig. 3 is calculated with $Q_{1,2} = 5$.

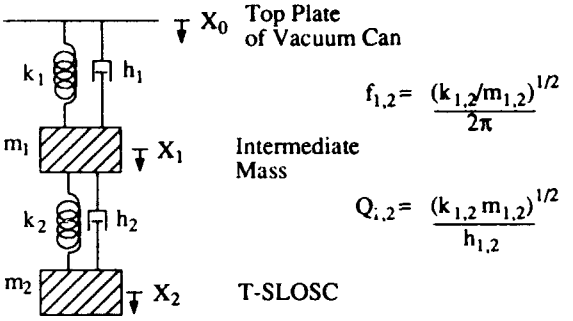


Fig 2. One dimensional model of isolation system.

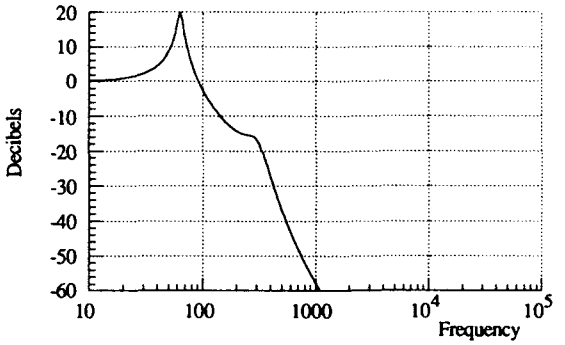


Fig. 3. Predicted transfer function of vibration isolation, from top plate to T-SLOSC.

Large low frequency vibrations have previously limited the useful tuning range. The current isolation is stiff enough to limit low frequency vibrations, while still achieving 50 dB attenuation at 700 Hz.

Phase Noise Measurement

A tunable reference oscillator is constructed by mixing a fixed frequency SLOSC with a HP 8662A synthesizer. The noise floor of the synthesizer was measured to be between -140 dBc and -135 dBc. Forth coming results reveal that amplifier flicker noise is present above this level, hence the reference oscillator below suffices for these measurements.To reduce the effects of flicker noise in the reference oscillator we have configured a cooled SLOSC with just one high gain narrow band amplifier in the feed back loop, also the signal is taken out of the filtered port.

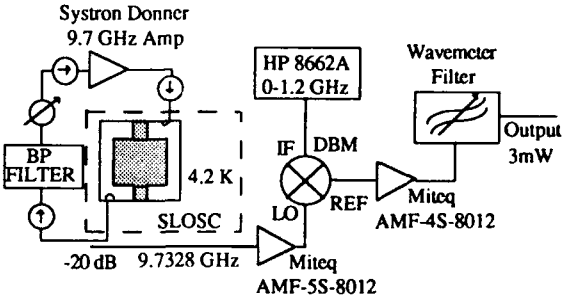


Fig. 4. Reference Oscillator Circuit

The circuit shown in fig. 5 shows the configuration used to construct a compact loop oscillator, where just the phase condition is successful in distinguishing modes of oscillation. This is useful as the feedback loop may be kept cold, and modes may be selected by just an external voltage control. However to keep the circuit versatile at cryogenic temperatures we must be able to move the probes to adjust couplings. This requires the use of thin flexible pieces of coaxial cable. Once they are added just phase shift alone is no longer good enough to distinguish between electromagnetic modes. Hence it was necessary to include a tunable wavemeter filter at room temperature in the feed back loop. The extra attenuation supplied by the room temperature part of the loop requires that two amplifiers must be used in the loop. Measurements at 77 K and 40 K made use of two cryogenic amplifiers, however at 4.2 K one of the amplifiers failed and it was necessary to include a room temperature amplifier in the feedback loop.

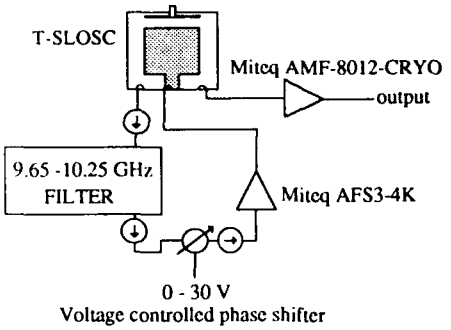


Fig. 5. Schematic of a compact room temperature circuit, where all modes between 9.65 and 10.25 GHz could be selected by just controlling the voltage controlled phase shifter.

For the phase noise measurements described below, we determine that the flicker noise in the amplifiers is the dominant effect. When flicker noise is dominant, the flicker noise level at 1 Hz due to the amplifiers, is given by Eq. 1 [5].

$$S_\phi(\text{Amps. @ 1 Hz}) = \frac{S_\phi(\text{Osc. @ 1 Hz})}{(\Delta f/2)^2} \qquad (1)$$

where Δf is the bandwidth of the resonator.

The phase noise measurements therefore allow us to determine the level of flicker noise in the feedback loop using Eq. 1.

Phase Noise Results

Room Temperature

The set up for room temperature measurements is shown in fig. 6. Phase noise results are given in fig. 7.

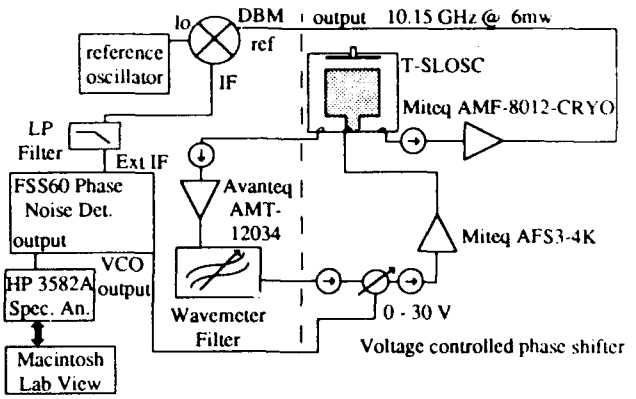


Fig 6. Configuration of the room temperature phase noise measurement setup. To the left of the dashed line is component outside the experimental vacuum can, and to the right are components inside the experimental vacuum can.

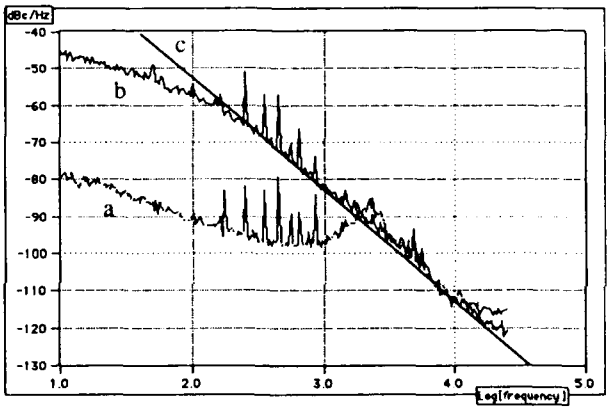


Fig 7. SSB phase noise of the T-SLOSC oscillator set up in fig. 6. a ; Phase noise with a high loop gain in the PLL. b; The effect of decreasing the PLL loop gain. c ; The extrapolated phase noise disregarding the effects of the PLL.

This measurement was achieved with the T-SLOSC oscillating on the $TE_{6\,1\,1+\delta}$ mode at 10.15 GHz, which had a loaded Q of $1.3 \cdot 10^5$. From Eq. 1, and fig. 7, we determine that; S_ϕ(Amps. @ 1 Hz) = - 84 dBc.

Liquid Nitrogen Temperature

The set up for liquid nitrogen temperature measurements is shown in fig. 8. Phase noise results are given in fig. 9.

At 77K the $TE_{6\,1\,1+\delta}$ mode is split into a doublet pair, at 10.21966 and 10.21957 GHz, with loaded Q values of $2.2 \cdot 10^6$ and $7.6 \cdot 10^5$ respectively. The phase noise measurement in fig. 9 was taken while oscillating on the lower Q mode. From Eq. 1, and fig. 7, we determine that; S_ϕ(Amps. @ 1 Hz) = - 96 dBc.

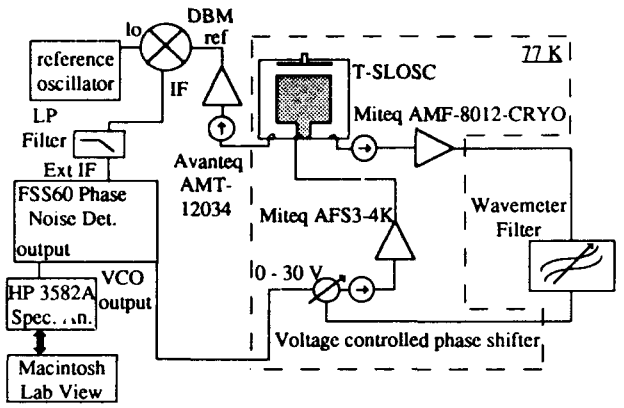


Fig 8. Configuration of the room temperature phase noise measurement setup.

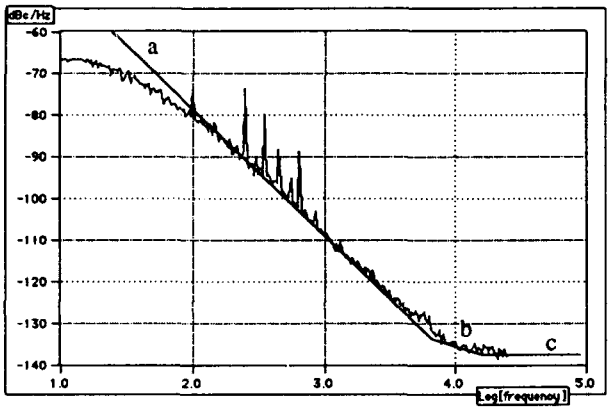


Fig 9. SSB phase noise of the T-SLOSC oscillator set up in fig. 8. a ; Flicker noise within the band width of the resonator, with 30 dB per decade roll of b; Flicker noise outside the band width of the resonator, with 10 dB per decade roll of c ; Noise floor of the HP 8662A synthesizer.

40 K

As the experiment shown in fig. 8 warmed up, at about 40K with no liquid helium left in the dewar, the phase noise was measured (fig. 10). The T-SLOSC was oscillating on the lower Q mode ($Q = 5 \cdot 10^6$), at 10.2209 GHz.

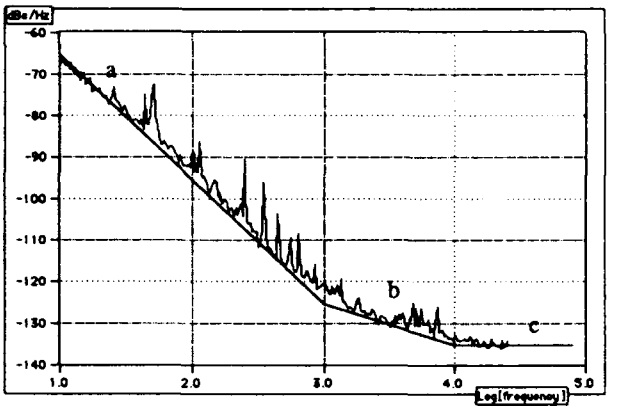


Fig 10. SSB phase noise of the T-SLOSC oscillator at 40 K. a ; Flicker noise within the band width of the resonator, with 30 dB per decade roll of b; Flicker noise outside the band width of the resonator, with 10 dB per decade roll of c ; Noise floor of the HP 8662A synthesizer.

From Eq. 1, and fig. 7, we determine that; S_ϕ(Amps. @ 1 Hz) = - 96 dBc.

Liquid Helium Temperature

At liquid helium temperature the Miteq AMF-8012-CRYO amplifier failed. This port was thus decoupled from the oscillator. A room temperature amplifier was combined with the tunable wavemeter filter to complete the feedback loop with the two remaining ports. A directional coupler was placed before the amplifiers so the filtered signal was measured.

Oscillation was achieved on the higher Q mode of the doublet. The loaded Q was measured to be $1.5 \cdot 10^8$ and frequency 10.22107 GHz.

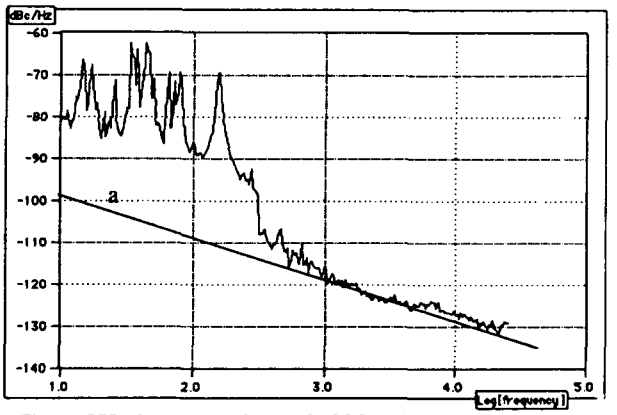


Fig 11. SSB phase noise of the T-SLOSC oscillator at liquid helium temperature. a ; Flicker noise outside the band width of the resonator, with 10 dB per decade roll of.

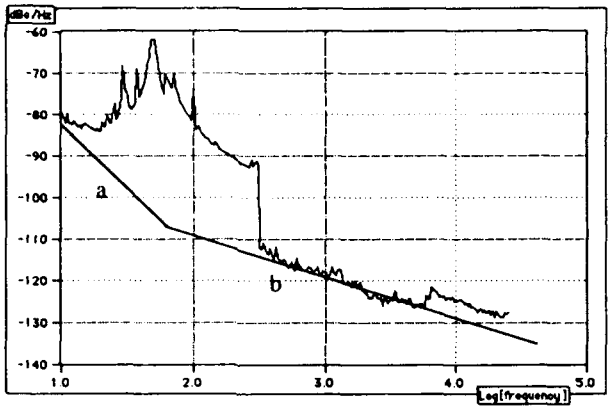


Fig 12. SSB phase noise of the T-SLOSC oscillator at liquid helium temperature when raised on springs. a ; Flicker noise within the band width of the resonator, with 30 dB per decade roll of b; Flicker noise outside the band width of the resonator, with 10 dB per decade roll of.

When the dewar is raised on soft springs the low frequency spectrum disappeared. Due to a leak in the experimental can a backing pump was continually running, the increase in tension in the rubber connection seemed to enhance frequencies between 20 and 300 Hz. When the pump was turned off this spectrum was seen to disappear, however the frequency drift due to the leak was enough to inhibit a phase noise measurement with the pump off. It is interesting to compare the structure of the low frequency noise in fig. 12 with the theoretical transfer function (fig. 3) of the vibration isolation designed. This is strong evidence that the vibration isolation works as designed.

From Eq. 1, and fig. 11, we determine that; S_ϕ(Amps. @ 1 Hz) = - 91 dBc. This is worse than the 40K and 77K results due to the room temperature amplifier that was required in the feedback loop.

Discussion

The reference oscillator is practically an absolute oscillator for the room temperature, 77K and 40K results. However the noise floor of the HP 8662A synthesizer at 77K and 40 K can be seen at frequencies greater than 10 kHz. At these frequencies the flicker noise at 4.2K is higher than this noise floor. This means that we are seeing the phase noise of the T-SLOSC. This extra phase noise is due to the inclusion of a *room temperature amplifier.*

The noise floor of the SLOSC oscillator part of the reference oscillator, was never seen. This is because it only uses one amplifier. The T-SLOSC uses two, which means the second amplifier will operate highly saturated, and will be susceptible to non linear upconversions of low frequency noise.

With one amplifier in the feedback loop we should be able to limit the flicker noise effect on the oscillator. We estimate that -120 dBc/Hz to -130 dBc/Hz flicker noise may be achievable [6], either by using a single cryogenic GaAsFET amplifier or a bipolar transistor. This would give us a SSB phase noise of -150 dBc to -160 dBc at 1 kHz.

Conclusion

We have successfully built and understood a cryogenic tunable X-band source with a SSB phase noise of -120 dBc at 1 kHz. By reconfiguring this oscillator we expect to be able to construct a oscillator with a SSB phase noise between -150 to -160 dBc at 1 kHz.

Acknowledgements

The authors would like to thank Adrian Giles, Anthony Mann and Andre Luiten for their general assistance. This research was supported by the Australian Research Council.

References

[1] A.J. Giles, S.K. Jones, D.G. Blair and M.J. Buckingham, "A high stability microwave oscillator based on a sapphire loaded superconducting cavity," in Proc. IEEE 43rd Annu. Symp. Freq. Contr., 1989. pp. 89-93.

[2] A.J. Giles, A.G. Mann, S.K. Jones, D.G. Blair and M.J. Buckingham, "A very high stability sapphire loaded superconducting cavity oscillator," Physica B, 165, pp. 145-146, 1990.

[3] M.E. Tobar and D.G. Blair, "A generalized equivalent circuit applied to a tunable superconducting resonator," IEEE Trans. Microwave Theory Tech. special issue on Microwave Applications of Superconductors, Sept, 1991.

[4] M.E. Tobar and A.G. Mann, "Resonant frequencies of higher order modes in cylindrical anisotropic dielectric resonators," in Proc. IEEE. MTT-S International Microwave Symposium, June. 1991.

[5] D.B. Leeson, "A simple model of feedback oscillator noise spectrum," in Proc. IEEE, Vol. 54, No. 2, Feb. 1966, pp. 329-330.

[6] G.K. Montress, T.E. Parker and M.J. Laboda, "Residual phase noise measurements of VHF, UHF and microwave components" in Proc. IEEE 43rd Annu. Symp. Freq. Contr., 1989. pp 349 - 359.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Laser Frequency Stability Requirements for Coherent Space Communications

Chien-Chung Chen and Moe Zaw Win

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive, Pasadena, California, 91109

Abstract

The required laser frequency stability for a coherent optical receiver is analyzed for both data communication and Doppler tracking applications. The analysis indicated that the short term frequency stability of diode-pumped solid state laser can easily meet the requirement for coherent data communication channel if the data rate exceeds 10 kbps. Coherent phase tracking loop can potentially be implemented with received signal power as low as several fW. The required signal power of the phase locked receiver can be further reduced provided that both transmit and LO lasers can be stabilized to an external reference. The long term frequency stability requirement is derived for a Doppler tracking subsystem. The results indicated that long term laser frequency stability better than one part in 10^{16} will be required to provide a performance comparable to that which is projected at Ka band.

I. Introduction

Coherent heterodyne optical communication technology can provide improved receiver sensitivity compared to direct detection systems [1-2]. By amplifying the weak incident signal with a strong local oscillator (LO) output, the system can effectively get above the thermal noise limit and achieve near quantum-limited detection sensitivity. Since the spectral filtering is performed at the intermediate frequency, where the bandwidth can be much more selective, heterodyne reception offers a better background noise rejection capability. The bandwidth selective nature of the coherent receiver can also lead to a more efficient use of the optical spectrum and the potential of multiple-access communications over a single lasing bandwidth. The performance of the communication channel can be further improved with the use of frequency and phase encoding schemes.

In order to achieve phase coherent reception with the low received signal power projected for a space-to-space link, however, stringent demands must be placed on the frequency stability of both the transmit and local oscillator lasers. Semiconductor lasers, because of their large linewidth enhancement factor, typically have linewidths on the order of several MHz [3]. Coherent phase tracking loops for diode-laser based systems will therefore require several megahertz of bandwidth and demand a minimum signal power on the order of 1 nW. However, the low signal power expected from some free-space links precludes the use of high bandwidth phase tracking loops. Recently, with the advances in diode-pumped solid state laser technology, sub-KHz linewidth frequency stabilized lasers have been developed [4-5]. The high spectral purity of the frequency stabilized solid state laser is very desirable in applications such as free-space optical communications, where synchronization and retrieval of the received signal phase in the presence of weak incident signal power is required.

In addition to data communications, the potential of obtaining high frequency stability at the optical frequency can also improve spacecraft navigation. Current spacecraft navigation relies heavily on the two-way Doppler tracking technique to measure the spacecraft velocity [6]. In a two-way Doppler measurement, the uplink signal is coherently received, multiplied, and retransmitted by the spacecraft. On the ground station, the frequency and phase of the up and down link signals are compared, and the Doppler shift in frequency is calculated. The dominating source of error for the RF Doppler measurement technique is the phase noise introduced by the plasma scintillation as the signal passes through the charge particle region near the Sun. Since the charge particle effect decreases as the square of carrier frequency, the optical signal is inherently less sensitive to the charge particle effect. Consequently, an optical carrier capable of achieving similar frequency stability to current-day RF system can potentially offer significant improvement in spacecraft navigation capabilities.

CH2965-2/91/0000-500 $1.00 © 1991 IEEE

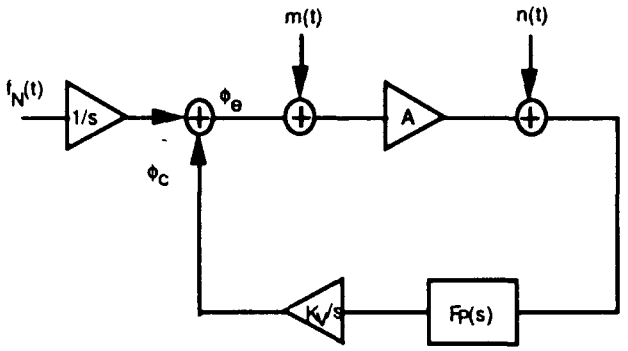


Fig.1 Block diagram of a linearized phase-locked loop showing various noise sources that affect loop performance.

The purpose of this paper is to outline the requirements for laser frequency stability for both data communication and Doppler tracking. The paper will be structured as follows. In Section 2, a brief overview of the coherent receiver will be provided. The application of the phase coherent reception in data communication will be addressed, and the required signal power level needed to achieve phase coherent reception will be calculated based on the laser frequency noise model measured for the Lightwave™ lasers. For data communications, it is the short term stability of the lasers that is of primary concern. The issues related to the short term frequency stability will be addressed in Section 3. The long term frequency stability problem and its requirements in Doppler measurement will be addressed in Section 4.

II. Coherent Carrier Phase Tracking

In general, coherent reception is accomplished by synchronizing the a local reference with the incoming carrier using a receiver phase-locked loop (PLL) [7-8]. The carrier is then mixed with the reference to recover the baseband data stream. Shown in Figure 1 is a simplified block diagram of the phase locked loop. The PLL consists of a phase detector, a loop filter, and a voltage controlled oscillator (VCO). The phase detector compares the phase between the incoming carrier and the local oscillator (LO) output and outputs a signal that is proportional to the phase error. This error signal is then filtered by a loop filter $F(s)$ and fed back into a VCO with tuning constant K_V.

The application of PLL in RF communication systems has been analyzed extensively [7-8]. Several studies have also been conducted for optical PLLs[9-11]. The PLL performance is affected by the noises present in the tracking loop. For an optical PLL, the noise sources include the additive noise in the detection process and electronics, $n(t)$, the oscillator frequency noise, $f_N(t)$, and the noise introduced by the modulation on the carrier, $m(t)$. It is generally assumed that the tracking loop bandwidth is small compared to the data rate such that the modulation-induced noise is negligible. Furthermore, when the loop is operating in the linear region, the effects of each noise can be determined separately and combined to obtain the net results [8]. The total phase error variance is therefore given by

$$\sigma_{\phi_e}^2 = \sigma_{\phi_n}^2 + \sigma_{\phi_f}^2. \quad (1)$$

where $\sigma_{\phi_n}^2$, $\sigma_{\phi_f}^2$ are phase error variances due to additive noise and oscillator frequency noise, respectively.

The additive noise $n(t)$ in the loop is due primarily to the local oscillator shot noise and pre-amplifier thermal noise, and can be modelled as an additive white Gaussian noise process with power spectral density (PSD) N_0. When the LO is sufficiently strong, the LO shot noise is generally the dominating source of additive noise. The phase error variance contributed by the additive noise can be written as [8]

$$\sigma_{\phi_n}^2 = \frac{1}{A^2}\int_0^\infty |H_1(j2\pi f)|^2 N_0\, df = \frac{N_0}{A^2} B_L, \quad (2)$$

where A is the baseband signal amplitude, $H_1(s)$ is the closed loop transfer function, and the parameter B_L is the one sided loop bandwidth defined by

$$B_L \equiv \int_0^\infty |H_1(j2\pi f)|^2\, df\,. \quad (3)$$

The closed loop transfer function $H_1(s)$ is related to the tuning constant of the VCO, K_V and the filter transfer function, $F(s)$ by by [8],

$$H_1(s) = \frac{AK_V F(s)}{s + AK_V F(s)}. \quad (4)$$

For a shot noise limited receiver, the signal to noise ratio, $A^2/2N_0$ can be related to the input signal power by

$$A^2/2N_0 = \eta_{het}\frac{\eta_D P_S}{h\nu}\,, \quad (5)$$

where η_{het} is the receiver heterodyne efficiency, η_D is the detector quantum efficiency, P_S is the incident signal power, and $h\nu$ is the photon energy.

The effect of oscillator frequency noise can also be analyzed based on the simplified loop block diagram

shown in Fig. 1 [8]. The transfer function from $f_N(t)$ to $\phi_e(t)$, $H_2(s)$, is related to $H_1(s)$ by

$$H_2(s) = [1 - H_1(s)]\frac{1}{s}, \qquad (6)$$

and the phase error variance contributed by the frequency noise can be written as

$$\sigma_{\phi_f}^2 = \int_0^\infty |H_2(j2\pi f)|^2 S_f(f) df , \qquad (7)$$

where $S_f(f)$ is the power spectral density of the oscillator frequency noise.

In general, the short term frequency noise of lasers consists of three major components: a $1/f$ or $1/f^2$ noise at low frequencies, a white frequency noise, and spectral peaks that correspond to relaxation oscillations at higher frequencies. For semiconductor laser-based PLLs, it is generally assumed that the tracking loop bandwidth is sufficiently wide to track out the low frequency noises. The relaxation oscillation peaks are usually at a frequency much higher than the tracking loop bandwidth such that their effect on the loop performance can be ignored. As a result, the white frequency noise dominates the phase tracking response of the loop. Typical semiconductor lasers have white frequency noise floor as high as several MHz [3].

In contrast, diode-pumped solid state lasers have been shown to possess superior frequency stability [3],[12]. The relaxation oscillation is located at several hundred kHz that makes electronic feedback compensation possible. Furthermore, frequency noise measurement conducted using Lightwave^{TM} Model 120-01 lasers have shown that diode-pumped lasers have an inherent low white frequency noise approaching the Schawlow-Townes limit. The measured frequency noise power spectral density of the Lightwave lasers is shown in Fig. 2 [12]. The spectral peak at 400 Hz is due to the FM calibration signals. Also shown in the figure is an empirical model of the laser frequency noise consisting a white component, a $1/f$ component, and a $1/f^2$ component. Unlike semiconductor lasers for which the white frequency noise dominates the frequency noise PSD, frequency stability of diode-pumped solid state lasers are dictated by the low frequency noise components. Furthermore, the magnitude of the white frequency noise is measured to be less than $0.2\ Hz^2/Hz$.

Based on the measured frequency noise shown in Fig. 2, the PSD of the laser frequency noise can be modelled as

$$S_f(f) = k_0 + \frac{k_1}{f} + \frac{k_2}{f^2} . \qquad (8)$$

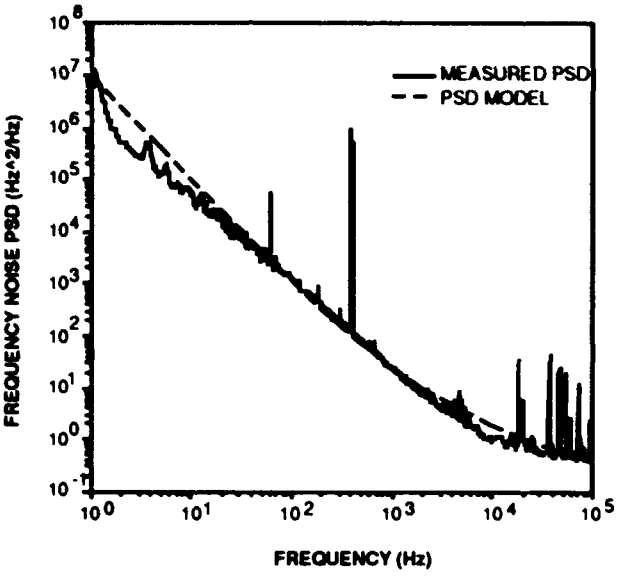


Fig.2 Measured power spectral density of the frequency noise of Lightwave Model 120-01A lasers. Also shown in the figure is the approximated frequency noise PSD including $1/f$ and $1/f^2$ noises.

The presence of $1/f$ and $1/f^2$ poles suggests that tracking loops with compensating zeroes at dc are needed to effectively track out the low frequency fluctuations. For our analysis, it will be assumed that a perfect second order loop is used to compensate for the frequency poles. A perfect second-order loop has the advantage of being unconditionally stable [7,8]. The unconditional stability ensures that the loop performance is less sensitive to small parameter changes. Furthermore, a perfect second order loop can track a constant Doppler signal without a static phase offset, and is thus much less sensitive to both the Doppler change and long term frequency drift of the lasers. Finally, a perfect second order loop has a theoretically infinite pull-in range [8], and can thus allow a more efficient phase acquisition. The transfer function of the loop filter is given by

$$F(s) = \frac{1 + \tau_2 s}{\tau_1 s} . \qquad (9)$$

By substituting the noise model in Eq.(8) and the filter transfer function into Eq.(7), the phase error variance due to frequency noise can be calculated in closed-forms. The results are:

$$\sigma_{\phi_f}^2 = k_0 g_0(\zeta)\frac{1}{B_L} + k_1 g_1(\zeta)\frac{1}{B_L^2} + k_2 g_2(\zeta)\frac{1}{B_L^3} \qquad (10)$$

where $g_0(\zeta)$, $g_1(\zeta)$, and $g_2(\zeta)$ are defined by [13]

$$g_0(\zeta) = \frac{4\zeta^2+1}{64\zeta^2}$$

$$g_1(\zeta) = \begin{cases} \frac{(4\zeta^2+1)^2}{512\zeta^3\sqrt{\zeta^2-1}} \ln\left[\frac{\zeta^2-\frac{1}{2}+\zeta\sqrt{\zeta^2-1}}{\zeta^2-\frac{1}{2}-\zeta\sqrt{\zeta^2-1}}\right], & (\zeta > 1) \\ \frac{25}{128}, (\zeta = 1) \\ \frac{(4\zeta^2+1)^2}{256\zeta^3\sqrt{1-\zeta^2}}\left[\frac{\pi}{2} - \tan^{-1}\left(\frac{\zeta^2-\frac{1}{2}}{\zeta\sqrt{1-\zeta^2}}\right)\right], \\ \qquad\qquad (\zeta < 1) \end{cases}$$

$$g_2(\zeta) = \frac{\pi^2(4\zeta^2+1)^3}{1024\zeta^4}$$

respectively. The damping factor, ζ, is given by [8]

$$\zeta = \frac{1}{2}\sqrt{\frac{\tau_2^2 A K_V}{\tau_1}}. \qquad (14)$$

Eq.(10) suggests that, given the choice of perfect second order loop filter, the loop architecture can be determined by its bandwidth and the damping ratio.

The choice of damping factor can be made by inspecting the values of the relative weight of functions $g_0(\zeta)$, $g_1(\zeta)$, and $g_2(\zeta)$. Plotted in Fig. 3 are values of these functions versus the damping factor, ζ. For a given loop bandwidth, tracking loops with damping factor less than 0.5 would result in a much higher phase tracking error variance than loops with damping factor greater than 0.5. Furthermore, depending on the relative weight of different frequency components, the damping factor can be chosen which minimizes the frequency noise contributions to output phase error variance.

Given the loop damping factor, the performance of the phase-locked loop can be calculated from Eqs.(2) and (10) as a function of the loop bandwidth and signal-to-noise ratio. Shown in Figure 4 are contributions to the total phase error variance from each of the noise components, as well as the total phase error variance. It is seen from Fig. 4 that performance of the PLL is frequency noise limited when the loop bandwidth is small. This is because small loop bandwidths, the loop admits very little additive noise, but also cannot effectively respond to input frequency fluctuations. As the loop bandwidth increases, the phase error variance decreases as more frequency fluctuation is tracked out by the loop. At higher loop bandwidths, the loop becomes additive shot noise limited and the phase error variance begins to increase with increasing loop bandwidth. Consequently, an optimal loop bandwidth exists which minimizes the phase error variance.

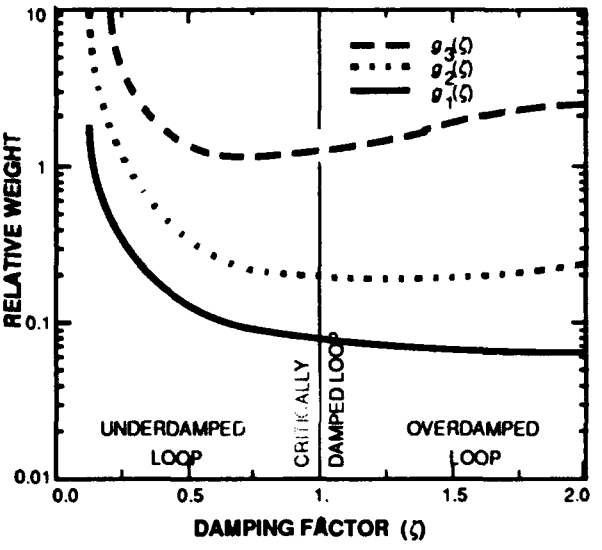


Fig. 3 Relative weights of the g-functions versus the loop damping factor of the loop. For $\zeta \ll 0.5$ the loop is sensitive to the frequency noise.

Furthermore, the optimal loop bandwidth increases as the SNR increases. This is because at higher SNR, better tracking performance can be obtained by widening the loop bandwidth.

III. Short Term Frequency Stabilization

As was mentioned in the previous section, the performance of the PLL is frequency noise limited at small loop bandwidths. Furthermore, the optimal loop bandwidth decreases with decreasing signal-to-noise ratio. For applications where the signal power is at a premium such as in a deep-space link, the low available signal power implies that lasers with excellent frequency stability will be needed to achieve phase coherent communications.

It is of great interest to determine the signal power requirement in optical phase tracking. By assuming that a PLL with optimal loop bandwidth is used to track the incoming signal, the phase error variance can be calculated as a function of the carrier SNR. Shown in Figure 5 is a plot of the optimal phase error variance at different carrier SNRs. The figure was generated by assuming that the transmit and LO lasers have a combined frequency noise spectrum equal to those of free-running Lightwave lasers. It is seen from Fig. 5 that an RMS phase tracking error of 0.3 rad can be achieved with a carrier SNR of 40 dB-Hz. If it is further assumed that an E_b/N_0 of 0 dB is needed to

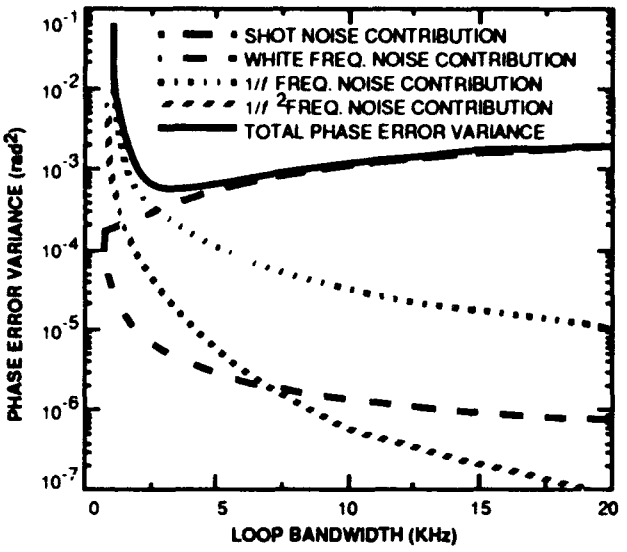


Fig.4 Contributions of each noise components to the total phase error variance versus the PLL loop bandwidth for A^2/N_0 of 70 dB-Hz. The frequency noise paraemters are $k_0 = 0.2$ Hz, $k_1 = 1.5 \times 10^4 \text{Hz}^2$, and $k_2 = 10^7 \text{Hz}^3$.

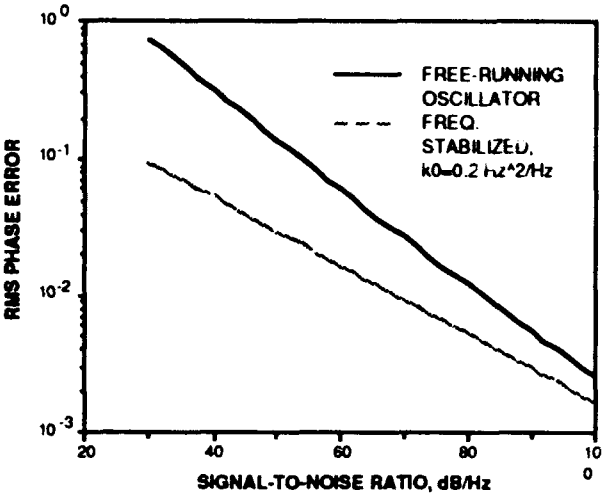


Fig.5 Optimal rms phase error variance versus the signal to noise ratio. The top curve is generated for PLL operating using un-stabilized laser with frequency noise shown in Fig. 1, and the bottom curve is generated for PLL using frequency stabilized lasers

effectively decode the incoming data, it follows that a 10 kbps link can easily meet the signal power requirement for coherent phase tracking. The absolute power requirement can also be estimated. By assuming that the receiver has a heterodyne efficiency of 50% and a detector quantum efficiency of 50%, it is seen that adequate phase locking can be achieved with signal power as low as 10 fW.

The performance calculated in Fig. 5 was for free-running lasers that exhibit large $1/f$ and $1/f^2$ frequency noises. As can be seen from Eq.(10), the higher order frequency contribution decreases rapidly with increasing frequency. Consequently, stabilization techniques can be applied which effectively compensates for these noises. Shown in Fig. 6 is a simplified block diagram of a frequency stabilization loop, which consists of a frequency discriminator, a loop filter, and a tunable LO laser. The resulting Allan variance of the stabilized loop can be calculated as [15]

$$\sigma_f^2(\tau) = 2\int_0^\infty |1 - H_f(f)|^2 S_f(f) \frac{\sin^4 \pi f\tau}{(\pi f\tau)^2} df + 2\frac{N_A}{G_f^2}\int_0^\infty |H_f(f)|^2 \frac{\sin^4 \pi f\tau}{(\pi f\tau)^2} df , \quad (15)$$

where τ is the measurement interval, $H_f(s)$ is the

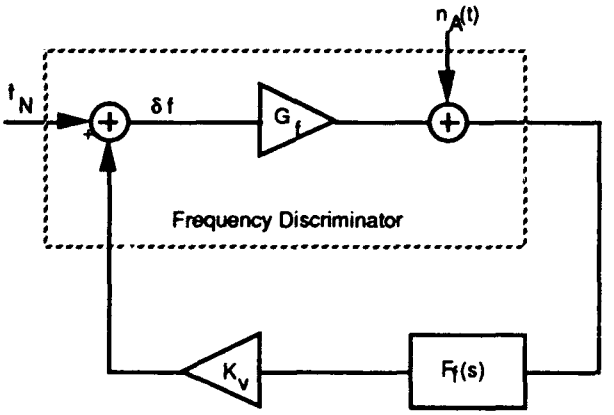


Fig.6 Block diagram of a frequency stabilization loop showing both amplitude and frequency noises.

transfer function of the stabilization loop, N_A is the PSD of the additive noise, and G_f is the gain of the frequency discriminator. The frequency discriminator can be implemented using either an AM scheme in which the transmitted(reflected) amplitude near a resonance is stabilized, or an FM sideband technique [14] in which the phase of the reflected (transmitted) signal from a resonance is detected.

The first term on the right hand side of Eq.(15)

is the residual frequency error due to incomplete compensation of the oscillator frequency noise. The frequency stabilization loop also introduce additional frequency noise since any amplitude noise at the VCO input will result in a frequency deviation at the oscillator output. The additive noise contribution to the output frequency error is summarized in the second term on the RHS of Eq.(15). Equivalently, the frequency stabilization process alters the output frequency spectrum of the laser to $S'_f(f) = |1 - H_f(f)|^2 S_f(f) + N_A |H_f(f)|^2 / G_f^2$. As was in the case of carrier phase tracking, a perfect second order tracking loop is desired to compensate for the $1/f$ and $1/f^2$ frequency poles.

The performance of the frequency stabilization loop shown in Fig. 6 can be easily analyzed in a manner similar to that used in the PLL analysis. It can be shown that the additive noise contribution increase with increasing loop bandwidth, whereas the effective frequency error due to the frequency noise decreases with increasing loop bandwidth. For systems operating with sufficiently high signal to noise ratio, the contribution due to additive noise can in general be ignored. In this case a wideband tracking loop may be applied to provide adequate frequency compensation. For the case of Lightwave™ lasers, calculation shows that an Allan deviation of approximation 1 Hz can be achieved with laser power as low as 1 mW using a 10 kHz tracking loop.

Since the phase error variance is dictated by the $1/f$ and $1/f^2$ poles at low signal power, the performance of the PLL can be significantly improved if both transmit and local oscillator are frequency stabilized. If it is further assumed that the stabilization process effectively removes the $1/f$ and $1/f^2$ poles such that the frequency noise consists only the white frequency noise with PSD k_0, then the optimal phase error variance can be calculated as

$$\sigma^2_{\phi,opt.} = 2\sqrt{\frac{k_0 g_0(\zeta)}{\rho}}, \qquad (16)$$

where $\rho = A^2/N_0$. The optimum RMS phase tracking error of a PLL operating with $k_0 = 0.2\text{Hz}^2/\text{Hz}$ was calculated and plotted in Fig. 5. Compared to the rms phase error for systems operating with un-stabilized lasers, It is seen that significant performance improvement can be expected if both transmit and LO lasers are frequency stabilized. In fact, the curve in Fig. 5 suggests that an adequate tracking performance can be expected for SNR as low as 20 dB-Hz, or equivalently, a signal power as low as 0.1 fW.

IV. Long Term Frequency Stabilization

In addition to data communications, the availability of a frequency stabilized reference can also benefit the spacecraft navigation effort. Current spacecraft navigation techniques rely heavily on Doppler tracking and range measurements to determine spacecraft position and velocity. Doppler signal extraction typically employes the two-way measurement technique. An Earth-based reference station transmits a signal to the spacecraft, where it is coherently multiplied in frequency, re-transmitted, and subsequently received at the original transmitter or another ground station. The two-way technique is preferred over the one-way measurement because the stable reference oscillator is more easily obtainable on the ground. For spacecraft navigation within the solar system, Doppler measurement can provide 5 of the 6 parameters describing the spacecraft location and velocity [6]. The radial distance can then be derived by integrating the equation of motion governing the spacecraft, or by using a ranging code.

In order to perform adequate Doppler tracking, the reference oscillator must remain stable over a time period that is comparable to the round trip light time between the ground station and spacecraft. Currently, the uncertainty in measuring the Doppler signal at S band (2.3 GHz) is limited by the plasma scintillation along the signal path to approximately one part in 10^{12} over 60 seconds of integration time [16]. This uncertainty is several orders of magnitude larger than that which is introduced by the oscillator, and is also several orders of magnitude higher than other systematic noises such as tropospheric scintillation. Since the effect due to charge particle decreases with $1/f^2$ [17], the uncertainty in Doppler measurement will be significantly lower at higher carrier frequencies. In fact, the projected Doppler uncertainty for the Cassini spacecraft, which operates at Ka band (32 GHz), is on the order of 10^{-16} over 1,000 seconds [17].

At the optical frequencies, the charge particle effect can be practically ignored; although the tropospheric attenuation does present a very serious problem at optical frequencies. In order to achieve comparable accuracy in measuring the spacecraft Doppler in S band, long term frequency stability better than one part in 10^{12} will be required. In the future, long term frequency stability better than one part in 10^{16} will be required to meet the projected Doppler tracking capability at Ka band. Currently, laser frequency stabilization using a stable tuned cavity can achieve a drift rate of approximately 1 Hz/sec [19]. A more

promising long term stabilization technique is via saturated absorption to an atomic resonance [18]. Both techniques can easily achieve the goal of one part in 10^{12} [18], although more research is needed to achieve a stability comparable to the RF oscillators.

V. Conclusions

Coherent space communications technology offers the potential of significantly improved data communication performances. By employing a phase coherent receiver aboard the spacecraft, the required signal power can be reduced by as much as 10-15 dB from direct-detection based systems. Furthermore, a coherent transponder at the optical frequency can offer potential improvement in Doppler measurements and spacecraft navigation. Analysis indicated that the short term frequency stability of diode-pumped solid state laser can easily meet the requirement for data communication channel if the data rate exceeds 10 kbps. In fact, coherent phase tracking can potentially be implemented with received signal power as low as several fW. The required signal power of the phase locked receiver can be further reduced provided that both transmit and LO lasers can be stabilized to stable external references.

The long term frequency stability requirement can be derived for a spacecraft Doppler tracking subsystem. The current S band system can achieve one part in 10^{12} accuracy. This performance can easily be met at the optical frequency since the dominating source of error, namely the plasma scintillation, is practically negligible at the optical frequency. However, a long term laser frequency stability better than one part in 10^{16} will be required to provide a performance comparable to that which is projected at Ka band.

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The author would like to acknowledge the helpful discussion of Dr. W. M. Folkner on the RF Doppler tracking subsystem.

References

[1] V. W. S. Chan, "Space coherent optical communication systems – An Introduction," *IEEE J. Lightwave Technol.*, Vol.LT-5, no. 4, pp.633-637, 1987.

[2] R. M. Gagliardi, and S. Karp, *Optical Communications*. New York: Wiley, 1975.

[3] C. H. Henry, "Theory of linewidths of semiconductor lasers," *IEEE J. Quantum Electron.*, Vol. QE-18, p. 259, 1982.

[4] T. Y. Fan and R. L. Byer, "Diode laser-pumped solid-state lasers," *IEEE J. Quantum Electron.*, vol. QE-24, no. 6, pp. 895-912, June, 1988.

[5] L. G. Kazovsky and D. A. Atlas, "Miniature Nd: YAG lasers: noise and modulation characteristics" *IEEE J. Lightwave Technol.*, vol. LT-8, no. 3, pp. 294-301, June, 1990.

[6] J. H. Yuen, *Deep Space Telecommunications Systems Engineering*, New York: Plenum, 1983.

[7] A. J. Viterbi, *Principles of Coherent Communications*, New York :McGraw-Hill, Inc., 1962.

[8] W. C. Lindsey, *Synchronization Systems in Communication and Control*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1972, pp.135, table 4-1.

[9] T. Day, A. Farinas, and R. Byer, "Demonstration of a low bandwidth 1.06 μm optical phase locked loop for coherent homodyne communication," *IEEE Photonic Technology Lett.*, vol. 2, no. 4, pp. 294-296, Apr., 1990.

[10] T. J. Kane, E. A. P. Cheng, "Fast frequency tuning and phase locking of diode-pumped Nd:YAG ring laser," *Opt. Lett.*, vol. 13, no. 11, pp. 970-972, Nov., 1988.

[11] L. G. Kazovsky, "Performance analysis and laser linewidth requirement for optical PSK heterodyne communication systems," *IEEE J. Lightwave Technol.*, vol. LT-4, no. 4, pp. 415-425, 1986.

[12] C.-C. Chen and M. Z.Win, "Frequency Noise Measurement of Diode-Pumped Nd:YAG Ring Lasers," *IEEE Photonic Technology Lett.*, vol. 2, no. 11, pp. 772-774, Nov., 1990.

[13] M. Z. Win, C. C. Chen, and R. A. Scholtz, "Optical phase locked loop for free space laser communications with heterodyne detection," *SPIE OE/LASE'91, Free-Space Laser Communication Technologies III*, Los Angeles, CA, January 1991.

[14] R. W. Drever, J. L. Hall, F. V. Kowalski,

J. Hough, G. Ford, A. J. Munley, and H. Ward, "Laser phase and frequency stabilization using an optical resonator," *Appl. Phys.*, vol. B31, pp. 77-105, 1983.

[15] J. A. Barnes, et al., "Characterization of Frequency Stability," *IEEE Transaction on Instrument and Measurement*, Vol. IM-20, no. 2, May 1971.

[16] W. E. Kirhofer, et al, "Galileo Tracking Data Success And Concerns," Internal Memorandum, Jet Propulsion Laboratory, IOM 314.3-904.

[17] A. L. Riley, et al., "Cassini Ka-Band Precision Doppler and Enhanced Telecommunications System Study," Jet Propulsion Laboratory, Internal Document.

[18] M. Nakazawa, J. Nakamura, and T. Musha, "FM eliminated CH_4 locked frequency satilization of 3.39 μm He-Ne laser in dual feedback control," *IEEE J. Quantum Electron.*, vol. QE-16, no. 8, pp. 854-859, 1980.

[19] L. Hollberg, Private Communications.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Stabilization of Oscillator Phase Using a Fiber-Optic Delay-Line

Ronald T. Logan, Jr., Lute Maleki, Mehdi Shadaram*

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109

Abstract - **We describe an electronic oscillator phase stabilizer based on a fiber-optic delay-line. Excellent agreement between calculated and first experimental results was obtained for this fiber-optic stabilized electronic oscillator (FOSO) at 100 MHz and 7.8 GHz. Expressions are derived for the FOSO phase noise performance and optimum fiber length, based on the noise contributions of the various components. The predicted performance of the FOSO expected with improved system components is also calculated. The FOSO is tunable over an extremely wide frequency range, with phase noise performance rivaling that of good resonant-cavity stabilized oscillators which are constrained to narrow band operation.**

Introduction

The first implementation of a scheme for stabilization of the phase and frequency fluctuations in the output of an electronic oscillator by using an ultra-stable fiber-optic delay-line is reported. The principle of operation of the oscillator is similar to conventional radio-frequency (RF) coaxial delay-line stabilized oscillators. However, the fiber-optic stabilized oscillator (FOSO) employs an ultra-stable intensity modulated fiber-optic link as the delay-line, similar to systems used in the JPL/NASA Deep Space Network to distribute hydrogen maser reference frequencies over distances of up to 29 kilometers with fractional frequency stability greater than 10^{-15} for 1000 second averaging times [1]. In the FOSO, the use of a long, stable, fiber-optic delay at RF and microwave frequencies provides much greater sensitivity to frequency fluctuations of the electronic oscillator, compared to conventional coaxial delay-line implementations, resulting in greatly improved phase stability.

In this paper, we develop an analytical model for the FOSO performance, based on the open-loop characteristics of the fiber-optic delay-line and other system components. An expression is derived for the optimum fiber length to be used, given the noise parameters of the system components. The predictions of this model are then compared with the results of the first experimental laboratory trials, and are shown to be in excellent agreement. These results provide confidence in the predictive ability of the model, which is then used to illustrate the potential performance of the FOSO based on anticipated improvements in fiber-optic systems. Finally, these results are compared to the performance of several state-of-the-art microwave signal sources.

System Description and Analysis

The FOSO system configuration is depicted in Figure 1. Frequency fluctuations of the voltage-controlled-oscillator (VCO) are converted to baseband voltage fluctuations by the fiber-optic discriminator. This baseband signal is amplified, inverted, filtered, and fed back to the frequency (or phase) control input of the VCO to reduce the frequency fluctuations.

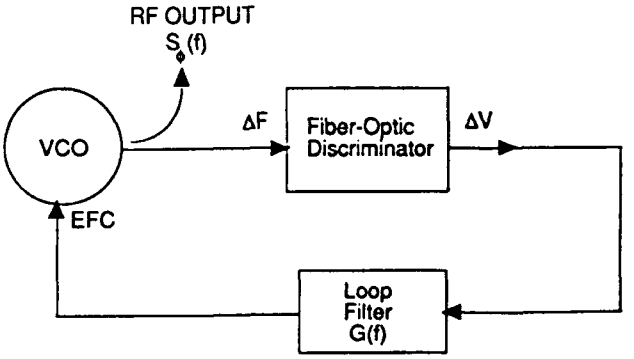


Figure 1. FOSO system configuration.

* This work represents one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract sponsored by the National Aeronautics and Space Administration.

** 1990 DOE/AWU Summer Faculty Research Fellow. Permanent Address: Electrical Engineering Department, University of Texas at El Paso, El Paso, Texas, 79968.

CH2965-2/91/0000-508 $1.00 © 1991 IEEE

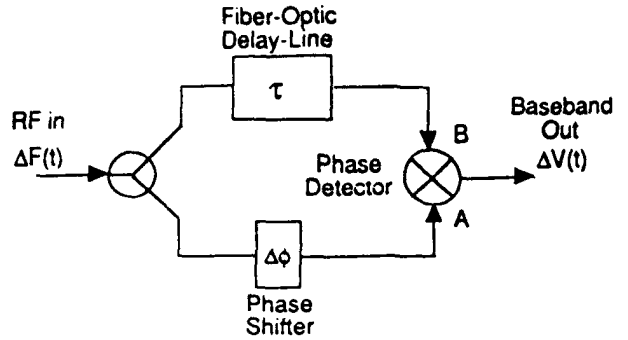


Figure 2. Fiber-optic frequency discriminator configuration.

In the fiber optic discriminator, shown in Figure 2, the RF input signal is split and compared to a delayed version of itself in a phase detector. The RF phase shifter is used to set the signals to 90-degree phase quadrature. The transfer function H(f) [V/Hz] of this configuration, assuming no loss, is:

$$H(f) = 2\pi\tau K_\phi \frac{\sin(\pi\tau f)}{\pi\tau f} e^{j(2\pi\tau f)} \qquad (1)$$

where f is the offset frequency of the frequency fluctuations from the RF signal , K_ϕ is the phase detector sensitivity in volts/radian, $\tau = nL/c$ is the delay time in the fiber which is L meters long with group index of refraction n, and c is the velocity of light in vacuum. For sufficiently low offset frequencies such that $f << 1/\tau$, this expression may be approximated by

$$H(f) \approx 2\pi\tau K_\phi \qquad (2)$$

Several points are important: First, the discriminator sensitivity increases linearly with delay time. Second, the delay-line discriminator can operate at any frequency within the bandwidth of the delay-line and phase detector, by simply setting phase quadrature at the phase detector inputs. The sensitivity is ultimately limited by the losses and bandwidth of the delay-line, but since optical fiber has approximately 1 dB/kilometer loss for modulation signals at microwave frequencies, multi-kilometer delay-lines at microwave frequencies are practical. Therefore, a highly sensitive, tunable frequency discriminator is possible when a fiber-optic delay-line is used.

To the extent that the frequency fluctuations of the VCO can be measured, they may be reduced. With appropriate feedback loop design, the phase noise of the VCO can be reduced to the equivalent phase noise of the fiber-optic discriminator, which is determined by the open-loop phase stability of its components and the loop parameters.

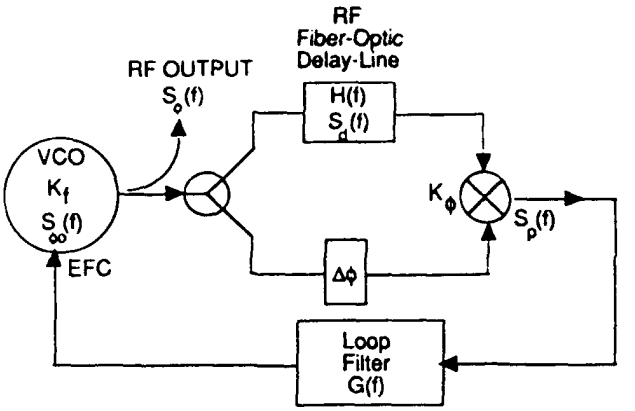


Figure 3. FOSO system with noise sources.

Figure 3 illustrates the FOSO with equivalent noise generator sources. The open-loop noise of the VCO is $S_{\phi o}(f)$, the fiber-optic delay-line phase noise is $S_{\phi d}(f)$, and the phase detector and loop amplifier combined equivalent phase noise is $S_{\phi p}(f)$. The baseband loop filter transfer function is G(f). For a VCO with a frequency tuning input, the phase noise of the FOSO calculated using standard loop analysis is:

$$S_\phi(f) = \frac{1}{(1 + G H K_f)^2} \cdot \left(S_{\phi o}(f) + |G K_f K_\phi|^2 \frac{(S_{\phi d}(f) + S_{\phi p}(f))}{f^2} \right) \qquad (3)$$

where K_f is the VCO tuning sensitivity in Hz/volt. The presence of the fiber-optic discriminator transfer function, H(f) in the denominator of Equation (3) is significant. Equation (1) indicates that H(f) has zeros at offset frequencies which are integral multiples of the reciprocal of the delay time, τ. Thus, the loop filter gain, G(f), must be adjusted to roll off at a low enough frequency so that the closed-loop response of the system will not exhibit oscillation at frequencies corresponding to the zeros of H(f).

For the presently available state-of-the-art, the fiber-optic phase noise, $S_{\phi d}(f)$, is much larger than the phase detector noise, $S_{\phi p}(f)$. Then, at offset frequencies low compared to $1/\tau$, in the limit of large loop gain, the expression for the closed-loop FOSO phase noise reduces to:

$$S_\phi(f) \approx \frac{S_{\phi d}(f)}{(2\pi\tau f)^2} \qquad (4)$$

As will be shown, the experimental results indicate that this simple expression gives a very good prediction of the phase noise performance of the FOSO, for offset frequencies close to the carrier (f less than 10 kHz).

Fiber Length Considerations

Equation (4) indicates that the open-loop phase noise of the free-running oscillator is reduced to a level limited by the fiber-optic link phase noise, and that the amount of improvement increases with the fiber length. However, the fiber-optic link phase noise is also a function of fiber length, due to attenuation and interaction between the laser linewidth and fiber dispersion. Considering only those losses due to attenuation, the phase noise at frequency offsets greater than about 1 kHz from the RF carrier has the functional form:

$$S_{\phi d}(f) \approx \frac{P_{ln}\, 10^{(x\,\alpha\, L\,/\,10)} + P_{th}}{P_{sig}\, 10^{(2\,\alpha\, L\,/\,10)}} \qquad (5)$$

where α is the loss of the optical fiber in dB/meter, L is the length in meters, P_{ln} is the received RF noise power due to the optical source in the fiber-optic transmitter, P_{th} is the thermal noise floor of the fiber-optic receiver, and P_{sig} is the RF power of the received modulation signal. For an optical source which has intensity noise above the shot-noise limit, x = 2; x = 1 for a shot-noise limited source. Thus, for a laser not operating at the shot-noise limit, such as the semiconductor laser used in the experiments, the expression for the FOSO phase noise depends on fiber length as

$$S_{\phi}(f) \approx \frac{c^2}{(4\,\pi\, L\, n)^2} \left(\frac{P_{ln} + P_{th}\, 10^{(\alpha\, L\,/\,5)}}{P_{sig}} \right). \qquad (6)$$

Equation (6) implies that an optimum fiber length exists to yield the lowest FOSO phase noise for given values of laser noise and fiber loss. Differentiating this expression with respect to L yields a transcendental equation which can be solved to find the optimum fiber length. A non-zero solution for L exists provided the laser noise P_{ln} is larger than the thermal noise floor of the receiver, P_{th}, which is typically the case.

Experimental Results

Experiments were performed with a synthesized signal generator at 100 MHz, using a phase-feedback configuration, and with a cavity-tuned signal generator at 7.8 GHz in a frequency-feedback configuration [2]. In both cases, the RF delay-line of the fiber-optic discriminator in Figure 2 was an intensity modulated analog fiber-optic link. In this fiber-optic link, the RF signal modulated the injection current of a Fabry-Perot semiconductor laser diode, producing a linear intensity modulation. This modulated optical carrier was launched into single-mode optical fiber: 2.2 km in the 100 MHz trial, and 6.2 km in the 7.8 GHz trial. The optical power was detected by a PIN photodiode receiver, amplified, and applied to the phase detector input. An adjustable coaxial phase shifter was used in the other RF path to achieve phase quadrature between the delayed and un-delayed signals at the phase detector inputs.

The baseband output of the phase detector was amplified, filtered, and fed back to the phase (frequency) modulation input of the signal generator. The filter was not optimized to completely eliminate the oscillations at the zeros of the fiber-optic discriminator, described earlier, so phase noise peaks are observed at offset frequencies corresponding to the inverse of the delay time. Loop locking was easily achieved at any arbitrary operation frequency within the system bandwidth, by adjusting for quadrature between the phase detector inputs. Also, smooth, continuous, high-resolution frequency tuning was observed when locked, by simply adjusting the phase shifter.

Figures 4 and 5 illustrate the calculated and measured phase noise results at 100 MHz. Figure 6 illustrates the 7.8 GHz results, which clearly shows the X-band oscillator phase noise is reduced by over 40 dB, to the level predicted by Equation (4) at low offset frequencies. In both cases, calculated and experimental results are in excellent agreement.

The calculations were made using the analysis described earlier, with the measured phase noise of the open-loop VCO and the fiber-optic delay-line as inputs. All of the phase noise measurements at 100 MHz were performed using the phase-detector technique by phase locking to a low-noise quartz crystal oscillator. At 7.8 GHz, the open-loop fiber-optic discriminator noise was measured using a multiplied low-noise quartz crystal. The open-loop signal generator noise was then measured using the fiber-optic discriminator. The closed-loop FOSO performance was measured using a low-noise sapphire "whispering-gallery-mode" resonator discriminator described in Reference [3], phase-locked to the FOSO by temperature-tuning the sapphire.

Performance Projections

In this preliminary series of investigations, the calculated and measured performance of the FOSO was limited by the fiber-optic discriminator noise. The fiber-optic delay-line used in these first experiments represents the state-of-the-art in injection current modulated semiconductor laser-based fiber optic links. However,

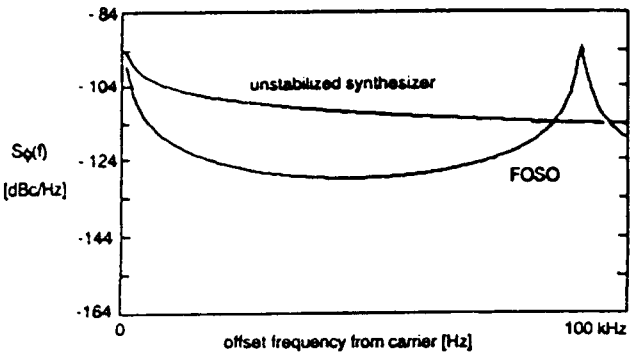


Figure 4. FOSO calculated phase noise performance at 100 MHz.

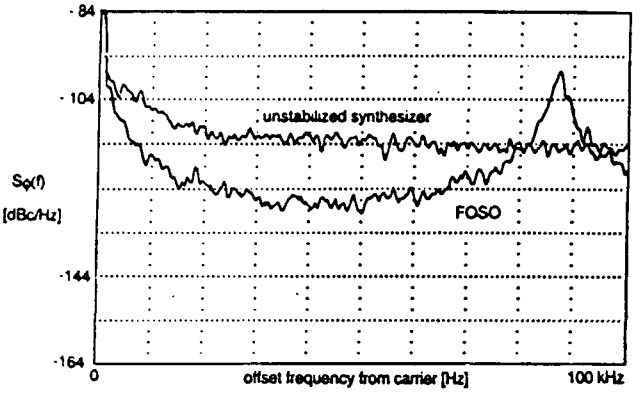


Figure 5. FOSO experimental results at 100 MHz.

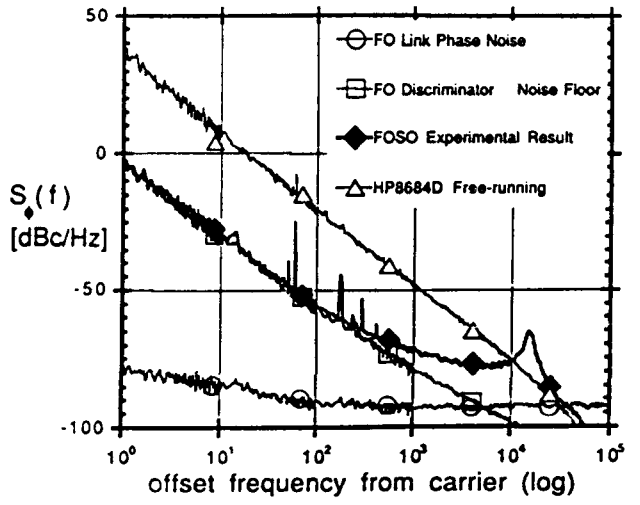


Figure 6. FOSO phase noise performance at 7.8 GHz.

with the development of improved fiber-optic links based on diode-pumped solid-state lasers and external intensity modulators, much higher signal-to-noise ratio (SNR) is possible [4, 5], and hence, improved phase noise performance.

Improved fiber-optic link phase noise performance has been recently demonstrated at 8.4 GHz (-108 dBc/Hz at f = 10 Hz, -130 dBc/Hz at f = 10 kHz) using an externally modulated diode-pumped Nd:YAG transmitter [6]. Calculations of the FOSO performance based on this improved fiber-optic delay-line are shown in Figure 7. Also shown are the previous FOSO experimental results, and typical phase noise performance of various signal sources at X-band. It is noted that the projected performance of the FOSO, a widely tunable source, rivals the performance of superconducting DRO sources, which are typically tunable only a few percent, and must operate at cryogenic temperatures.

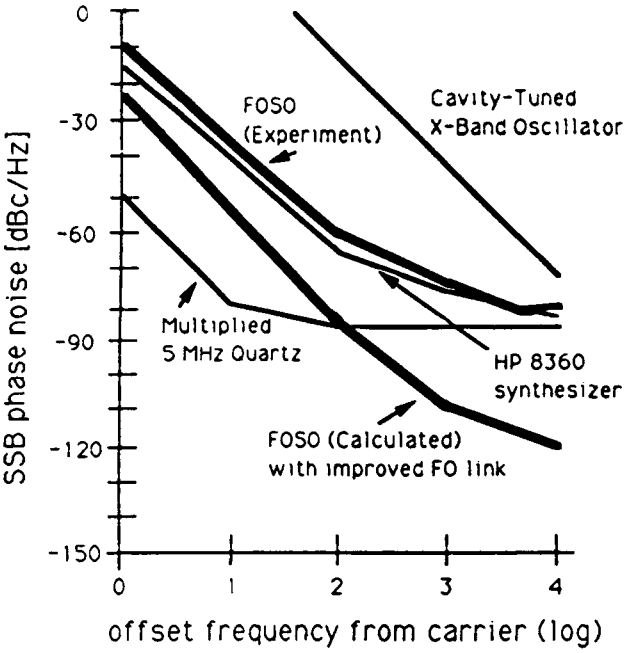


Figure 7. Comparison of FOSO performance at 7.8 GHz: Calculated FOSO performance using improved fiber-optic link, experimental FOSO data of Figure 6, and phase noise of various microwave signal sources.

A noteworthy asset of the fiber-optic delay-line discriminator is the fact that it is not a resonant cavity. Therefore, it can be used to measure and stabilize the frequency fluctuations of a widely tunable oscillator. This is a capability that resonant-cavity discriminators, which are constrained to operate at a fixed design frequency, fundamentally do not possess. Also, the delay-line discriminator "Q" increases linearly with delay length as well as frequency. Cavity resonators, however, typically exhibit "Q" roll-off as $1/f^2$ with increasing frequency due

to electromagnetic losses. The modulation bandwidth of present fiber-optic links is limited by the electro-optic modulation and detection devices. At the time of writing, frequencies as high as 64 GHz have been demonstrated for electro-optic modulators [7], and 60 GHz photodiodes are commercially available [8]. This high bandwidth potential of fiber-optic delay-lines makes the FOSO technique also attractive for stabilization of tunable millimeter-wave sources.

Summary

First results were reported for a fiber-optic stabilized electronic oscillator (FOSO). An analytical model was developed and used to predict phase noise performance at 100 MHz and 7.8 GHz that was in excellent agreement with experimental results. It was shown that the FOSO performance was limited by the fiber-optic delay-line phase noise, and that an optimum fiber length exists for given system parameters. The derived model was used to predict future FOSO performance based on a fiber-optic delay-line with improved phase noise performance. The phase noise performance of the experimental trial at 7.8 GHz was shown to be comparable to state-of-the-art tunable microwave sources for offset frequencies less than 10 kHz, while the projected phase noise performance is 15 to 30 dB better.

The FOSO technique can be used to stabilize a tunable electronic oscillator over the entire bandwidth of the fiber-optic delay-line, which can be tens of GHz. This is an important advantage compared to resonator-based oscillators which are constrained to operate at a fixed design frequency. Long fiber-optic delay-lines can be contained in small, light-weight, rugged packages whose components are all solid-state, low-power devices. These features may make the FOSO an attractive candidate for a stable, frequency-tunable microwave and millimeter-wave oscillator for many different applications.

Acknowledgements

The assistance of David Santiago, John Dick, George Lutes and Bill Diener and Ruthann Melbourne in the form of helpful discussions, measurement apparatus, and computer expertise is thankfully acknowledged.

References

[1] L. Primas, G. F. Lutes, R. T. Logan, Jr., "Ultra-Stable Fiber Optic Systems for Frequency Distribution," Proceedings of the 42nd Annual IEEE Frequency Control Symposium, pp. 202-211, Denver, CO, 31 May - 2 June, 1989.

[2] Signal generators used in the experiments: (1) Hewlett-Packard 8660C Synthesized Signal Generator at 100 MHz, using phase modulation input, (2) Hewlett-Packard 8684D Cavity-tuned Signal Generator at 7.8 GHz, using frequency modulation input.

[3] G.J. Dick, J. Saunders, T. Tucker, "Ultra-Low Noise Microwave Phase Stabilizer Using Sapphire Ring Resonator," Proceedings of the 44th Annual IEEE Frequency Control Symposium, Baltimore, MD, pp. 577-584, 23-25 May, 1990.

[4] R. T. Logan, Jr., G. F. Lutes, L. Maleki, "Microwave Analog Fiber Optic Link for use in the Deep Space Network," JPL/NASA TDA Progress Report, Pasadena, CA, pp. 21-33, February 15, 1990.

[5] G. E. Betts, C. H. Cox, III, K. G. Ray, "20 GHz Optical Analog Link Using an External Modulator," IEEE Photonics Technology Letters, Vol. 2, No. 12, pp. 923-925, December 1990.

[6] R.T. Logan, Jr., L. E. Primas, G. F. Lutes, L. Maleki, "Modulation Signal Stability Considerations in Analog Fiber Optic Systems," talk given at the First Annual DARPA/RADC Symposium on Photonics Systems for Antenna Applications, Monterey, CA, 13 December 1990.

[7] W. B. Bridges, F. T. Shechy, J. H. Schaffner, "Wave-Coupled $LiNbO_3$ Electrooptic Modulator for Microwave and Millimeter-Wave Modulation," IEEE Photonics Technology Letters, Vol. 3, No. 2, pp. 133-135, February 1991.

[8] 60 GHz photodiodes have recently become available commercially from New Focus, Inc., 340 Pioneer Way, Mountain View, CA 94041.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PROTOTYPE OF AN OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD

Kenji HISADOME and Masami KIHARA

NTT Transmission Systems Laboratories
1-2356 Take, Yokosuka, Kanagawa, 238-03 Japan

ABSTRACT

The prototype of an optically pumped cesium beam frequency standard is composed of three modules: a cesium beam tube, a microwave module, and a laser diode module. In the cesium beam tube, background lights from the windows are reduced. In the microwave module, an automatic level controller is used in order to reduce long-term frequency instabilities caused by the microwave level fluctuation. In the laser diode module, three laser diodes are connected to the cesium beam tube by polarization maintaining fibers and fiber couplers.

The short-term frequency stability (two-sample standard deviation), which is estimated from the preliminary experimental results gained from the prototype, is about $\sigma_y(\tau) = 1 \times 10^{-12}/\sqrt{\tau}$.

1. INTRODUCTION

Cesium beam frequency standards generate an accurate and stable frequency synchronized to the "clock transition" frequency (9.192631770 GHz) of cesium atoms [1]. Optically pumped cesium beam frequency standards use lasers instead of conventional deflection magnets. Their frequency accuracy, long-term frequency stability, and short-term frequency stability are expected to be better than conventional cesium beam frequency standards. With recent advances in laser diode technology, several laboratories are in the process of developing optically pumped cesium beam frequency standards [2]-[6].

NTT is developing a small optically pumped cesium beam frequency standard as a clock frequency source for synchronous digital communication networks [7]. In this paper, a prototype of the optically pumped cesium beam frequency standard is introduced. The prototype is composed of three modules: a cesium beam tube, a microwave module, and a laser diode module. Each of these modules is explained in Sections 2 through 4. Preliminary experimental results gained from the prototype are reported in Section 5.

2. CESIUM BEAM TUBE

The scheme of the cesium beam tube is shown in Fig. 1. The cesium beam tube is 590 mm long and is 90 mm shorter than the previous cesium beam tube [8]. It has two cesium beam ovens, four pairs of windows, and four spheroidal mirrors for fluorescence detection. These windows are made of anti-reflection (AR) coated fused silica.

Although windows in the previous cesium beam tube also were AR-coated, the measured background light levels were very large, and the background lights were noise sources. It is found that reflected lights from rear windows and scattered lights from front windows caused the background lights. In the new cesium beam tube, the rear windows are inclined and the reflected lights are absorbed by carbons installed inside the cesium beam tube. The front windows are moved away from the spheroidal mirrors. As a result of these improvements, the measured background light levels are reduced by 90% and the background light

CH2965-2/91/0000-513 $1.00 © 1991 IEEE

level to input laser power ratio is about 4×10^{-6}.

An aluminium alloy, which is nonmagnetic and emits little gas, is used as the material for the cesium beam tube. Two ion pumps, whose evacuation speeds are 8×10^3 cm^3/sec, are used. The vacuum level in the beam tube can be held to less than 10^{-7} Torr.

A three layered permalloy magnetic shield is used. In order to improve uniformity of the magnetic field in the shield, the cesium beam tube, with the exception of the ion pumps, is enclosed by the shield and connected to other modules by a coaxial cable and optical fibers.

switch and is digitized by an AD converter. The computer reads the digitized resonant signals and modulation signals, derives frequency discrimination signals from the read signals, integrates the discrimination signals, and outputs the integrated discrimination signals as the frequency control signals. The frequency control signals are DA-converted and are applied to a VCXO.

The short-term frequency stability, whose averaging time is shorter than time constant of the frequency servo loop, is determined by the VCXO. An oven controlled electrodeless quartz crystal oscillator, whose nominal frequency is 5 MHz, is used as the VCXO of this optically

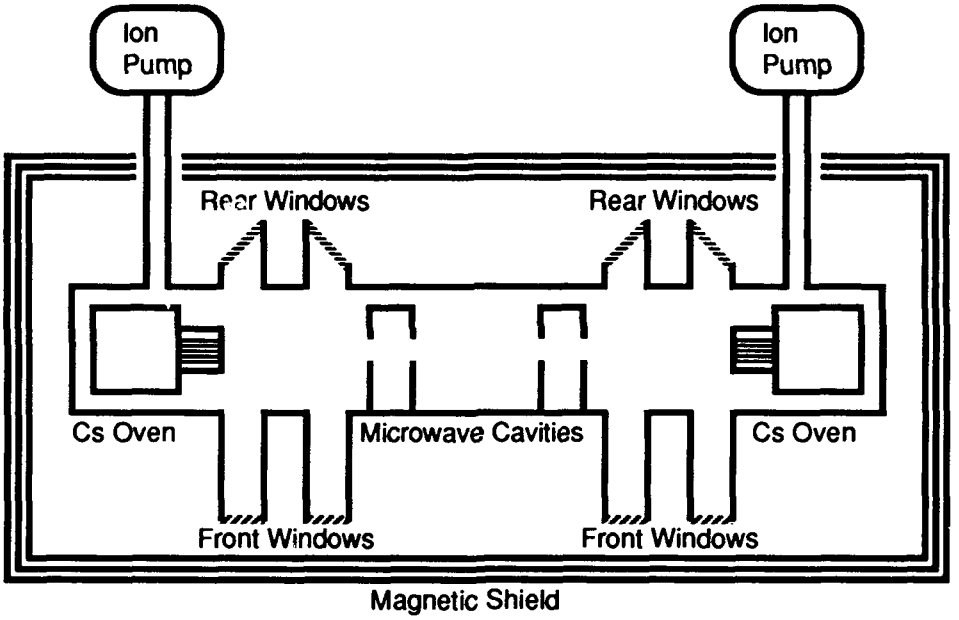


Fig. 1. Schematic diagram of the cesium beam tube.

3. MICROWAVE MODULE

The scheme of the microwave module is shown in Fig. 2. A digital frequency servo loop is employed in the microwave module. In the prototype, an external personal computer is used as a controller. One of two resonant signals from the cesium beam tube is selected by an input

pumped cesium beam frequency standard. According to the specifications, the two-sample standard deviation of the VCXO is less than 3×10^{-12} from 0.2 to 30 sec.

A 9.2 GHz oscillator and two 7.37 MHz synthesizers are phase-locked to the VCXO. Two 7.37 MHz synthesizer frequencies can be tuned by the computer, and the frequency difference between the two 7.37 MHz synthe-

sizer frequencies is equal to the full width at half maximum of the Ramsey fringe. One of the two 7.37 MHz synthesizer signals is selected by a modulation switch according to the modulation signal. This switching does not cause any modulation distortion. A double balanced mixer generates a 9.19263 GHz microwave from the 9.2 GHz microwave and the selected 7.37 MHz synthesizer signal. The 9.19263 GHz microwave is supplied to the cesium beam tube.

The 9.19263 GHz microwave level fluctuation changes several microwave resonant frequency offset values within the cesium beam tube. This phenomenon is called "power shift" and causes long-term frequency instabilities of the cesium beam frequency standards [9]. In order

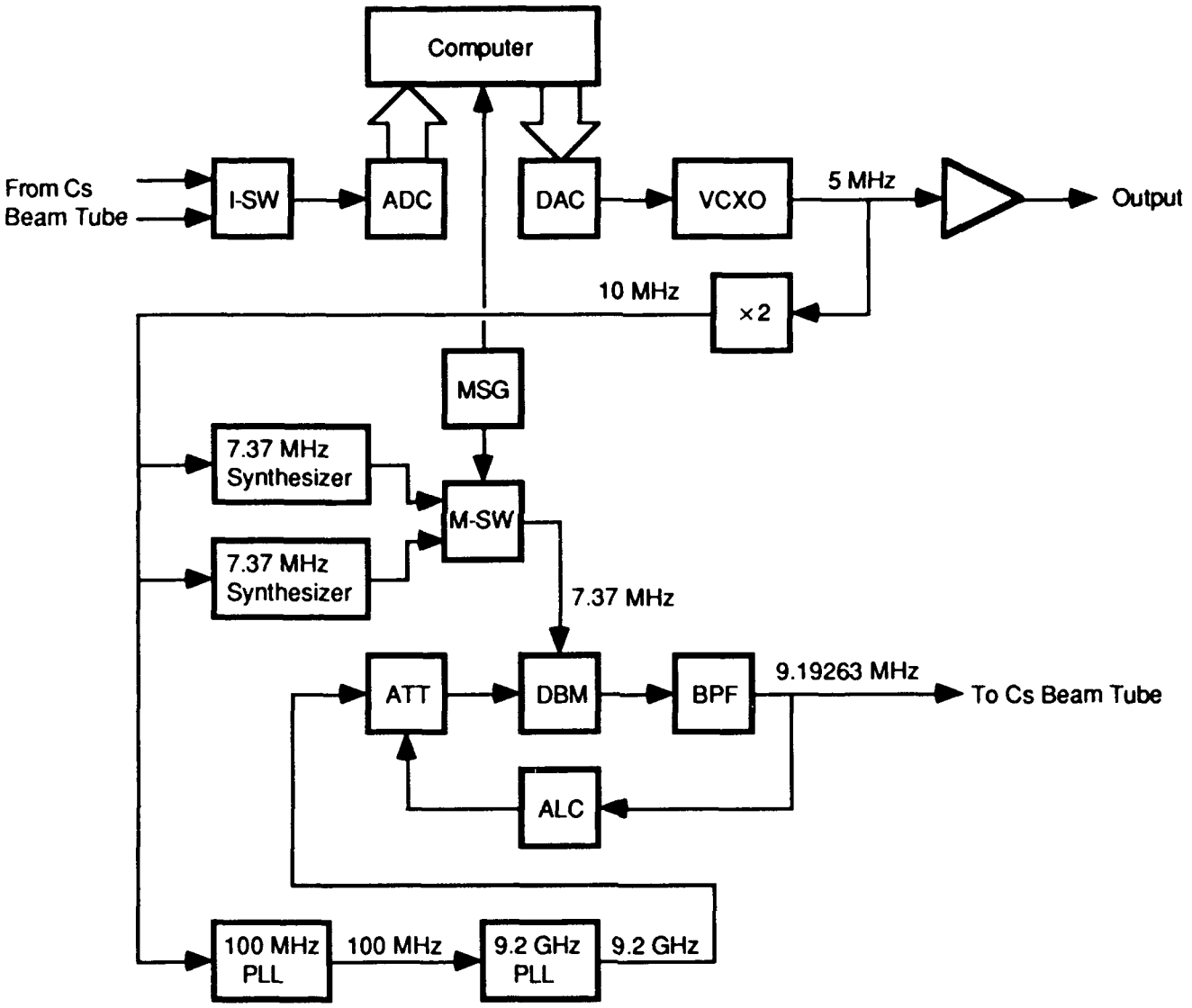


Fig. 2. Schematic diagram of the microwave module. In this figure, I-SW is an input switch, ADC is an analog to digital converter, DAC is a digital to analog converter, VCXO is a voltage controlled crystal oscillator, MSG is a modulation signal generator, M-SW is a modulation switch, ATT is an attenuator, DBM is a double balanced mixer, BPF is a band-pass filter, ALC is a automatic level controller, and PLLs are phase-locked loops.

to reduce the 9.19263 GHz microwave level fluctuations, an automatic level controller (ALC) is employed. The ALC measures the 9.19263 GHz microwave level, compares it with a reference level, and attenuates the 9.2 GHz microwave level to make the 9.19263 GHz microwave level coincide with the reference level. The reference level is set by the computer. As a result of the ALC, the 9.19263 GHz microwave level fluctuation is reduced to less than 0.2 dB from 10 to 40°C.

4. LASER DIODE MODULE

The scheme of the laser diode module is shown in Fig. 3. In the laser diode module, three commercial Fabry-Perot type GaAlAs semiconductor laser diodes are installed. Two laser diodes are used as pump lasers and the other one is used as a probe laser. External resonators, to narrow the

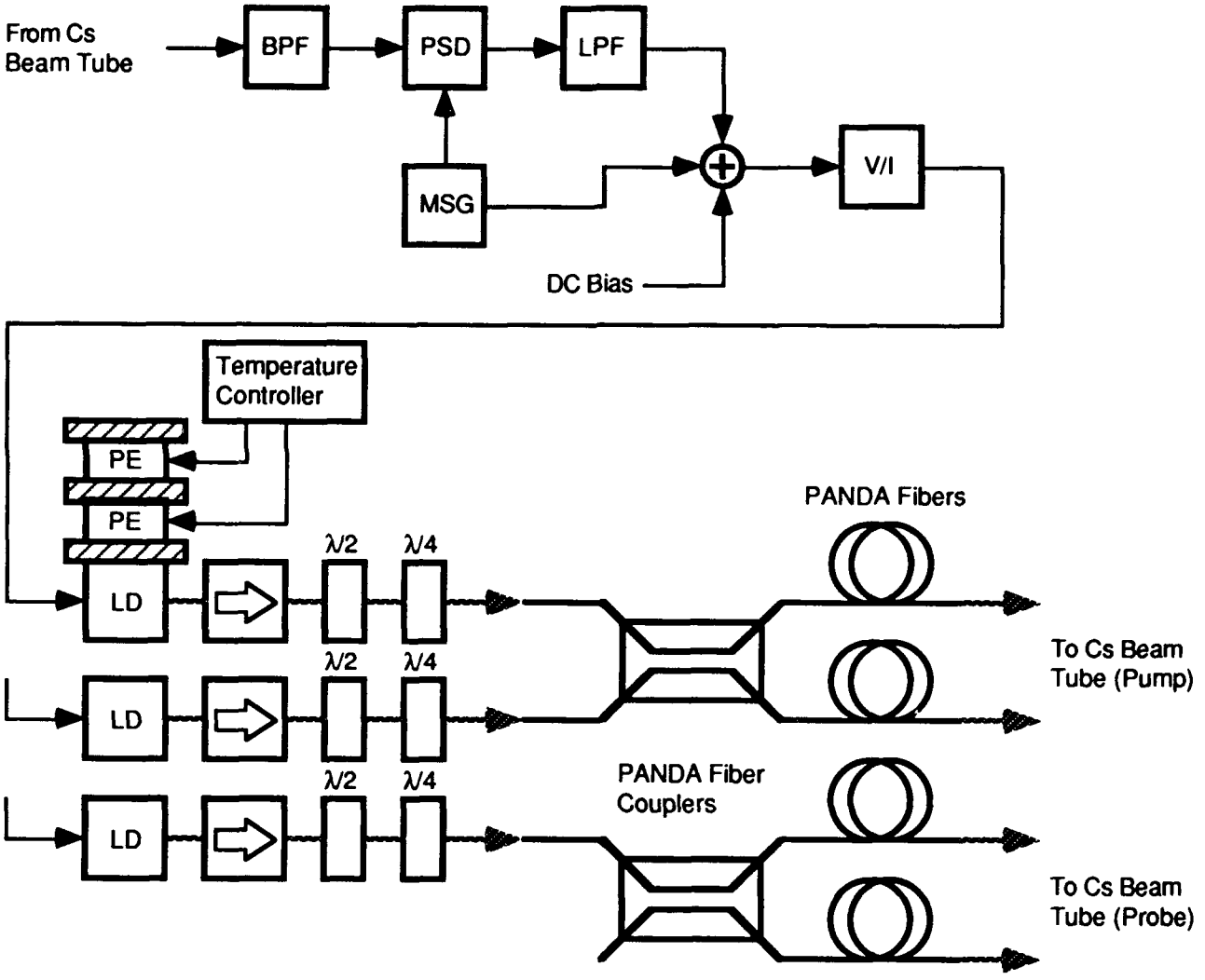


Fig. 3. Schematic diagram of the laser diode module. In this figure, BPF is a band-pass filter, PSD is a phase sensitive detector, LFP is a low-pass filter, MSG is a modulation signal generator, V/I is a voltage to current converter, LDs are laser diodes, PEs are Peltier elements, l/2s are half-wave plates, l/4s are quarter-wave plates, and PANDA fibers are polarization-maintaining and absorption-reducing fibers.

spectrum linewidth of the laser diodes, are not used at the present time. The temperature of each laser diode is stabilized within 10^{-2}°C by a two-stage Peltier cooler. The wavelength of each laser diode can be locked to one of the fluorescence peaks from the cesium beam or absorption lines in cesium vapor cell by an injection current controller. This controller is composed of a phase sensitive detector, a low-pass filter, and a voltage to current converter.

In the prototype, optical fibers are used to connect the laser diodes to the cesium beam tube. In general, the laser beam polarization state are changed by various disturbances, in a regular single-mode or multi-mode optical fiber. When the pump laser beam is elliptically or circularly polarized, σ^+-transition probabilities are not equal to σ^--transition probabilities and the populations in Zeeman sublevels become asymmetric. Even if the pump laser beam is linearly polarized, the pumping efficiency depends on the angle between the electrical field of the laser beam and the static magnetic field (C-field) applied to the cesium beam [10]. Therefore, the laser beam polarization fluctuations cause noise in the cesium beam tube. It is necessary to optimize the laser beam polarization states and to maintain the optimized polarization states. Polarization-maintaining and absorption-reducing (PANDA) fibers and fiber couplers are employed to maintain the laser beam polarization states. Half-wave plates and quarter-wave plates are used to optimize the polarization states at the output ends of the optical fibers. 60 dB optical isolators are necessary to reject the returned light from the above optical elements to the laser diodes. This is done because the returned lights to the laser diodes cause intensity and frequency instabilities in the laser diodes.

Although the output laser powers from the optical fibers are more than 150 μW, the powers are not sufficient. In order to increase the output laser powers, it is necessary to reduce optical coupling losses between the laser diodes and the optical fibers.

5. PRELIMINARY EXPERIMENTAL RESULTS

First, the fundamental optical pumping experiment is done. The experimental setup is shown in Fig. 4. In this experiment, one-laser pumping was used. The pump laser wavelength was swept over six cesium D2 transition lines by sweeping the DC bias voltage. The probe laser wavelength was locked to the cycling transition line F = 4

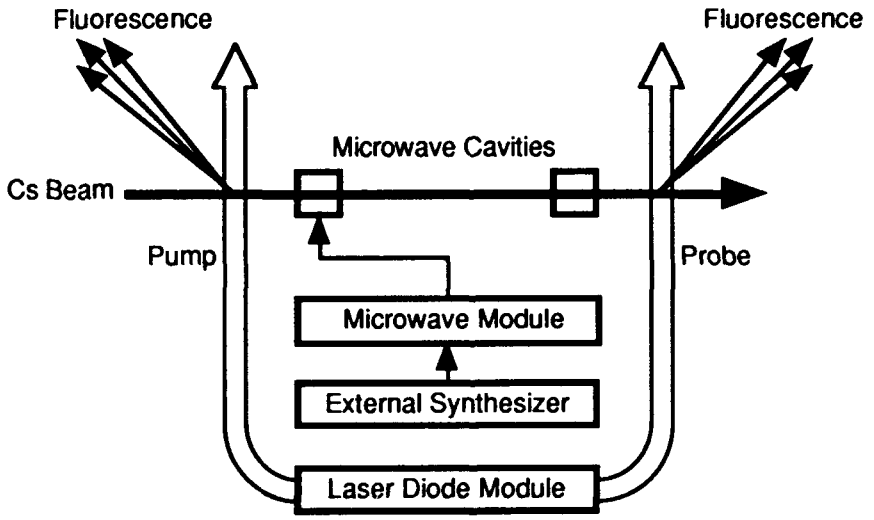


Fig. 4. Schematic diagram of the preliminary experiment setup.

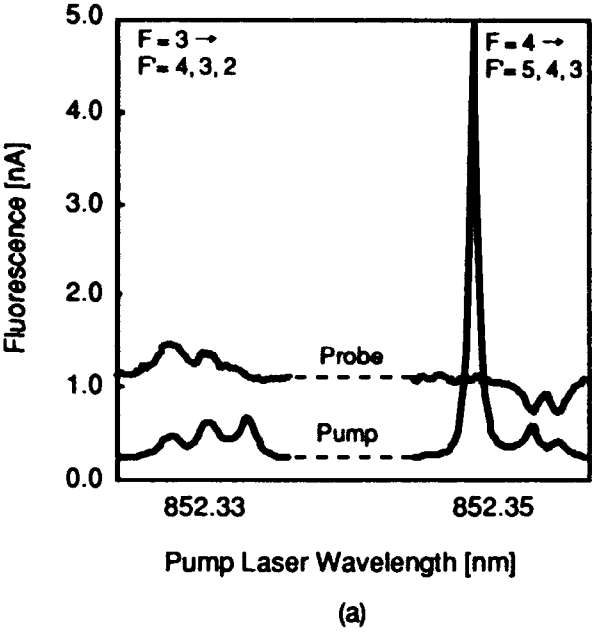
→ F' = 5. The microwave was not supplied. F and F' indicate total angular momentum quantum numbers in the ground and excited states respectively.

The experimental results are shown in Fig. 5 (a). In this experiment, the cesium beam oven temperature was 120°C. When the probe laser wavelength was not tuned to any transition line, the observed background light level was 0.6 nA. When it was tuned to the cycling transition line, the observed light level was 1.1 nA. Therefore, the net fluorescence level in this measurement was 0.5 nA. When the pump laser wavelength was tuned to pumping transition lines F = 3 → F' = 3 or 2, the probe fluorescence level increased. Also, when it was tuned to pumping transition lines F = 4 → F' = 3 or 4, the probe fluorescence level decreased. A magnified portion of the Fig. 5 (a) is shown in Fig. 5 (b). In this figure, the background light level (0.6 nA) has been offset. When the pump laser wavelength was tuned to the pumping transition line F = 3 → F' = 2, the probe fluorescence level was about 0.8 nA. If the coupling loss between the laser diode and the optical fiber was smaller, the probe fluorescence level was about 1 nA.

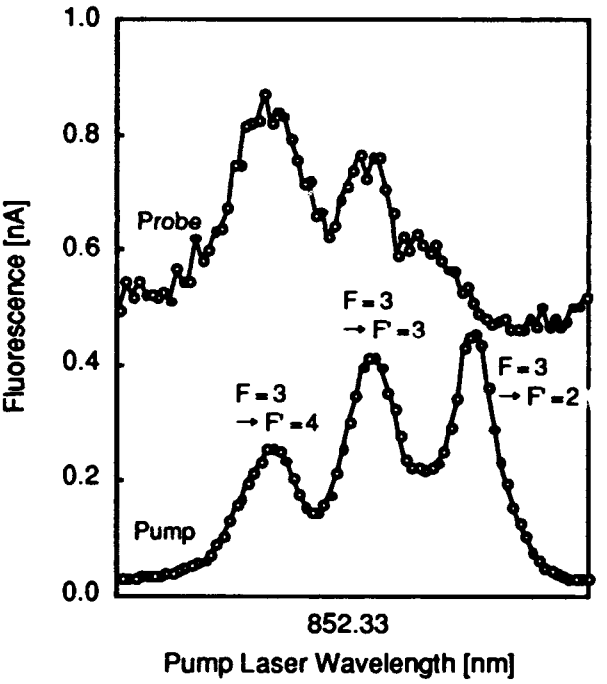
Next, the microwave resonant experiments were done. Only the Rabi resonance was measured because a U-shaped waveguide, which is necessary to observe the Ramsey resonance, is not available at the present time. In this experiment, one-laser pumping was also used. The pump laser wavelength was locked to the pumping transition line F = 3 → F' = 2. The microwave frequency was swept by sweeping the external synthesizer frequency. The probe laser wavelength was locked to the cycling transition lines F = 4 → F' = 5. Zeeman sublevels were degenerate because the C-field was not applied.

The experimental results are shown in Fig. 6. When the microwave frequency was tuned to the clock transition frequency, the optically pumped cesium atoms were stimulated to transit F = 4 → F = 3, and the probe fluorescence decreased to 0.5 nA. Observed full width at half maximum of the Rabi resonance was 24 kHz and agreed with the theoretical value.

Using the measured beam current value, the short-term frequency stability of the prototype can be estimated. In this estimation, four assumptions are made: (1) two-



(a)



(b)

Fig. 5. (a) Optical pumping experimental results and (b) magnified portion.

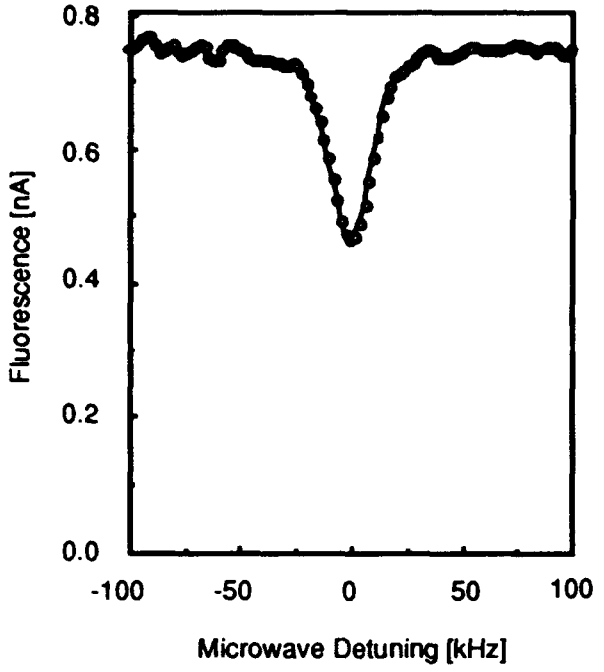


Fig. 6. Rabi resonant spectrum gained from the prototype. The background light level has been offset.

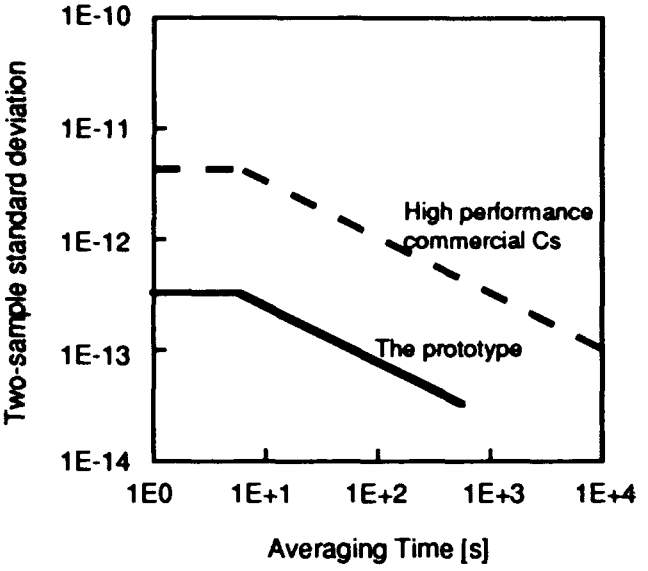


Fig. 7. Estimated short-term frequency stability of the prototype of the optically pumped cesium beam frequency standard. The short-term frequency stability of a conventional high performance commercial cesium beam frequency standard is also presented.

laser pumping is used, (2) the full width at half maximum of the Ramsey fringe is 600 Hz, (3) the product of photon number per atom, fluorescence collection efficiency of the spheroidal mirror, and the quantum efficiency of the photodiode is 30, and (4) the beam shot noise is dominant. The estimated short-term frequency stability (two-sample standard deviation) is about $\sigma_y(\tau) = 1 \times 10^{-12}/\sqrt{\tau}$, and the values are shown in Fig. 7.

6. SUMMARY

The prototype of an optically pumped cesium beam frequency standard is created. It is composed of a cesium beam tube, a microwave module, and a laser diode module. The prototype's features are as follows:

1. The measured background light levels in the cesium beam tube are reduced to less than 1/10 of the background light levels in the previous cesium beam tube.
2. To reduce long-term frequency fluctuations, the microwave level fluctuation is reduced to less than 0.2 dB from 10 to 40°C by the automatic level controller.
3. To optimize and maintain the laser beam polarization states, polarization maintaining optical fibers and optical fiber couplers are used.

Using the preliminary experimental results gained from the prototype, it is estimated that the short-term frequency stability of this cesium beam frequency standard is about $\sigma_y(\tau) = 1 \times 10^{-12}/\sqrt{\tau}$.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Tetsuya Miki, Executive Manager and Mr. Katsuaki Kikuchi, Research Group Leader of NTT Transmission Systems Laboratories, for their encouragement and suggestions.

REFERENCES

[1] J. Vanier and C. Audoin, *The quantum physics of atomic frequency standards*, Bristol, England: Adam Hilger, 1989, Chapter 5.

[2] S. Ohshima, Y. Nakadan, T. Ikegami, Y. Koga, R. Drullinger, and L. Hollberg, "Characteristics of an optically pumped Cs frequency standard at the NRLM", *IEEE Trans. Instrum. Meas.*, IM38, 2, pp. 533-536, 1989.

[3] V. Gandelier, V. Giordano, A. Hamel, G. Théobald, P. Cérez, and C. Audoin, "Frequency stability of an optically pumped cesium beam frequency standard", *Appl. Phys.*, B49, pp. 365-370, 1989.

[4] P. Thomann, H. Schweda, and G. Busca, "Preliminary results on small optically pumped cesium resonator", in *Frequency standards and metrology*, ed. A. DeMarchi, pp. 392-394, Berlin, German: Springer-Verlag, 1989.

[5] L. L. Lewis, "Low noise laser for optically pumped cesium beam standards", in *Proc. 43rd Annual Symposium on Frequency Control*, Denver, Colorado, pp. 151-157, 1989.

[6] R. E. Drullinger, "Optically pumped primary frequency standard", in *Proc. 44th Annual Symposium on Frequency Control*, Baltimore, Maryland, pp. 76-81, 1990.

[7] M. Kihara and K. Hisadome, "Design of high performance, portable, optically pumped cesium beam standards", in *Digest of 1988 Conference on Precision Electromagnetic Measurements (CPEM '88)*, Tsukuba, Japan, pp. 248-249, 1988.

[8] K. Hisadome and M. Kihara, "Frequency locking of laser diodes using an optically pumped cesium beam tube", in *Proc. 44th Annual Symposium on Frequency Control*, Baltimore, Maryland, pp. 44-52, 1990.

[9] A. DeMarchi, "New insights into causes and cures of frequency instabilities (drift and long term noise) in cesium beam frequency standards", in *Proc. 41st Annual Frequency Control Symposium*, Philadelphia, Pennsylvania, pp. 53-58, 1987.

[10] G. Avila, V. Giordano, V. Candelier, E. de Clercq, G. Théobald, and P. Cérez, "State selection in a caesium beam by laser-diode optical pumping", *Phys. Rev. A*, 36, 8, pp.3719-3728, 1987.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

MINIATURE OPTICALLY PUMPED CESIUM STANDARDS

L. L. Lewis*
Ball Communications Systems Division
P.O. Box 1235
Broomfield CO 80038-1235

Abstract

We have designed, constructed, and tested two different types of optically pumped cesium frequency standards. One standard uses saturated absorption double resonance spectroscopy in a closed cesium cell. The second standard uses a more traditional optically pumped cesium beam, but with a very short microwave cavity, and optical pumping using the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=5 transition at 852 nm.

Introduction

The first laboratory studies of optical pumping of cesium using laser diodes were made as early as 1971 [1], and the first laboratory cesium beam frequency standard using laser diode optical pumping was constructed by 1980 [2]. Since that time, a number of groups [3,4,5] have sought to improve the accuracy of primary frequency standards by using optical pumping of cesium. Optically pumped cesium beams can have improved spatial symmetry, uniform population distributions, more easily predicted velocity distributions, and increased signal-to-noise ratios (SNR). It is possible to prepare the cesium atoms in a single magnetic sublevel. Cesium beams have been cooled by lasers to produce very slow atoms with narrow velocity distributions, which result in narrow microwave resonances with little uncertainty in the second order Doppler shift. All of these features can be used to improve the accuracy of a primary frequency standard.

On the other hand, other groups [6,7,8] have attempted to improve the characteristics of small, field usable frequency standards by means of optical pumping. The features of optical pumping which are desirable in a primary frequency standard are generally attained with large, elaborate, and environmentally sensitive laboratory instruments, which are impractical for use in a field standard. The advantages of optical pumping for small frequency standards stem mostly from the higher SNR [6]. In this case, the higher SNR obtained with optical pumping and optical state detection may allow a reduction in size or an improvement in performance of a small standard. Several workers [6,9,10] report the importance of low noise lasers in achieving a high SNR. The work reported here uses an external cavity diode laser in a folded Littrow configuration [11]. This laser has a (voltage) SNR on detection of the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=5 cycling transition in excess of $10^5/Hz^{-1/2}$. Such high SNR is critical to the success of miniature frequency standards.

In order to study the limitations of miniature cesium standards, we constructed several devices of different geometries. Two of these devices are of particular interest. In the first, which uses saturated absorption spectroscopy, an effective beam of atoms is selected within a closed cell. In the second apparatus, a very short microwave region is used with a more traditional cesium beam. In a beam tube which is carefully designed to remove all known frequency shifts, any unknown systematic effects might be difficult to observe. Our beam tube was fabricated with relatively low precision in the interest of time and cost, and in order to quickly identify any unanticipated frequency shifts associated with a small optically pumped device.

Saturated Absorption Double Resonance

Background

A traditional cesium beam frequency standard uses a vacuum envelope with cesium oven, graphite getters, and an ion pump. The standard would be greatly simplified and reduced in size if these elements could be eliminated, and a simple cell of cesium vapor used in-

*Work supported in part by U.S. Army Laboratory Command, Contract No. DAAL01-88-C-0805

CH2965-2/91/0000-521 $1.00 © 1991 IEEE

stead. This is done in a rubidium frequency standard, where the alkali vapor is contained within a glass cell. However, in order to avoid wall collisions and satisfy the Dicke condition, rubidium cells use buffer gases to confine the rubidium atoms to a small region of space. This also prolongs the interaction of the atoms with the microwave field, and (through the Dicke criterion) preserves the phase of the microwave-atom interaction. Unfortunately, the buffer gases also produce frequency shifts of the microwave transition [12]. A way to retain the advantages of a closed cell without the disadvantages of a buffer gas may be to use saturated absorption spectroscopy [13,14].

In a saturated absorption experiment, Fig. 1, one laser beam is split into two counterpropagating laser beams, which pass through a cesium cell. Wavelength dependent features appear in both the optical fluorescence and absorption spectra when a particular velocity class of atoms interacts with both lasers. This interaction occurs when the atomic velocity results in a positive Doppler shift of one laser's frequency to an upper energy level, while simultaneously Doppler shifting the second laser beam's frequency to a lower level. If the frequency separation between two optical transitions is $\Delta\nu$, then a saturated absorption feature appears when the component of atomic velocity along the laser beam direction is approximately

$$v = \lambda \frac{\Delta\nu}{2}$$

where λ is the laser wavelength. The laser is tuned to a frequency precisely midway between the two energy levels. Thus saturated absorption features appear on each optical transition, corresponding to zero atomic velocity along the laser beam. Additional features appear midway between any two allowed optical transitions, indicated by a "$\otimes$" symbol. The $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=3 $\otimes$ 5 transition, for example, requires an atomic velocity of about 194 m/s. This compares well with 200 m/s, the most probable velocity in a cesium cell at 50 deg C. For the cesium D2 transition, there are a total of six saturated absorption transitions associated with each ground state hyperfine level. Figure 2 shows a typical saturated absorption spectrum for optical transitions originating in the $^2S_{1/2}$,F=4 ground state. When one of the laser transitions is an optical pumping transition, the saturated absorption feature is enhanced, since very little laser power is required to remove an atom from interaction with the second laser.

When a microwave field is applied to the cesium atoms, the saturated absorption spectrum is further modified. By way of example, consider the $^2S_{1/2}$,F = 4 $\leftrightarrow$ $^2P_{3/2}$,F'=3 $\otimes$ 4 transition. An atom travelling along the laser beam axis with velocity component 86.5 m/s will be Doppler shifted into resonance for the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'= 3 transition for one laser beam, and the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'= 4 transition for the other beam. The saturated absorption feature is associated with an effective beam of atoms within the cell, travelling collinearly with the laser beams. Both of these optical transitions optically pump the atoms into the $^2S_{1/2}$,F=3 ground state hyperfine level. If the atoms now make a microwave transition from the $^2S_{1/2}$,F=3,m_F=0 sublevel to the $^2S_{1/2}$,F=4,m_F=0 sublevel, they are once again available to interact with the laser beams. Consequently, when a microwave transition occurs, the optical absorption increases. The absorption is observed as a microwave frequency dependent dip in the detected light at the photodetector (Figure 1), which may be used to control the frequency of a local oscillator. We refer to this method as Saturated Absorption Double Resonance Spectroscopy (SADR).

Microwave Measurements

The microwave resonator used for the SADR measurements consists of a copper cavity 5.1 cm ID x 2.4 cm long, coaxially loaded with a quartz cell 1.3 cm OD x .96 cm ID x 2.5 cm long. A pinch-off tube is brought outside of the copper resonator, and the bulk of the cesium metal is collected there. We obtain a clean microwave resonance in the cavity at 9.2 GHz, with a Q of 1500, as long as the deposited cesium is completely external to the microwave cavity. Fine tuning is possible by inserting an alumina rod parallel to the long axis of the microwave cavity, at a radius near the outside copper wall. This gives a tuning range in excess of 20 MHz at 9.2 GHz. However, a transparent layer of cesium on the quartz reduces the cavity Q to about 100. In this case, greater microwave power is required, but the fine tuning rod is not necessary.

A sketch is given in Figure 3 of the microwave source used for our measurements with the SADR cavity. A low phase noise microwave source was derived from a Hewlett-Packard 5061 cesium standard, exclusive of the cesium beam tube. Frequency modulation and frequency sweeping were obtained by substituting an external synthesizer for the internal 12.63177 MHz synthesizer of the 5061. Greater microwave power to drive the low Q cavity came from an external oscillator which was phase locked to the output of the 5061. We frequency modulated the external synthesizer at a 1 kHz rate, and used a lock-in amplifier to detect the resultant 1 kHz modulation of the SADR optical

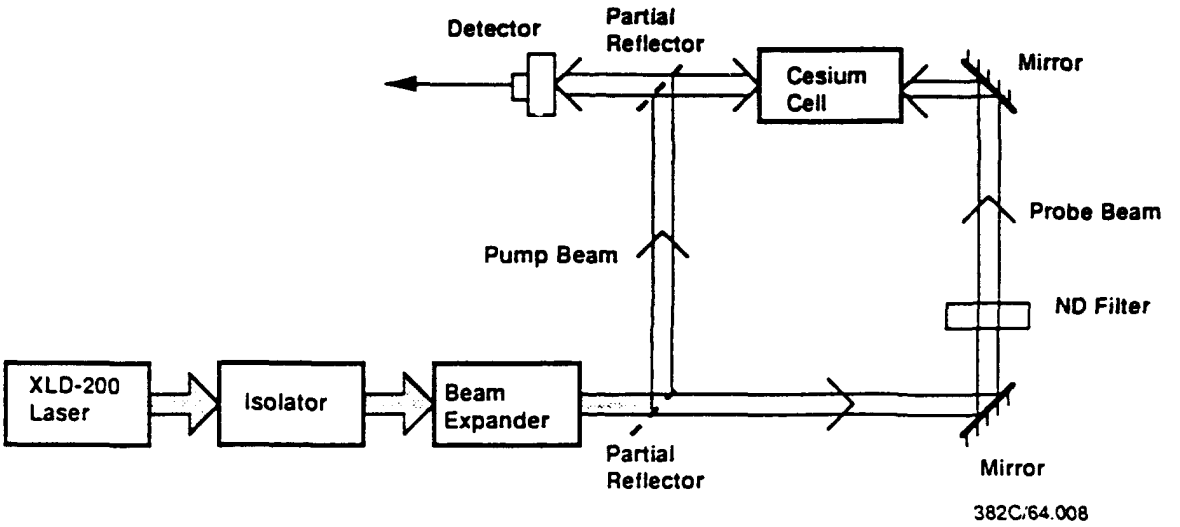


Figure 1: Saturated absorption experimental arrangement.

signal.

The laser source is sketched in Figure 4. Details of the XLD-200 laser are given in [11]. For the measurements reported here, a number of different laser beam cross sections were used. A typical laser beam cross-section was 10 mm x 6 mm, with 310 μW in the pump beam, and 33 μW in the probe beam. This gave a maximum detected signal of 5 μW. Smaller laser cross sections and powers were also used with success. An optical isolator was not used. The laser light was plane polarized. The pump power, probe power, microwave power and modulation index, laser frequency, and C-field current were all varied for these measurements.

Figure 5 gives a representative SADR microwave spectrum for the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=3 $\otimes$ 5 saturated absorption transition, with a small C-field applied, so that Figure 5 shows the $^2S_{1/2}$,F=3,m_F=0 $\leftrightarrow$ $^2S_{1/2}$,F',m_F=0 microwave transition only. The displayed curve is approximately the derivative of the microwave resonance, because of the use of a phase sensitive detector. For this particular set of parameters, the SNR is about 240 in a one Hz bandwidth. The microwave line width here is 14.3 kHz, while the predicted line width for a Rabi cavity of 2.4 cm length is 8.67 kHz. An accurate prediction of the line width for the SADR device should include the optical pumping process. It is likely that the true microwave interaction length is much shorter than 2.4 cm, since atoms are optically pumped throughout the laser path. The observed microwave line width would produce a short term stability of $\sigma_y = 2.6 \times 10^{-9}\tau^{-1/2}$ in a standard with a closed servo loop. The noise limitation here is laser noise. Figure 6 shows the Zeeman splitting of the SADR microwave transition for the same conditions as in Figure 5. Since the $\Delta m_F = \pm 1$ transitions are not evident in this spectrum, the magnetic field vector of the microwaves in the SADR cavity must be reasonably parallel to the C-field vector. Additionally, the observed symmetry of the various magnetic sublevel transitions is indicative of a lack of optical pumping among magnetic sublevels.

In order to further separate the parameters involved in the SADR frequency standard, we also locked the laser to saturated absorption features in a cell independent of the cell used for the microwave measurement. This changed the physics of the microwave resonance somewhat, as well, and allowed us to vary the laser intensity without affecting the laser frequency.

By locking the microwave source to the $^2S_{1/2}$, F=3,m_F = 0 $\leftrightarrow$ $^2S_{1/2}$,F',m_F = 0 cesium clock transition, we were able to measure the shift in the clock's frequency as a function of the laser frequency. We attribute this shift to the ac Stark shift (or "light shift"). It is characterized by dispersive features centered on each of the saturated absorption lines. It is one of the most important systematic effects of concern in an

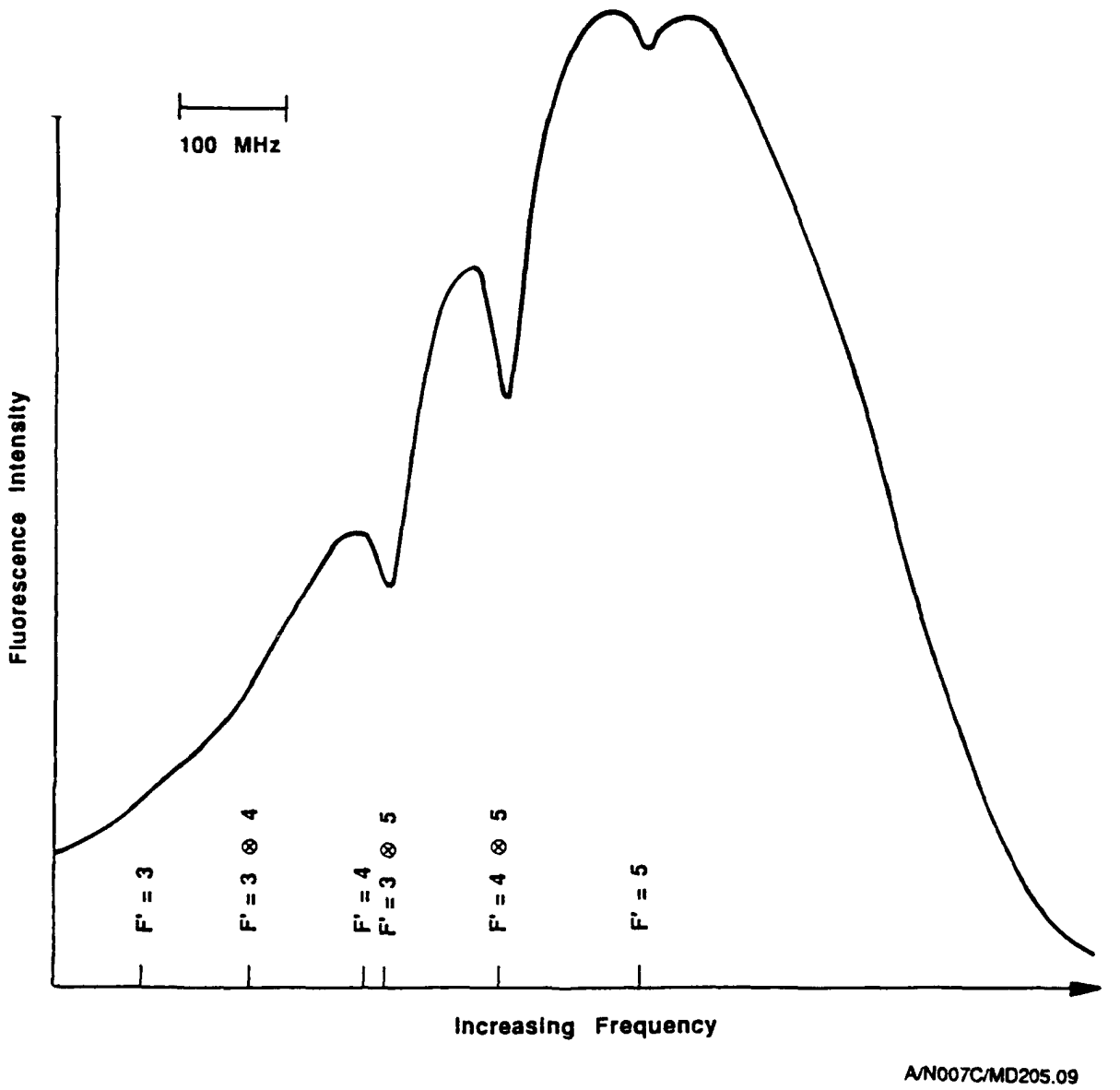


Figure 2: Saturated absorption spectra, F=4↔F′.

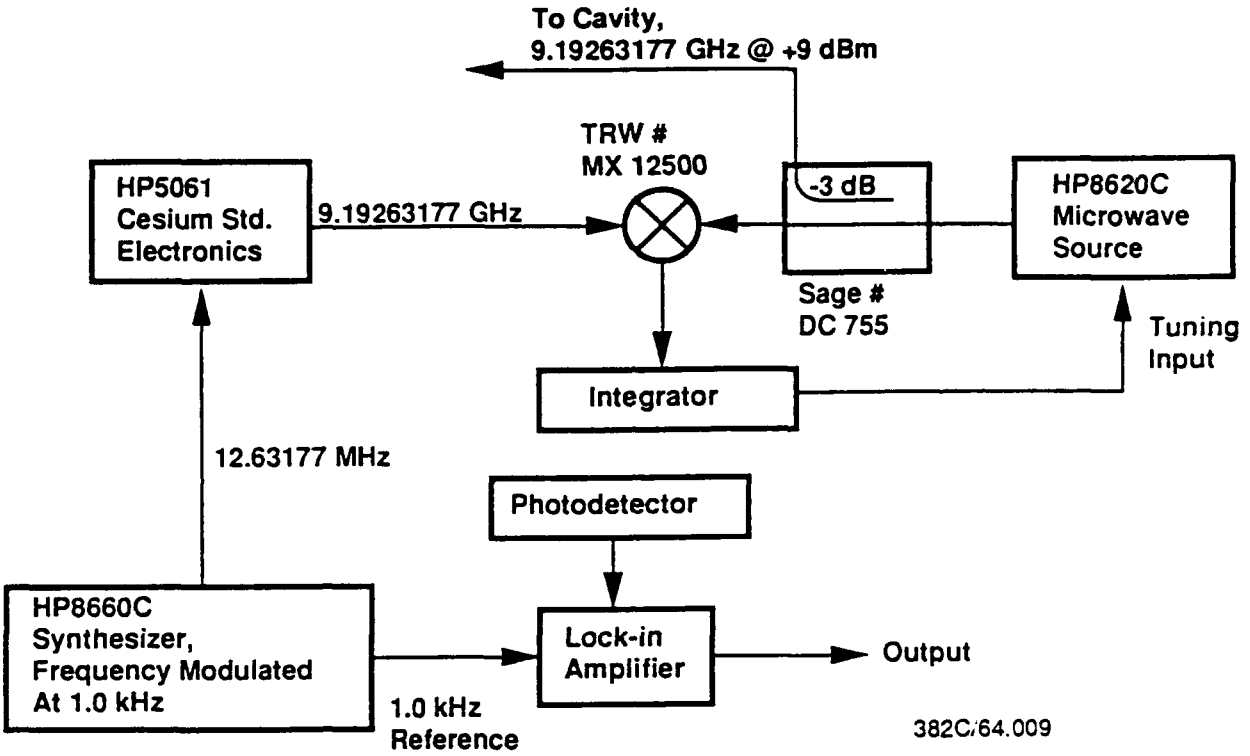


Figure 3: Microwave source for SADR.

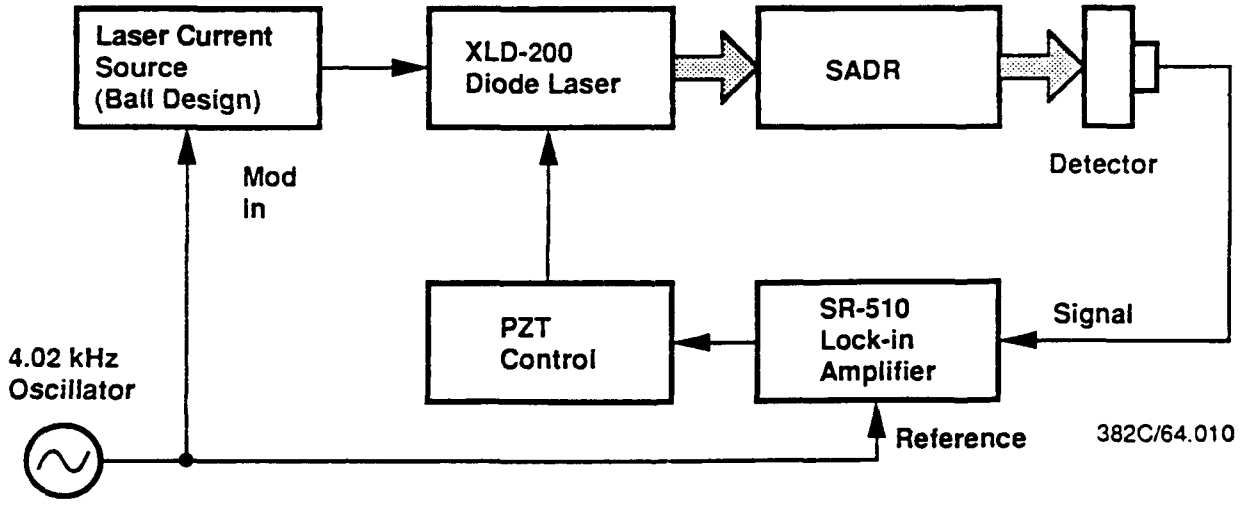


Figure 4: Laser source for SADR.

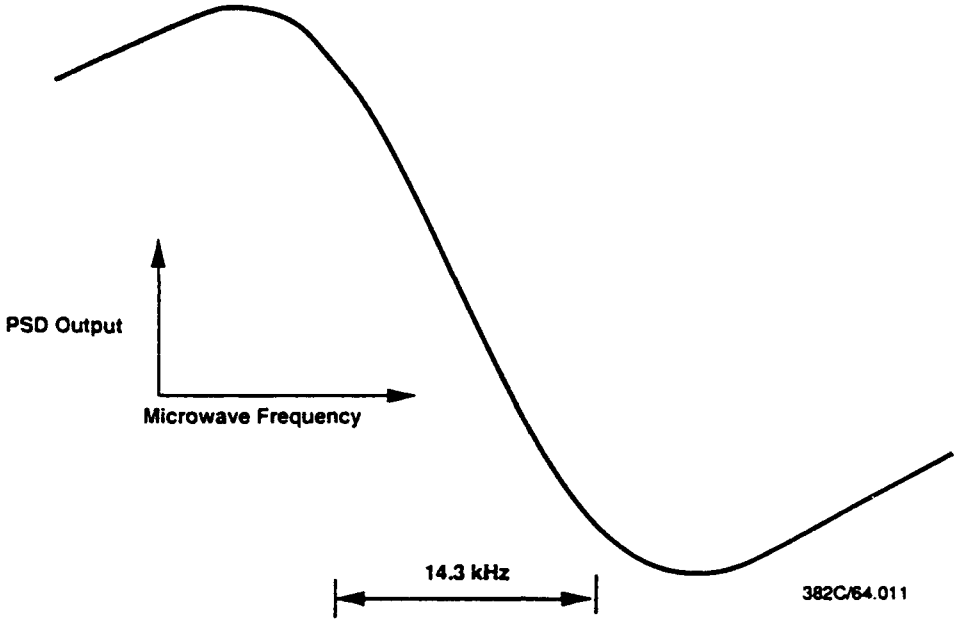


Figure 5: SADR microwave resonance.

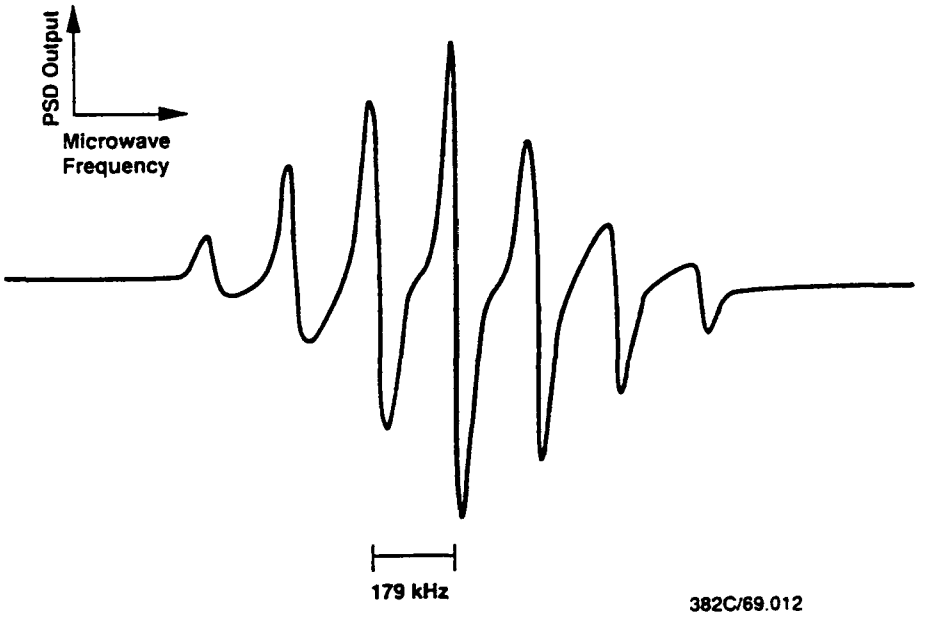


Figure 6: SADR microwave resonance.

optically pumped frequency standard.

For the light shift measurement, the expanded laser beam size was 6 mm x 10 mm, with 37 μW in the pump beam, and 14 μW in the probe beam. The detected laser power (off the optical resonance) was 3.0 μW. The microwave source was frequency modulated at a 1000 Hz rate, with a 5 kHz modulation depth. The microwave power delivered to the cavity was -1dBm. The relatively large microwave power used here is associated with the low Q microwave cavity, and the large microwave line width, which requires a high microwave power in order to reach a maximum signal.

The microwave light shift as a function of laser frequency is shown in Figure 7. The lower curve is the saturated absorption spectrum, showing all six lines originating in the $^2S_{1/2}$,F=4 ground state. The upper curve shows the locked microwave frequency as a function of the laser's optical frequency. The two curves are not exactly aligned with each other because the laser, which was swept in a free running mode, tended to drift during the measurements. The curves clearly show the dispersive peaks in the SADR frequency associated with the light shift at each of the saturated absorption lines. This curve is the clearest, most convincing evidence that we are indeed observing microwave resonances associated with a "beam" of cesium atoms within a closed cell. That is, we observe features in the microwave spectra which are associated with the sub-Doppler optical resonances. These features are due to the velocity class of cesium atoms selected by the saturated absorption process.

Unfortunately, the size of the light shift presents serious difficulties for this version of a SADR cesium frequency standard. The $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=3 $\otimes$ 4 transition, which probably has the best SNR of the six lines for microwave detection, has a relatively small light shift. This makes this transition a good candidate for use in the SADR frequency standard. The peak-to-peak frequency shift is 2.0 kHz, with a half-width-at-half maximum of about 4.4 MHz at the optical frequency. The combined laser power of the pump and the probe beams is about 85 μW/cm^2, which gives a light shift of 5.3 kHz/(MHz-mW/cm^2). At the power els used here, it would be necessary to stabilize the laser's optical frequency to about 200 Hz (out of 3.5×10^{14} Hz) in order to maintain a microwave fractional frequency stability of 1×10^{-11}. We believe that a laser stability of a few thousand Hz would be possible, considering the width of the saturated absorption feature, and the projected environment of this frequency standard. Therefore, we consider the 200 Hz stability requirement to be impractical for this application, in its present form. While we have concentrated our continuing work on the more traditional beam tube geometry as described below, we believe that the SADR technique still has merit. The sealed cell, with no buffer gas, greatly simplifies construction of a cesium frequency standard. Elimination of the vacuum pump, cesium oven, and graphite getters should improve reliability and lifetime of the standard. The method may be used in applications where the large light shift is tolerable. Alternatively, methods should be explored to reduce the light shift. For example, operation in a pulsed mode, where the microwave frequency is measured in the dark, may be possible.

Short Beam Tube

Apparatus

We also made measurements on an optically pumped cesium beam standard where the entire microwave cavity was contained within the vacuum system. A miniature microwave cavity was placed in the region between the pump laser beam and the detection laser beam of an optically pumped cesium beam of traditional design. The cavity used was a dielectrically loaded quasi-TE_{011} resonator, consisting of a copper cylinder 2.69 cm ID x 3.0 cm long. The dielectric was an alumina cylinder 1.91 cm OD x 1.28 cm ID x 3.0 cm long. The linewidth of the microwave cavity resonance at 9.2 GHz was 30 MHz, centered at a frequency 5 MHz higher than the exact cesium resonance frequency. Coupling to the cavity was achieved using magnetic loops in the sidewall.

For these measurements, the two laser beams were approximately 1 mm in diameter, with 314 μW upstream (pump laser), and 18.7 μW downstream (detection laser). The cesium beam diameter is also 1 mm in diameter. The cesium flux at the downstream detector was about 2×10^9 atoms/s, based on the fluorescence signal and predicted collection efficiency (32 percent) of the detector. This figure agrees with predictions based on the cesium oven geometry.

Microwave Spectra

In order to make a Δm_F=0 microwave transition, the rf magnetic field must be in the same direction as the static C-field. For a true TE_{011} mode microwave cavity, with the atomic beam on axis, this direction is along the longitudinal axis of the cavity. We expected the same to be true of the dielectrically loaded TE_{011} cavity used in these measurements. However, with a

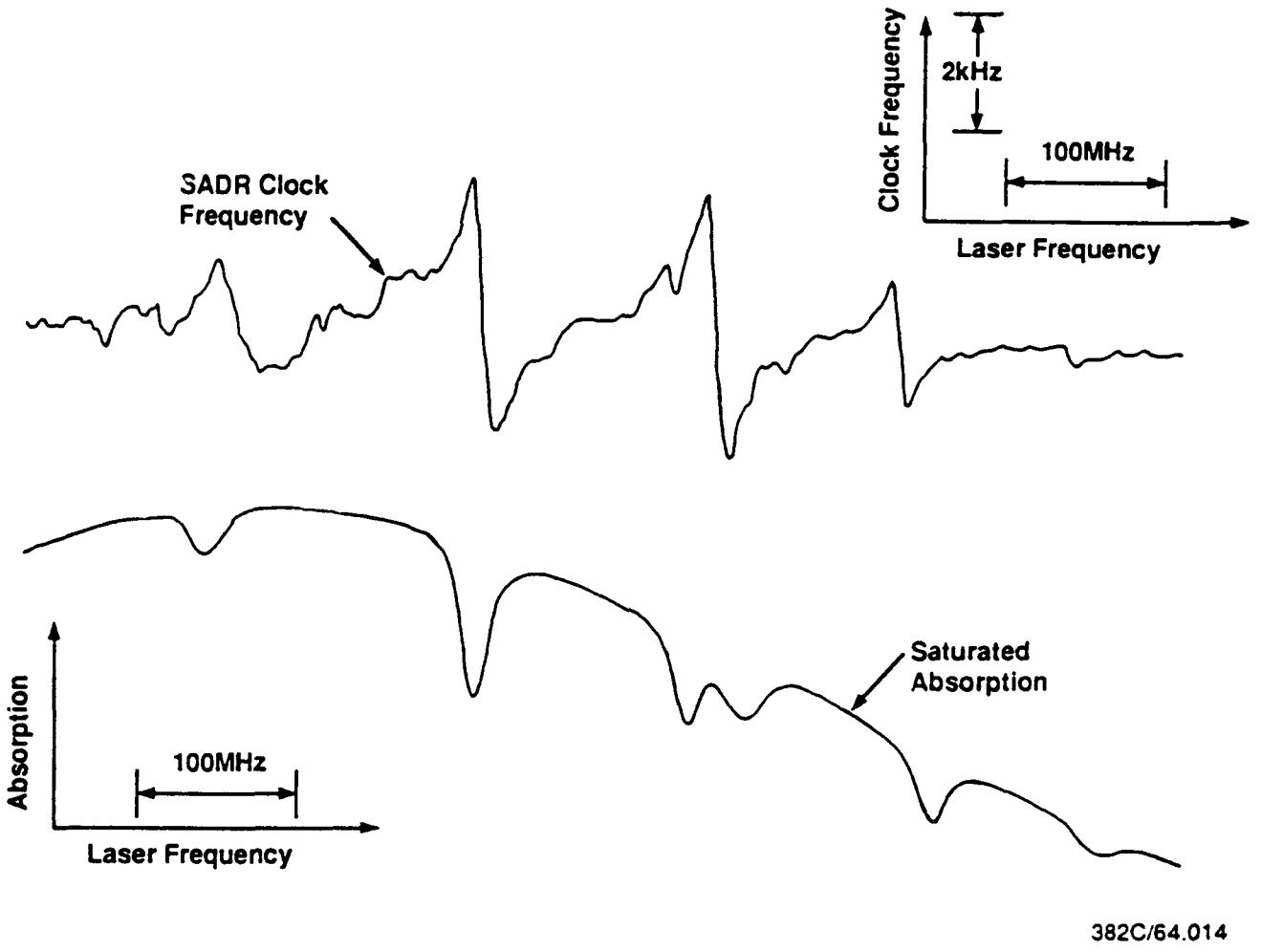


Figure 7: F=4→F′, showing the light shift of SADR.

moderately uniform longitudinal C-field we observed little or no $\Delta m_F = 0$ transitions. The $\Delta m_F = \pm 1$ transitions, on the other hand, were quite distinct. When an additional static field was applied in a transverse direction, we discovered that we could produce strong $\Delta m_F = 0$ transitions. Figure 8 shows the phase sensitive detector's output for the $^2S_{1/2}$,F=3,m_F=0 $\leftrightarrow$ $^2P_{3/2}$,F'=4, m_F=0 clock transition when a transverse field is applied. The same laser , locked to the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=4 optical transition, was used for both pumping and detection. Figure 9 shows the same spectrum using the output of the downstream detector directly, without phase sensitive detection. Figure 8 is just the derivative of Figure 9. Notable in these curves is the large asymmetry about line center, and the presence of two peaks. In fact, the shape of Figure 9, exclusive of the asymmetry, agrees well with an inverted Ramsey pattern, with a line width of about 11 kHz. This line width corresponds to a Ramsey interaction length of 1.4 cm, about half the cavity length used in this device. The maximum fluorescence observed on this microwave transition is about 16.6 percent of the maximum fluorescence observed downstream when the pump laser is removed upstream. This figure agrees well with values obtained in earlier experiments at NBS [6].

Figure 10 gives most of the Zeeman spectra for this device, taken under similar conditions to those in Figure 8. Both Δm_F=0 and $\Delta m_F = \pm 1$ transitions are clearly visible in this curve, indicating that the C-field is not parallel to the rf magnetic field, and that the C-field is moderately uniform (or else the higher frequency Zeeman transitions would broaden considerably more).

We make the following interpretation of these spectra. Due to variations in the wall thickness of the alumina dielectric in the microwave cavity, the effective center of the resonator is displaced from the geometric center of the cavity. Consequently, the atomic beam travels a line which crosses transverse rf magnetic fields at the entrance and exit regions of the cavity. The phase of the rf fields reverses in passing from the entrance to the exit. This produces the inverted Ramsey pattern which we observe. The effective length of this Ramsey cavity is less than the length of the whole cavity because of the extended nature of the microwave fields.

While unexpected, this result may actually be desirable. It may be useful to use this unusual Ramsey cavity in a small package which is simple to fabricate. By choosing materials and dimensions properly, the effect can be controlled and reliably reproduced. The technique can be used with a rectangular cavity, instead of the circular resonator used here. In the case of the rectangular cavity, the microwave magnetic field lines are aligned uniformly in a direction transverse to the atomic beam axis. This is an improvement over the cylindrical cavity used in this work, where the magnetic field direction changes along the atomic beam axis.

This interpretation of a Ramsey pattern does not explain the large asymmetry present in the microwave spectrum. Cutler [15] has analyzed the case in which the rf field direction is not held parallel to the C-field. He finds that large frequency shifts are possible under such conditions. This may eventually explain the large asymmetry present in Figure 9. At present, though, our experimental situation is too incompletely characterized to relate it to Cutler's theory.

The details of the microwave spectra are surprising. Very narrow features appear in the microwave spectra when the transverse and longitudinal C-fields are varied in magnitude and direction over the cavity length. An example of this phenomenon is shown in Figure 11. This curve has features which are only 1.5 kHz wide, which would correspond to a Ramsey interaction length of about 10 cm. Hashimoto [16] has described narrow features in rubidium frequency standards due to the modulation transfer effect. It seems unlikely that modulation transfer is responsible for the narrow features which we observe, since the optical fields are physically separated from the microwave fields. While our experimental arrangement in its present form is too variable to model theoretically, it is important to note that it is possible to generate very large and narrow features in the microwave spectra. Researchers constructing primary frequency standards are setting fractional frequency accuracy limits of less than 1×10^{-14}. The narrow features described above can produce frequency shifts as large as 1×10^{-7} in our apparatus. The laboratory primary frequency standards are generally 100 times longer, and the field uniformity is extremely good. Nonetheless, there is some question whether these features are rejected by factors of 10^7 or more.

Frequency Stability

We define a quantity S, which is a measure of short term stability:

$$S = \frac{1}{(2\pi)^{1/2}} \frac{\Delta\nu_\mu}{\nu_\mu(\mathrm{SNR})} \tau^{-1/2}$$

where $\Delta\nu_\mu$ is the atomic resonance linewidth, ν_μ is the microwave frequency, SNR is the (voltage) signal-

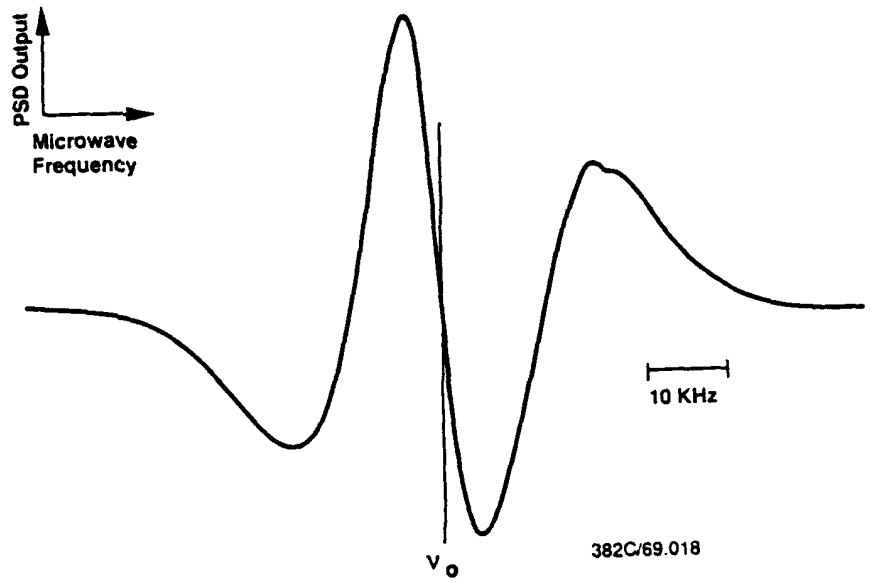


Figure 8: PSD output of miniature cesium clock transition.

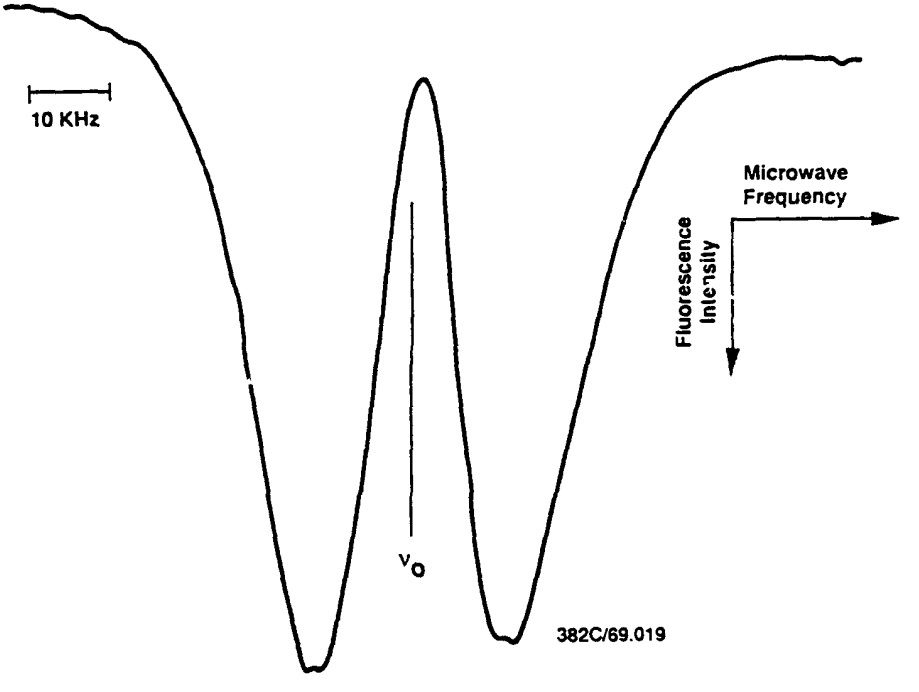


Figure 9: Detected fluorescence of miniature cesium clock transition.

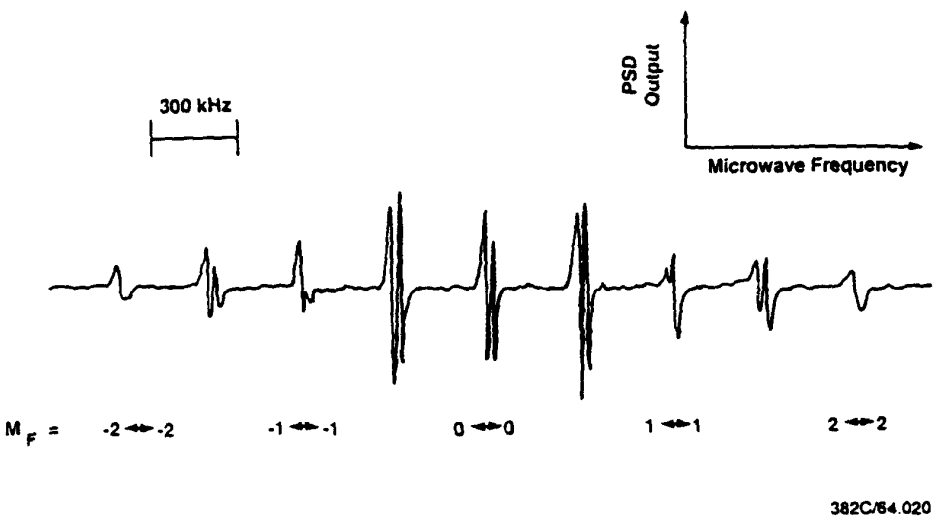


Figure 10: Zeeman spectra of miniature cesium.

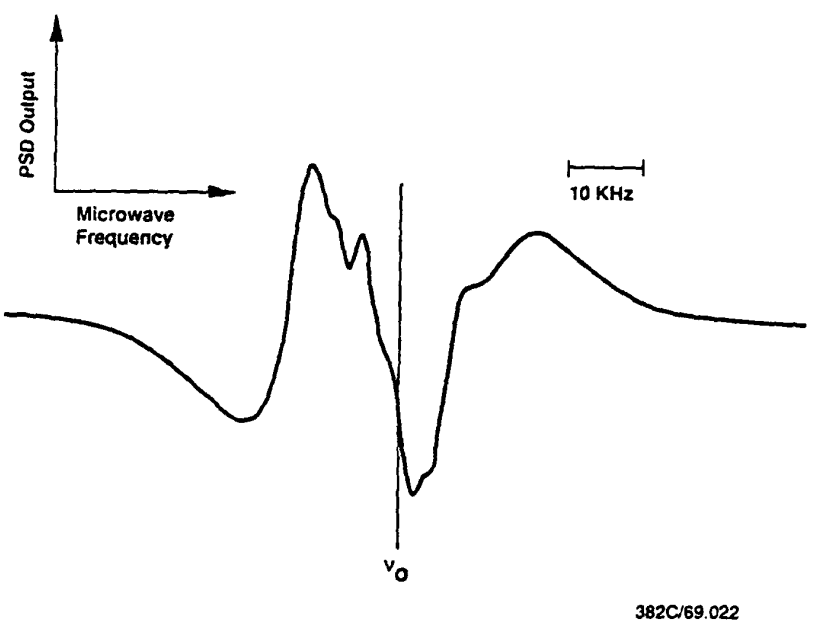


Figure 11: PSD output of miniature cesium clock transition with detailed structure.

to-noise ratio in a one Hz bandwidth, and τ is the measurement interval. This quantity approximately equals the short term stability of an atomic frequency standard [17].

In the case of our optically pumped cesium beam, with a short microwave cavity, typical measured values of $\Delta\nu_\mu$ are 5500 Hz, ν_μ is 9.2 GHz, and SNR is 624. This results in a value of S of $3.8 \times 10^{-10}\tau^{-1/2}$. The measured $\sigma_y(\tau)$ is closer to $5 \times 10^{-10}\tau^{-1/2}$.

These numbers were obtained for detection using the $^2S_{1/2}$,F=4 $\leftrightarrow$$^2P_{3/2}$, F'=4 transition. In this case, the SNR is limited by our detector noise. By switching to the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=5 cycling transition, the detection noise is limited by the cesium beam "shot" noise. Locking the laser to the downstream, $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=5 transition, we measure SNR=13,400, still limited by detector noise. Using this value in the expression for S, we obtain $\sigma_y(\tau)=1.8 \times 10^{-11}\tau^{-1/2}$. This is the short term stability to be expected if a separate laser is used to detect the downstream fluorescence. In addition, switching to a longer cavity configuration (such as the TE_{013} cavity [18]), increasing the oven temperature (a 20 deg C rise in temperature gives a factor of 4 increase in cesium flux, and a factor of 2 or more in SNR), and increasing the emitting area of the cesium oven (a factor of 4 larger area is practical) would improve the value of S by a factor of 12. This indicates that a short term stability of $1.5 \times 10^{-12}\tau^{-1/2}$ is possible in a miniature cesium standard. This gives us confidence in our ability to meet our goal of $1 \times 10^{-11}\tau^{-1/2}$.

Light Shift

Notwithstanding the large frequency shifts of unknown origin, we placed some limits on the size of the offsets due to light shift. We optically pumped the atoms with the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=5 optical transition, and measured the clock frequency. We then attenuated the pump laser power by a factor of 2.3, and repeated the frequency measurement. The measured frequency shift was $(-6.9 \pm 2)\times10^{-10}$. This implies a total light shift for the full power optical pumping of less than $(+12.1 \pm 3.5)\times10^{-10}$. It is interesting to note that for this measurement the total frequency shift associated with the asymmetric line is 800×10^{-10}. It is likely that the frequency shift observed when the pump laser is attenuated is also associated with a small change in the asymmetry of the microwave resonance, rather than a true light shift.

Even assuming that the measured shift is due to light shift, if one uses the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=4 optical pumping transition for state preparation, the effect is reduced. The smaller fluorescence from the latter tansition should produce less shift. Scaling the limit by the measured size of the fluorescence present when the $^2S_{1/2}$,F=4 $\leftrightarrow$ $^2P_{3/2}$,F'=4 transition is used for optical pumping, and accounting for the different light shift coeficients of the two optical transitions, the light shift limit is less than 2×10^{-10}. With a 10 percent stability of fluorescence intensity, the maximum instability due to light shift is therefore $\pm2 \times 10^{-11}$.

Based on theoretical calculations [19], we expect the light shift to be much smaller. However, a more sensitive measurement must await correction of the large asymmetry in the microwave resonance.

Alternative Pumping Method

As mentioned above, we did not always use the conventional optical pumping transitions [4]. A static magnetic field was applied along the direction of the pump laser beam, and the laser light was circularly polarized. In this manner, the atoms in the $^2S_{1/2}$,F=4 ground state hyperfine level were pumped into either the $^2S_{1/2}$,F=4,m_F=+4 or -4 sublevels, depending upon the sign of circular polarization. Atoms which make the $^2S_{1/2}$,F=3,m_F=0 $\leftrightarrow$ $^2S_{1/2}$,F'=4,m_F=0 microwave transition then contribute to increased fluorescence in the downstream detection region.

The advantage of this approach is that the same laser can be used for optical pumping upstream as is used for fluorescence detection on a cycling transition downstream, thereby helping to overcome the limitations of detector noise. One disadvantage is that atoms are removed from the $^2S_{1/2}$,F=4, m_F=0 sublevel, while the atoms in the $^2S_{1/2}$,F=3,m_F=0 sublevel remain unchanged by the optical pumping process. This is compared with the conventional pumping method, which increases the population difference by adding atoms to the $^2S_{1/2}$,F=3, m_F=0 sublevel, as well as depleting the $^2S_{1/2}$,F=4, m_F=0 sublevel. In addition, the atoms in the $^2S_{1/2}$,F=4, $m_F = \pm4$ sublevels can contribute background fluorescence in the detection region, which increases the noise appreciably. This latter difficulty may be ameliorated by applying a relatively large magnetic field in the detection region, which shifts the $^2S_{1/2}$,F=4, $m_F = \pm4$ sublevels out of the laser line width.

Acknowledgements

The author gratefully acknowledges helpful comments by T. English, S.R. Stein, P. Chantry, and A. DeMarchi.

References

[1] G. Singh, P. DiLavore, and C.O. Alley, "GaAs-Laser-Induced Population Inversion in the Ground-State Hyperfine Levels of Cs^{133}," *IEEE J. Quantum Electron.*,vol. QE-7, pp. 196-198, 1971.

[2] M. Arditi and J.-L. Picque, "A cesium beam atomic clock using laser optical pumping. Preliminary tests," *J. Physique (Paris)*, vol. L-41, pp. L-379-381, 1980.

[3] R.E. Drullinger, D.J. Glaze, J.L. Lowe, and J.H. Shirley, "The NIST Optically Pumped Cesium Frequency Standard," *IEEE Trans. Inst.and Meas.*, vol. 40, pp. 162-164, 1991.

[4] P. Cerez, G. Theobald, V. Giordano, N. Dimarcq, and M.deLabachelerie, "Laser Diode Optically Pumped Cesium Beam Frequency Standard Investigations at LHA," *IEEE Trans. Inst. and Meas.*, vol.40, pp. 137-141, 1991.

[5] S.-I. Ohshima, Y. Nakadan, Y. Koga, "Development of an Optically Pumped Cs Frequency Standard at the NRLM," *IEEE Trans. Instr. and Meas.*, vol. 37, pp. 409-413, 1988.

[6] A. Derbyshire, *et al.*,"Optically Pumped Small Cesium Beam Standards; A Status Report," in *Proc. 39th Annual Symposium on Frequency Control*, pp. 18-21, 1985.

[7] P. Thomann and F. Hadorn, "Short-term Stability of a Small Optically Pumped Cesium Resonator," presented at the 3rd European Frequency and Time Forum, Besancon, France, March, 1989.

[8] J.C. Camparo and R.P. Frueholz,"Fundamental stability limits for the diode laser pumped rubidium atomic frequency standard," *J. Appl. Phys.*, vol. 59, pp. 3313-3317, 1986.

[9] T. Maeno, H. Saito, J. Umezu, Y. Ohta, M. Kajita, R. Hayashi, and E. Morikawa, "Characteristics of a Laser Diode System for an Optically Pumped Cesium Atomic Clock," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 146-148, 1991.

[10] D.-H. Yang and Y.-Q. Wang,"Detection by cycling transition in an optically pumped cesium beam frequency standard," *Optics Comm.*, vol. 73, pp. 285-288, 1989.

[11] L.L. Lewis, "Low noise laser for optically pumped cesium standards," in *Proc. 43rd Annual Symposium on Frequency Control*, pp. 151-157, 1989.

[12] L.G. Bernier, A. Brisson, M. Tetu, J.Y. Savard, and J. Vanier, "Possible avenues of Improvement of the Short and Long Term Stability of Optically Pumped Passive Rubidium Frequency Standards," in *Proc. 33rd Annual Symposium on Frequency Control*, pp. 376-383, 1980.

[13] W.R. Bennett, Jr., "Hole Burning Effects in a He-Ne Optical Maser," *Phys. Rev.*, vol. 126, pp. 580-593, 1962.

[14] P.G. Pappas, M.M.Burns, D.D. Hinshelwood, M.S. Feld, and D.E. Murnick, "Saturation spectroscopy with laser optical pumping in atomic barium," *Phys.Rev.A*, vol. 21, pp. 1955-1968, 1980.

[15] L.S. Cutler, C.A. Flory, R.P. Giffard, and A. DeMarchi, "Frequency pulling by hyperfine σ transitions in cesium beam atomic frequency standards," *J. Appl. Phys.*, vol. 69, pp. 2780-2792, 1991.

[16] M.Hashimoto and M. Ohtsu, "Modulation transfer and optical Stark effect in a rubidium atomic clock pumped by a semiconductor laser," *J.Opt.Soc.America*, vol. B6, pp. 1777-1789, 1989.

[17] C. Audoin and J. Vanier, "Atomic frequency standards and clocks," *J. Phys. E*, vol. 9, pp. 697-720, Sept. 1976.

[18] A. DeMarchi, R.E. Drullinger, and J.H. Shirley, "Interference Fringes from Single-Cavity Excitation of an Atomic Beam," in *Proc. 44th Annual Symposium on Frequency Control*, pp. 34-38, 1990.

[19] J. Shirley, "Fluorescent Light Shift in Optically Pumped Cesium Standards," in *Proc. 39th Annual Symposium on Frequency Control*, pp. 22-23, 1985.

SINGLE ION OPTICAL FREQUENCY STANDARD*

J. C. Bergquist, Wayne M. Itano, D. J. Wineland, F. Diedrich†,
F. Elsner††, and M. G. Raizen

National Institute of Standards & Technology
325 Broadway
Boulder, CO 80303

Abstract

Experimental research at NIST toward the realization of an optical frequency standard of high accuracy is briefly reviewed. Our studies have concentrated on single, laser-cooled ions since they offer several attractive features toward the achievement of high accuracy (better than 1 part in 10^{17}). These features include long storage times which eliminate transit-time broadening, confinement that is nearly nonperturbative to the internal level structure of the ion, laser-cooling that reduces motional shifts to small values, and "electron-shelving" whereby transitions to long-lived states can be detected with unit probability. We have studied spectroscopically the electric-quadrupole allowed transition at 282 nm ($\sim 1 \times 10^{15}$ Hz) in a single, laser-cooled $^{199}Hg^+$ ion stored in an rf Paul trap with extremely high resolution. The measured linewidth is limited presently by the spectral purity of the laser to about 80 Hz. Possible improvements and future directions will be discussed.

Introduction

A single atom "at rest in space" has been promoted as an ideal system for spectroscopic measurements.[1] Certainly, a motionless atom that is free of any perturbations must be considered as an ideal reference for a frequency (and time) standard with high accuracy. A close approximation to this system is a single, laser-cooled ion in a miniature radio-frequency (rf) Paul trap. It appears realistic that a single, trapped ion could be free from perturbations due to collisions, Doppler shifts and electric or magnetic fields to the order of 10^{-18}.[1,2] However, a reference based on one atom is limited to a signal-to-noise ratio of unity (at best) for a single measurement cycle.[2-4] Therefore, the stability of a frequency standard is compromised by using single ions to achieve high accuracy. The loss of stability caused by using a single ion can be partially recovered by locking to a narrow transition with a resonance frequency that lies in the optical region of the spectrum. The stability of an oscillator locked to N atoms using the Ramsey method of interrogation (assuming 100% detection efficiency) is given by [5]

$$\sigma_y(\tau) = (\tau N T_R \omega_0^2)^{-1/2} \quad (\tau > T_R) \tag{1}$$

where τ is the averaging time, T_R is the Ramsey interrogation period and ω_0 is the transition frequency in radians per second. If the atomic transition is in the middle of the visible spectrum ($\omega_0/2\pi \simeq 5 \times 10^{14}$) and the Ramsey period is 1 s, the stability for a single atom could be as good as $3.2 \times 10^{-16}\tau^{-1/2}$. The potential high accuracy and stability of a single-ion optical frequency standard make it attractive as a clock. An important problem, however, is the difficulty in generating time from optical frequencies. The technology to connect microwave frequencies to visible frequencies has been demonstrated,[6] but it is cumbersome, laser intensive and narrowband. Schemes for broadband division from the optical region to the microwave region[7,8] and for simpler multiplication have been proposed and are being investigated.[8] These would help realize the full benefits of a single ion, optical frequency standard.

Experiment

Presently at NIST we are studying the $5d^{10}6s\ ^2S_{1/2}(F=0,m_F=0) \rightarrow 5d^96s^2\ ^2D_{5/2}\ (F=2,m_F=0)$ electric-quadrupole transition ($\omega_0/2\pi \simeq 1 \times 10^{15}$ Hz) in $^{199}Hg^+$ as an optical frequency standard.[2,5,9,10] There are other ions with credentials to compete as an optical frequency standard[1,11] and several groups throughout the international community are investigating some of these choices. A ^{199}Hg atom is ionized and trapped in the harmonic pseudopotential well created by an rf potential applied between the electrodes of a miniature Paul trap (endcap separation $2z_0 \simeq 660$ μm). The amplitude of the rf potential (at frequency $\Omega/2\pi \simeq 21$ MHz) could be varied up to 1.2 kV. The ion is laser-cooled to a few millikelvins by a few microwatts of radiation from two 194 nm sources.

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

The necessity of using two radiation sources to laser-cool $^{199}Hg^+$ is caused by its hyperfine structure. A mercury isotope with hyperfine structure is required in order to have a first-order field-independent clock transition near B = 0. The frequency of one source is tuned slightly below the resonance of the strongly allowed transition from the $5d^{10}6s\ ^2S_{1/2}(F=1)$ level to the $5d^{10}6p\ ^2P_{1/2}(F=0)$ level (see Fig. 1). Since transitions from the F=0 excited level to the F=0 ground level are forbidden, the $^2S_{1/2}(F=1)$ and the $^2P_{1/2}(F=0)$ levels nearly comprise a two-level system. However weak off-resonance excitation in the Lorentzian tail of the $^2S_{1/2}(F=1) \rightarrow {}^2P_{1/2}(F=1)$ transition causes pumping into the F=0 ground level through the F=1 excited level. The second 194 nm source is tuned to the $^2S_{1/2}(F=0)$ to $^2P_{1/2}(F=1)$ transition in order to optically pump the ion back to the F=1 hyperfine level in the ground state. In this way, the ion can be cooled to near the Doppler cooling limit of 1.7 mK by scattering photons at 194 nm. Transitions to other levels are detected by measuring the fluorescence count rate from the 194 nm photons that are scattered by the ion and collected in a solid angle of about $5 \times 10^{-3} \times 4\pi$ sr. For $^{199}Hg^+$, we achieve a peak count rate of about 25,000/s with a detector efficiency of 10%.

The narrow $^2S_{1/2}(F=0,m_F=0)$ - $^2D_{5/2}(F=2,m_F=0)$ clock transition is coherently driven by radiation at 282 nm obtained by frequency doubling the radiation from a cw dye laser that is stabilized to a high finesse cavity.[9,10] The frequency of the fundamental radiation at 563 nm could be offset from the cavity resonance and tuned through the ion resonance by means of an acousto-optic modulator. The 282 nm radiation and the 194 nm radiation are turned off and on sequentially in order to prevent broadening and shifts of the narrow S-D transition. Optical-optical double resonance[1,3,12] (electron shelving[1,13]) is used to detect each transition to the metastable D state as a function of the frequency of the 282 nm laser. At the beginning of each measurement, the ion is prepared in the F=0 hyperfine state in the ground level by blocking the 194 nm source tuned to the $^2S_{1/2}(F=0)$-$^2P_{1/2}$ (F=1) transition for a period of 5 ms. During this time the 194 nm source tuned to the $^2S_{1/2}(F=1)$ - $^2P_{1/2}$ (F=0) transition optically pumps the ions into the F=0 ground state. After this period, both sources at 194 nm are blocked and the 282 nm light is permitted to radiate the ion for a period that was varied up to 15 ms. The frequency of the 282 nm radiation was tuned to resonance or near resonance with the $^2S_{1/2}(F=0,m_F=0)$ - $^2D_{5/2}(F=2,m_F=0)$ transition. At the end of the probe period, the 282 nm radiation was turned off and both 194 nm sources were turned back on. If 194 nm fluorescence was detected no transition to the D state was recorded; if no fluorescence was detected, a transition to the D state was recorded. The data was digitized, 1 for fluorescence and 0 for no fluorescence, and then averaged with the previous results at this frequency. Then the frequency of the 282 nm radiation was stepped and the measurement cycle repeated. With the high fluorescence rate at 194 nm, it was possible to detect each transition with nearly no ambiguity in 10 ms.

Since there were long term drifts in the frequency of the narrow-band 282 nm radiation, we locked its frequency to the narrow S-D transition with an attack time of a few seconds.[10] To do this, we began each measurement cycle by stepping the 282 nm radiation to near the half power point on each side of the resonance N times (N varied from 8 to 32). At each step, we probed for 5 ms and then looked for any transition with the electron-shelving technique. We averaged the N results from each side of the resonance line, took the difference and corrected the frequency of the 282 nm laser. In this way, variations in the frequency of the 282 nm laser for time periods exceeding a few seconds were reduced.

In Fig. 1, we show a spectrum obtained of the resonance of the narrow S-D transition. For this figure 138 consecutive scans were made where the 282 nm probe period was 15 ms, and the step size was 15 Hz at 563 nm (30 Hz at 282 nm). The resonance shows a clearly resolved triplet with the linewidth of each component less than 40 Hz (< 80 Hz at 282 nm). We believe that this triplet structure is caused by Rabi power broadening. The 282 nm radiation is focussed on the ion to a spot size of about 25 μm; therefore, on resonance, a laser flux of fewer than 10^6 photons/s will saturate the transition. Below the data is a theoretical lineshape calculated for an ion at rest, for no broadening due to collisions or laser bandwidth, for a pulse length of 15 ns and for sufficient power to give a 3.5 π-pulse. Qualitatively the figures compare well. The fluctuations from measurement cycle to measurement cycle in the quantum-occupation number of the ion in the harmonic well of the trap cause variations in the transition probability of the ions. This, and the finite linewidth of the laser, likely cause the general broadening and weakening of the signal. We plan to study the lineshape and the effects of power broadening in more detail in future experiments.

We have discussed so far the possibility of improving the frequency stability in a single-ion frequency standard by increasing the frequency of the clock transition. From Eq. 1, we see that it is also possible to improve the stability of the frequency device as the square root of the number of trapped ions. However, for two or more laser-cooled ions in a quadrupole rf trap, the Coulomb force between ions repels them apart and away from trap center. If the ions do not reside at the center of the trap, they experience a force from the rf trapping potential whose magnitude increases with the distance away from trap center. This force causes motion of the ion at the trapping field frequency. Thus the ions are no longer at rest in space and the second-order Doppler shift due to the motion of ions at the rf frequency (termed "micromotion") grows

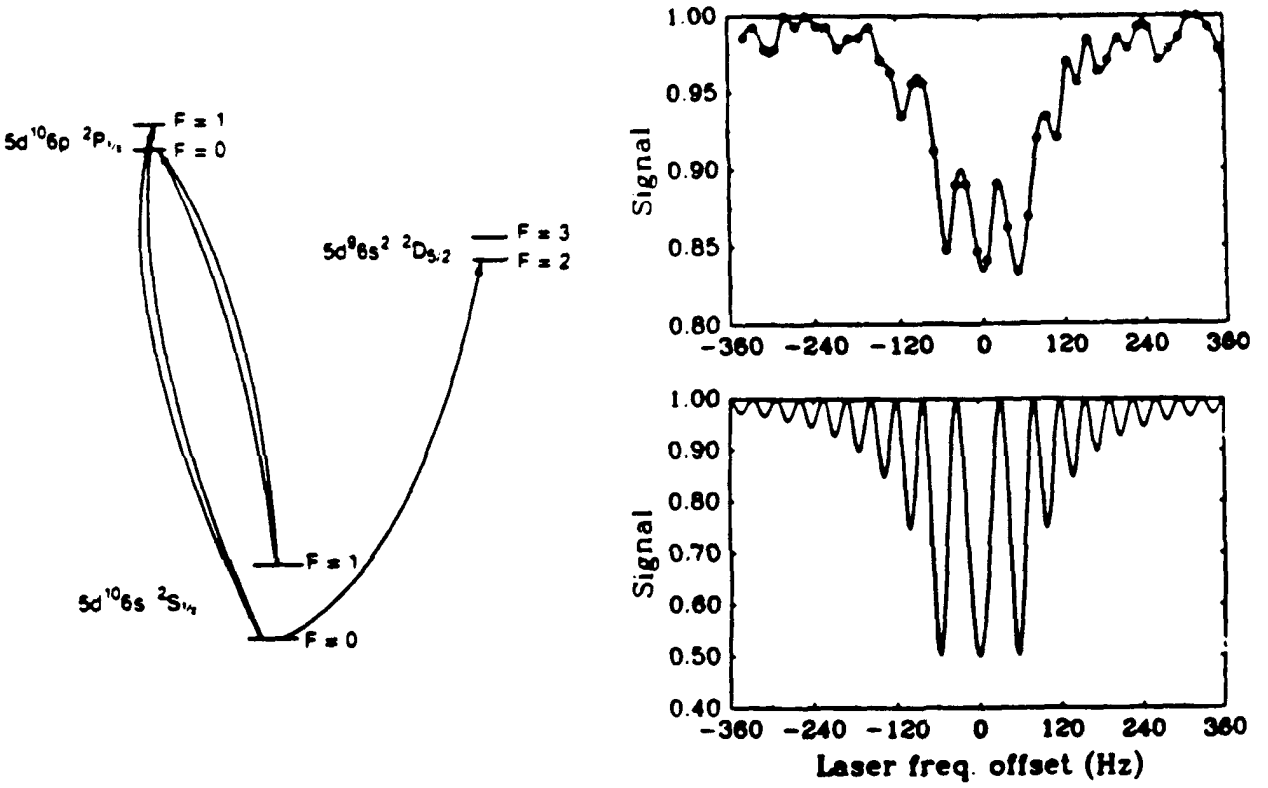


Figure 1. On the left is a simplified energy level diagram for $^{199}Hg^{+}$ at zero magnetic field. The quantized signal of the power-broadened lineshape obtained by scanning through the Doppler-free resonance of the $^{2}S_{1/2}(F=0,m_F=0)$ - $^{2}D_{5/2}(F=2,m_F=0)$ transition in a single laser-cooled $^{199}Hg^{+}$ ion is shown in the upper figure on the right. The frequency of a narrowband laser at 563 nm is doubled and long-term stabilized by locking to the S-D transition in the ion. The frequency of the laser is then stepped through the resonance in 15 Hz increments (30 Hz increments at 282 nm) for 138 consecutive sweeps. The lower-right figure shows the lineshape calculated for conditions similar to the experimental conditions for the upper figure, except that the ion is assumed to have zero temperature and the laser is assumed to have zero linewidth.

substantially larger than 2×10^{-18}.[2]

Another type of trap, the linear rf trap,[5,14-18] may allow the possibility to trap many ions while at the same time keeping the amplitude of their micromotion small. Ideally, the energy in the micromotion should be approximately equal to the energy in the laser-cooled secular motion of the ion. If so, then, even with many cold ions, it will be possible to achieve small second-order Doppler shifts ($< 2 \times 10^{-18}$). In fact, the fractional second-order Doppler shift should be somewhat less in a linear rf trap than in a quadrupole rf trap[5,16] because of the absence of rf micromotion along the symmetry axis of the linear trap. A schematic of a linear rf trap is shown in Fig. 2. An rf potential is applied to a pair of opposing rods while the other pair is held at ground potential. This produces a time averaged force which confines ions in the radial direction. An electrostatic potential applied to the end sections of the electrodes confines the ions along the trap axis. In a linear trap the rf fields approach 0 along a line (ideally along the trap axis) rather than only at a point as in the quadrupole rf trap.

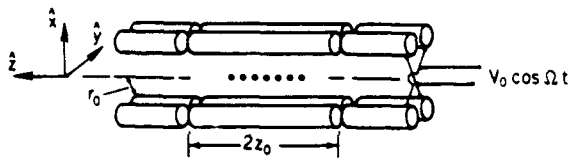


Figure 2. An example of a possible electrode configuration for a linear rf trap. In this configuration, if the length of the electrodes is much greater than the linear extent of the ions, the rf potential at the position of the ions has no z component. Thus, the energy in the micromotion of ions that are trapped along the axis of symmetry (the z-axis) can be on the order of the secular motion energy.

Use of a linear rf trap would allow us to store many laser-cooled ions ($N \simeq 50$), each nearly at rest in a benign environment and each acting as an independent clock.[5,14,15] Hence, the limitation to accuracy caused by second-order Doppler shifts in a frequency standard based on a linear array of trapped ions should be equivalent to a single ion frequency standard and the stability should be better by the square root of the number of stored ions. Thus, a linear trap frequency standard becomes attractive even for microwave frequencies. For example, from Eq. 1, for $T_R = 100$ s, $N = 50$ and $\omega_0/2\pi = 40.5$ GHz (the ground state hyperfine splitting in $^{199}Hg^+$) the frequency stability would be $\sigma_y(\tau) = 5.5 \times 10^{-14}\tau^{-1/2}$. We have begun our studies of both a microwave and optical frequency standard based on a linear trap with up to approximately 30 laser-cooled ions. We have observed crystallization of the cold ions along the trap axis, and observed the 40.5 GHz microwave transition with a fractional resolution of about 6×10^{-12}. Future possibilities include using the nearly motionless string of ions in tests of various principles in fundamental physics such as interference experiments and cavity QED.

We gratefully acknowledge the support of the Office of Naval Research and the Air Force Office of Scientific Research.

†Present address: Gsänger Optoelektronik GmbH, Robert-Koch-StraBe 1a, D-8033 Planegg 1, Germany
††Present address: Universität Hamburg, D-2000 Hamburg, Germany

References

[1] H. Dehmelt, "Coherent Spectroscopy on Single Atomic System at Rest in Free Space," in Proc. of the Second Frequency Standards and Metrology Symposium, 1976, pp. 357-361, and H. Dehmelt, "Coherent Spectroscopy on Single Atomic System at Rest in Free Space II," J. Phys. (Paris) Colloq., vol. 42, pp. C8-299-C8-305, Dec. 1981.

[2] D. J. Wineland, W. M. Itano, J. C. Bergquist, and R. S. Hulet, "Laser-cooling Limits and Single Ion Spectroscopy," Phys. Rev. A, vol. 36, pp. 2220-2232, Sept. 1987.

[3] J. C. Bergquist, W. M. Itano, and D. J. Wineland, "Recoilless Optical Absorption and Doppler Sidebands of a Single Trapped Ion," Phys. Rev. A, vol. 36, pp. 428-431, July 1987.

[4] D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, "Proposed Stored $^{201}Hg^+$ Ion Frequency Standards," in Proc. of the 35th Annual Frequency Control Symposium, 1981, pp. 602-611.

[5] D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano, D. J. Heinzen, S. L. Gilbert, C. H. Manney, and M. S. Raizen, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 37, pp. 515-523, Nov. 1990.

[6] C. R. Pollock, D. A. Jennings, F. R. Peterson, J. S. Wells, R. E. Drullinger, E. C. Beaty, and K. M. Evenson, "Direct Frequency Measurements of Transitions at 520 THz (576 nm) in Iodine and 260 THz (1.15 μm) in Neon," Opt. Lett., vol. 8, pp. 133-135, Mar 1983, and D. A. Jennings, C. R. Pollock, F. R. Peterson, R. E. Drullinger, K. M. Evenson, J. S. Wells, J. L. Hall, and H. P. Layer, "Direct Frequency Measurement of the I_2-stabilized He-Ne 473-THz (633-nm) Laser, Opt. Lett., vol. 8, 136-138.

[7] D. J. Wineland, "Laser-to-microwave Frequency Division using Synchrotron Radiation," J. Appl. Phys., vol. 50, pp. 2528-2532, April 1979.

[8] H. R Telle, D. Meschede, and T. W. Hänsch, "Realization of a New Concept for Visible Frequency Division; Phase Locking of Harmonic and Sum Frequencies," Opt. Lett., vol. 15, pp. 532-534, May 1990.

[9] J. C. Bergquist, F. Diedrich, W. M. Itano, and D. J. Wineland, "Hg^+ Single Ion Spectroscopy," in Proc. of the Ninth International Conference on Laser Spectroscopy, 1989, pp. 274-277.

[10] J. C. Bergquist, W. M. Itano, F. Elsner, M. G. Raizen and D. J. Wineland, "Single Ion Optical Spectroscopy," in Proc. of the International Workshop on Light Induced Kinetic Effects on Atoms, ions, and Molecules, to be published.

[11] See for example, D. J. Wineland, W. M. Itano, J. C. Bergquist, J. J. Bollinger, and H. Hemmati, "Frequency Standard Research Using Stored Ions," Prog. Quant. Electr., vol. 8, pp. 139-142, 1984, G. Werth, "Lifetime Measurements of Metastable States in Ions," in Frequency Standards and Metrology, ed by A. DeMarchi (Springer-Verlag, Berlin) 1989, pp. 293-299, R. Blatt, R. Casdorff, V. Enders, W. Neuhauser, and P. E. Toschek, "New Frequency Standards Based on Yb^+, in Frequency Standards and Metrology, ed. by A. DeMarchi (Springer Verlag, Berlin) 1989, pp. 306-311, G. P. Barwood, A. S. Bell, P. Gill, and H. A. Klein, "Trapped Yb^+ as a Potential Optical Frequency Standard," in Frequency Standards and Metrology, ed. by A. De Marchi (Springer-Verlag, Berlin) 1989, pp. 451-452.

[12] D. J. Wineland and W. M. Itano, "Spectroscopy of a Single Mg^+ Ion," Phys. Lett., vol. 82A, pp. 75-78, March 1981.

[13] H. Dehmelt, "Proposed $10^{14}\ \Delta\nu < \nu$ Laser Fluorescence Spectroscopy on $T\ell^+$ Mono-Ion Oscillator II," Bull. Am. Phys. Soc., vol. 20, pp. 60, 1975.

[14] H. Dehmelt, "Coherent Spectroscopy on a Single Atomic System at Rest in Free Space III," in Frequency Standards and Metrology, ed. by A. DeMarchi (Springer-Verlag, Berlin) 1989, pp. 15-19.

[15] M. G. Raizen, J. C. Bergquist, J. M. Gilligan, W. M. Itano, and D. J. Wineland, "Linear Trap for High Accuracy Spectroscopy of Stored Ions," submitted for publication in the Journal of Modern Optics.

[16] D. J. Wineland, "Spectroscopy of Stored Ions," in Precision Measurement and Fundamental Constants II, ed. by B. N. Taylor and W. D. Phillips, National Bureau of Standards (U.S.) Spec. Publ. 617, 1984, pp.83-92.

[17] D. A. Church, "Storage-Ring Ion Trap Derived from the Linear Quadrupole Radio-Frequency Mass Filter," J. Appl. Phys., vol. 40, pp. 3127-3134, July 1969.

[18] J. D. Prestage, G. R. Janik, G. J. Dick and L. Maleki, "Linear Ion Trap for Second-Order Doppler Shift REduction in Frequency STandard Applications," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 37, pp. 535-542, Nov. 1990.

[19] H. Walther and W. Quint, private communication.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INTENSITY AND FREQUENCY NOISE IN SEMICONDUCTOR LASERS

Kerry J. Vahala

California Institute of Technology

An overview of the physics governing frequency and intensity noise in semiconductor lasers will be discussed and the noise performance of state-of-the-art laser diodes will be reviewed.

Semiconductor lasers have now found their way into several large commercial markets. Brightness, diffraction limited spot size, power efficiency, reliability, and cost per component are the overriding concerns in most of these applications. A sole exception is their application to fiber optic telecommunication systems. Research and product-development activities in this area continue to set impressive device performance records concerning spectral purity, tunability, modulation speed, and relative intensity noise levels. As a result of this effort commercial semiconductor lasers are nearly ideal in terms of their physical properties. Their intensity noise spectra and short-term frequency stability are governed almost exclusively by quantum mechanical effects, and these, in turn, determine system performance levels. In this paper we will review the physics governing field fluctuations in semiconductor lasers, discuss the performance levels that have been achieved in state-of-the-art devices, and try to forecast future performance levels and novel structures/materials that may one-day be used in these devices.

CH2965-2/91/0000-539 $1.00 © 1991 IEEE

A SURVEY OF OPTICAL TECHNIQUES WITH AN EMPHASIS ON FREQUENCY CONTROL (PHASE STABILITY) ASPECTS

Harry T.M. Wang

Hughes Research Laboratories
3011 Malibu Canyon Road
Malibu, CA 90265

Abstract

An introductory survey of fiber optic techniques for signal transmission is given. Both intensity modulation/direct detection (IM/DD) and coherent approaches will be discussed. IM/DD is widely used in telecommunication and for RF/microwave signal transmission. Coherent technique has more stringent laser phase stability requirements but provides greater sensitivity and enhanced frequency selectivity. The system performances of both approaches are greatly enhanced by an inline low noise fiber optical amplifier. Mixing of coherent laser beams has been used to generate narrow linewidth microwave signals.

I. Introduction

Fiber optics has made astounding progress since the first proposal to fabricate ultra low loss dielectric optical waveguide was made in 1966.[1] The achievement of CW operation at 0.8 μm in semiconductor GaAlAs lasers at about the same time greatly facilitated development of the technology. The progress in device technology has been very impressive indeed. Fiber transmission attenuation has been reduced from decibels per meter to a fraction of a decibel per kilometer. Furthermore, because modal dispersion can be used to compensate for material dispersion, single mode silica fibers — with zero dispersion occurring at a wavelength (around 1550 nm) where the fiber attenuation (0.2 db/km) is also the lowest, the so-called dispersion shifted fiber—are readily available.[2] The combination of low attenuation and zero dispersion enables long distance repeater-free high frequency signal transmission. The other critical components of a fiber optic link, the laser transmitter and the photodiode detector, have made equally impressive gains. Progress in device technology resulted in compact and rugged components with multigigahertz frequency response capabilities, ushering in a new era of signal transmission. The demonstration of erbium-doped fiber optical amplifier operating in the 1550 nm wavelength region in 1987[3,4] added fuel to an already explosive developmental environment, providing additional exciting possibilities. The high gain, low noise figure, wide bandwidth and polarization insensitive inline fiber optical amplifier has effectively alleviated loss limitation in most systems.

In this paper, we give an introductry survey of the fiber optic technique. The intensity modulation/direct detection technique, the most widely used approach at present, will be discussed in the next section. Applications in addition to deployment in telecommunication systems include reference frequency distribution, antenna remoting and signal feed for wideband microwave phased array antennae. A brief discussion of the erbium-doped fiber optical amplifier and its impact on long distance signal transmission will also be given, followed by a discussion on coherent technique and its requirements on laser linewidth and frequency stability. We will also discuss recent developments in narrow linewidth lasers, laser frequency stabilization, and generation of microwave signals using coherent mixing of two laser beams.

II. Intensity Modulation/Direct Detection Technique

The ability to use high frequency current to modulate the optical output of a laser diode and to recover the information using a photodiode detector makes direct laser modulation a very simple and convenient approach to a fiber optic link. It is thus not very surprising that today most of the deployed fiber optic systems employ the laser intensity modulation and direct detection technique. The approach is limited by (1) laser intensity noise, and (2) the modulation response of the laser. In addition to nonlinearities that introduce intermodulation products, direct modulation of the laser also broadens its emission spectral linewidth, which enhances the dispersion effect.

Laser intensity noise is the dominant source of noise in a short fiber optic link. It arises from photon number and carrier density fluctuations and is defined as the ratio of the mean square amplitude of these fluctuations to the square of the average intensity level, or

$$RIN = \langle \delta P^2 \rangle / P_o^2 \qquad (1)$$

The spectral characteristics of relative intensity noise (RIN) is an important consideration in RF and microwave signal transmission applications. RIN has a resonant peak at the laser relaxation frequency ω_r given by:[5]

$$\omega_r = \sqrt{((I / I_{th} - 1) / \tau_s \tau_p)} \qquad (2)$$

CH2965-2/91/0000-540 $1.00 © 1991 IEEE

where I/I_{th} is the laser injection current normalized to its threshold value, and τ_s and τ_p are the carrier and the photon lifetimes with values of the order of 3 ns and 1 ps, respectively. Thus, peak RIN occurs at a frequency of a few gigahertz, limiting the high frequency operation of the laser. As can be seen from Eq. (2), the laser modulation bandwidth can be increased by operating the laser at higher output levels.[6] More recent devices have operated at a modulation frequency up to 24 GHz.[7] At the low frequency end, the measured intensity noise is also enhanced because of coupling between longitudinal modes.[8] By beating with the modulation frequency, this noise could be translated into the signal band, causing signal to noise degradation.

The process of modulating the RF signal onto the optical carrier could involve significant losses. In addition to the optical coupling losses from the laser to the fiber and from the fiber to the detector, the modulation process could also encounter losses from impedance mismatch. The RF-to-optical-to-RF-link insertion loss could be as high as 50 dB; typically between 10 to 30 dB. A systematic optimization of the circuit model of a directly modulated fiber optic link using a laser with a RIN of -136 dB/Hz shows an insertion loss of 13.6 dB, a noise figure of 33 dB, and an intermodulation free dynamic range of 91 dB.[9] For greater modulation bandwidth (up to 100 GHz)[10] an external modulator could be used. This is typically a Mach-Zehnder interferometer where the differential phase shifts produced by the RF signal in the two arms of the interferometer are combined to give an intensity modulated signal. Since in this case the laser is operating CW, there is no modulation broadening of the laser emission spectra. A low noise high power output laser, such as a diode-laser-pumped Nd:YAG laser, could be used to provide a low noise link. Incorporating linearization and compensation designs, the external modulator could also provide a greater distortion free dynamic range. The main drawback of the external modulator is greater system complexity.

A few examples of applications of the IM/DD technique with time and frequency control implications are (1) Maser reference frequency distribution,[11] which takes advantage of the low coefficient of delay of the silica fiber. Stable reference signal distribution over multi-kilometer distance has been demonstrated by using the bidirectional transmission capability of the fiber and a phase conjugate delay compensation technique. (2) Optical steering of a dual-band microwave phased array antenna[12] has been demonstrated. The approach takes advantage of the fact that transmission through a fiber optic feed provides true time delay. The programmed phase of the signal at the radiating element is thus independent of frequency, enabling multiband antenna operation. (3) A fiber optic radar noise test set at X-band[13] has demonstrated superior measurement capability. Antenna remoting, telecommunication, and fiber optic sensors are other well-established applications of the IM/DD technique.

III. Erbium-Doped Fiber Optical Amplifier

An amplifier is a critical component of a signal transmission system. The ability to fabricate an optical amplifier has existed since the demonstration of the very first laser. After all, the acronym LASER stands for Light Amplification by Stimulated Emission of Radiation. By proper fabrication techniques to control the facet reflectivity, we can make a semiconductor laser operate below the threshold for lasing as an amplifier. Although semiconductor optical amplifiers have several desirable characteristics, such as simple current injection pumping, and can be designed to operate at any wavelength, the coupling losses reduce the effective amplifier gain and, more importantly, the gain is sensitive to the signal state of polarization. This latter factor is a distinct drawback for semiconductor optical amplifiers, which has placed severe constraints on applications. Thus, even though significant progress has been made in reducing polarization sensitivity, semiconductor optical amplifiers have not created the same kind of excitement as the newer erbium-doped fiber amplifier (EDFA), which operates only in the 1550 nm region.

An erbium-doped fiber can be pumped at any one of several different wavelengths to obtain optical amplification at 1550 nm.[14] The main interest is centered on 980 and 1480 nm sources because of the availability of high power semiconductor lasers at these wavelengths. The most attractive features of an EDFA are: (1) high gain (45 dB), (2) high bandwidth (40 nm) operation, (3) nonsensitivity to signal polarization, and (4) the amplifier's capability of having a near-quantum-limit noise figure (3 dB). The device is simple and is easily spliced into the fiber transmission line, thus alleviating the need for the cumbersome and lossy optoelectronic repeaters. Since the first demonstration of the EDFA device in 1987, development has been rapid and impressive. Indeed, transmission of a 2.5 GB/s signal over 2200 km of fiber using an EDFA chain has been demonstrated.[15] The feasibility of transmitting a 5 GB/s signal over 9000 km, and 2.5 GB/s signal over 21000 km, using EDFAs in a recirculating loop configuration has also been demonstrated.[16] As a pre-amplifier, an EDFA will greatly improve the sensitivity of a photoreceiver. At high bit rates, the sensitivity of an IM/DD receiver with an EDFA pre-amplifier is comparable to that for an existing, much more complex, coherent optical receiver.[17]

IV. Coherent Technique

A coherent optical communication system offers greater frequency selectivity and receiver sensitivity. The former effectively increases the capacity of the system by enabling greater utilization of the optical bandwidth for information transfer. The latter is achieved using the high power local oscillator laser to suppress the detector noise, analogous to that in the electronic heterodyne receiver. Sensitivity improvement depends on the modulation scheme and the receiver design.[18] To obtain these desirable benefits, the technical challenges are reduction of laser phase noise, for both the transmitter and local oscillator lasers, and maintaining polarization matching between the received signal and the local oscillator beam. In the latter case, polarization-maintaining fiber, as well as other polarization determining devices, have been employed.[19] Polarization diversity receivers in which the received signal is processed in two different channels corresponding to two polarization states have also been investigated.[20] The processed signals could then be recombined to recover the transmitted information.

Laser phase noise is a major technical challenge in a coherent system. A measure of phase noise is the laser emission linewidth, and it is generally recognized that narrow linewidth lasers are required for coherent communication. The requirement also depends on the information transfer rate, the

modulation technique, and error tolerance.[21] Thus, for a given bit error rate, the greatest receiver sensitivity is achieved using a homodyne phase shift keying technique. The required laser linewidth has been calculated to be 3.1×10^{-4} times the information bit rate. The corresponding ratio for a frequency shift keying approach is in a more modest range of 7.2×10^{-3} to 9×10^{-2}. For a 100 MB/s system, the required laser linewidths are 31, 720 to 9000 kHz, respectively. Until recently, such linewidth requirements were considered to be difficult to attain in semiconductor lasers (although laboratory gas and dye lasers have shown much narrower linewidths). The progress in device technology is such that within a short time span we have seen: (1) a diode-laser-pumped Nd:YAG ring laser with a linewidth of 3 kHz;[22] (2) a distributed feedback multi-quantum-well laser using corrugation pitch modulated grating to obtain 170 kHz linewidth;[23] (3) a distributed Bragg reflector MQW laser with a long cavity to obtain 85 kHz linewidth;[24] and (4) an erbium-doped fiber ring laser with a linewidth of 60 kHz.[25] Other techniques, such as external cavity and coupled cavity lasers, have shown similarly impressive results but are considered to be too susceptible to perturbations for system deployment.

Another aspect of the laser phase noise is the stability of the laser emission frequency. Semiconductor lasers exhibit typical temperature sensitivity of about 10 GHz/°K and injection current sensitivity of about 0.1 GHz/mA. These parameters are utilized with a frequency reference for electronic stabilization of the laser frequency. For short terms and less critical requirements, a Fabry Perot resonator provides a simple frequency discriminator. For longer terms and more critical requirements, atomic and molecular transitions have been employed as references.[26]

V. Optical Generation of Microwaves

Mixing of two laser beams can be used to generate microwave signals.[27] The upper frequency of the signal is limited by the detector response, which has been shown to be as high as 100 GHz. The linewidth and stability of the signal depend on the spectral characteristics of the mixing beams. If the mixing beams are coherent, very narrow linewidth signals can be generated. An example is the generation of a 35 GHz signal with a measurement limited linewidth of 10 Hz.[28] The experiment utilizes a 5.8-GHz frequency modulated master laser. The upper and lower third harmonic sidebands are used to optical-injection-lock two slave lasers, whose outputs are then coherent with each other (and with the master laser) and separated by the 35 GHz frequency of the signal to be generated. An alternate and more versatile way to obtain phase coherence between two frequency offset laser beams is by optical mixing and employing electronic phase lock loop techniques.[29] In this case, narrow linewidth lasers, such as diode-laser-pumped Nd:YAG lasers, are essential. The low frequency phase fluctuations can be effectively tracked and corrected by electronic feedback in one of the lasers. A 30 GHz beat signal with a linewidth of 1 mHz has been demonstrated. Using a tunable frequency reference, one can obtain coherent laser beams at variable offset frequencies.

VI. Conclusion

In this brief survey of fiber optic techniques we have indicated some exciting developments in the field. Progress in device technology is occurring at a brisk pace. The technique is expected to have a revolutionary impact on signal transmission.

References

[1] K.C. Kao and G.A. Hockham, "Dielectric Surface Waveguide for Optical Frequencies," Proc. IEE 133, 1151-1158, 1966.

[2] L.B. Jeunhomme, Single Mode Fiber Optics, 2nd Ed. (Dekker, N.Y., 1990).

[3] R.J. Mears, L. Reekie, I.M. Jauncey, and D.N. Payne, "Low Noise Erbium Doped Fiber Amplifier Operating at 1.54 μm," Electron. Lett. 23, 388, 1987.

[4] E. Desurvire, J.R. Simpson, and P.C. Becker, "High Gain Erbium-Doped Traveling Wave Fiber Amplifier," Opt. Lett. 12, 888, 1987.

[5] T. Ikegami and Y. Suematsu, "Resonance-Like Characteristics of the Direct Modulation of a Junction Laser," Proc. IEEE 55, 122, 1967.

[6] W. Ng, R. Craig, and H.W. Yen, "Dynamic Characteristics of High Speed p-Substrate GaInAsP Buried Crescent Laser," IEEE J. Lightwave Tech., LT-7(3), 560, 1989.

[7] E. Meland et al., "Extremely High Frequency (24 GHz) INGaAsP Diode Lasers with Excellent Modulation Efficiency," Electron. Lett. 26, 1827, 1990.

[8] C.B. Su, J. Schlafer, and R.B. Lauer, "Explanation of Low Frequencyt Relative Intensity Noise in Semiconductor Lasers," Appl. Phys, Lett. 57(9), 849, 1990.

[9] C.H. Cox III, G.E. Betts, and L.M. Johnson, "An Analytic and Experimental Comparison of Direct and External Modulation in Analog Fiber Optic Links," IEEE Trans. MTT-38(5), 501, 1990.

[10] J. Nees, S. Williamson, and G. Mouron, "100 GHz Traveling Wave Electro-optic Phase Modulator," Appl. Phys. Lett. 54(20), 1962, 1989.

[11] L. Primas et al., "Distribution of Ultra-Stable Reference Frequency Signals Over Fiber Optic Cable," IEEE MTT-S Digest, 241-244, 1990.

[12] W. Ng, A. Walston, G. Tangonan, J. Lee, and I. Newber, "Optical Steering of Dual Band Microwave Phased Array Antenna Using Semiconductor Laser Switching," Electron. Lett. 26(12), 791, 1990.

[13] I. Newberg, C. Gee, G. Thurmond, and H. Yen, "Radar Measurement Applications of Fiber Optic Links," Proc. 42nd Ann. Fre. Contr. Symp., 453-455, 1988.

[14] R.I. Laming, M.C. Farries, P.R. Morkel, L. Reekie, and D.N. Payne, "Efficient Pump Wavelengths of Erbium Doped Fiber Optical Amplifier," Electron. Lett. 25, 12, 1989.

[15] S. Saito, T. Imai, and T. Ito, "An Over 2200-km Coherent Transmission Experiment at 2.5 Gb/s Using Erbium Doped Fiber In-Line Amplifiers," J. Lightwave Tech. LT-9(2), 161, 1991.

[16] N.S. Bergano et al., "A 9000 km 5 Gb/s and 21000 km 2.4 Gb/s Feasibility Demonstration of Transoceanic EDFA Systems Using a Circulating Loop," Optic Fiber Conf. '91, PD13 (San Diego, CA, 18-27 Feb, 1991).

[17] K. Nagagawa and S. Shimada, "Optical Amplifiers in Future Optical Communication Systems," IEEE LCS 57, Nov. 1990.

[18] T. Ososhi and K. Kikuchi, "Heterodyne Type Optical Fiber Communication," J. Opt. Comm. 2(3), 82, 1981.

[19] T. Okoshi, "Recent Advances in Coherent Optical Fiber Communication Systems," J. Lightwave Tech. LT-5(1), 44, 1987.

[20] L.G. Kazovsky, "Phase- and Polarization Diversity Coherent Optical Techniques," J. Lightwave Tech. LT-7(2), 279, 1989.

[21] L.G. Kazovsky, "Performance Analysis and Lase Linewidth Requirements for Optical PSK Heterodyne Communication Systems," J. Lightwave Tech. LT-4(4), 415, 1986.

[22] R.L. Byer, "Diode Laser Pumped Solid State Lasers," Science 239, 742, 1988.

[23] M. Okai, T. Tsuchiya, K. Uomi, N. Chinone, and T. Harada, "Corrugation-Pitch-Modulated MQW DFB Laser with Narrow Spectral Linewidth (170 kHz)," IEEE Photo. Tech. Lett. 2(8), 529, 1990.

[24] T. Kuni, Y. Matsui, H. Horikawa, T. Kamijoh, and T. Nonaka, "Narrow Linewidth (85 kHz) Operation in a Long Cavity 1.5 μm MQW DBR Laser," Electron. Lett. 27(9), 691, 1991.

[25] P.R. Morkel, G.J. Cowle, and D.N. Payne, "Traveling Wave Erbium Fiber Ring Laser with 60 kHz Linewidth," Electron. Lett. 26(10), 632, 1991.

[26] M. Ohtsu, M. Hashimoto, and H. Ozawa, "A Highly Stabilized Semiconductor Laser and its Application to Optically Pumped Rb Atomic Clock," Proc. 39th Ann. Symp. on Freq. Contr., pp. 43-53, 1985; S.G. Menocal et al., "Frequency Locking of 1.5 μm DFB Laser to a Neon Indicator Lamp using the O_1 togalvanic Effect," IEEE Phot. Tech. Lett. 1(10), 285, 1990; S. Sudo et al., "Frequency Stabilized DFB Laser Module using 1.53159 μm Absorption Line of C_2H_2, IEEE Photo. Tech. Lett. 1(10), 281, 1990.

[27] R.G. Hunsperger and M.A. Mentzer, "Optical Control of Microwave Devices," Proc. SPIE 993, 204-224, 1988.

[28] L. Goldberg et al., "FM Sideband Injection Locking of Laser Diodes," Electron. Lett. 18, 1019, 1982.

[29] K.J. Williams et al., "6-34 GHz Offset Phase Locking of Nd:YAG 1319 nm NonPlanar Ring Lasers," Electron. Lett. 25, 1242, 1989.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY PULLING IN CESIUM BEAM FREQUENCY STANDARDS DUE TO DELTA M=±1 (SIGMA) TRANSITIONS

L. S. Cutler, C. A. Flory, and R. P. Giffard
Hewlett-Packard Laboratories
P. O. Box 10350
Palo Alto, California 94303-0867
and
A. De Marchi
Politecnico di Torino
Torino, Italy

Abstract: Measurements have shown that cesium tubes with appreciable $\Delta M=\pm 1$ (sigma) transitions have significant frequency departures from the Breit-Rabi value as a function of C-field. The behavior is quite different from the usual pulling by $\Delta M=0$ (pi) transitions (Rabi pulling). A detailed theoretical analysis of the pulling by both sigma and pi transitions using a six level model has been carried out and some of the results are presented here. The most important result is that the contributions to the clock transition probability from the closest sigma transitions have the Ramsey shape but with a shifted center when the state populations are asymmetric. Thus the pulling by these sigma transitions, which we call Ramsey pulling, cannot be reduced by third harmonic detection or other background slope removal techniques that are effective in Rabi pulling. Ramsey pulling can only be reduced by decreasing the sigma transitions and the population asymmetries. Optical pumping could be arranged that would have clear advantage with regard to population asymmetries.

Introduction

Recent measurements [1] on cesium atomic beam tubes known to have appreciable $\Delta M=\pm 1$ (sigma) transitions have shown significant oscillatory departures from the expected Breit-Rabi value as a function of C-field. The pulling behavior in these tubes is quite different from the usual Rabi pulling which is caused by the residual background slope due to incomplete cancellation of the skirts of the neighboring $\Delta M=0$ (pi) transitions. (In this paper and in reference [2] we have used pi and sigma to designate parallel and perpendicular, respectively, orientation of the microwave magnetic field with respect to the C-field. This is in contrast to the usual optical definition in which pi and sigma mean that the optical electric field is parallel and perpendicular, respectively, to the static magnetic field.) Rabi pulling can be greatly reduced by techniques that can detect the residual background slope such as third harmonic detection of sinusoidal modulation. In contrast, the pulling in tubes with significant sigma transitions cannot be reduced greatly by background slope detection techniques. Also, the period of the oscillatory departure versus C-field in these tubes is about twice what is expected from Rabi pulling, but this is not surprising in itself if the bulk of the effect is due to the sigma transitions since they are only half as far away.

It was suspected that the significant microwave sigma transitions might differ in behavior from the pi transitions since the sigma transitions adjacent to the clock transition always involve at least one of the clock states (F=3, M=0 or F=4, M=0) while the neighboring pi transitions are between levels of the same M value and do not involve clock states. Therefore the pi transitions can be treated to very good approximation as completely independent. A detailed theoretical analysis was performed [2] and this suspicion was confirmed. The analysis will not be described in detail here but some of the results will be presented as well as some experimental verification of the theory. We call the pulling due to sigma transitions Ramsey pulling.

It is now clear that these sigma transitions are detrimental to the performance of cesium beam and other types of frequency standards and should be minimized if highest accuracy and best long term stability are to be achieved.

CH2965-2/91/0000-544 $1.00 © 1991 IEEE

Physical System

The six significant levels in the hyperfine structure of cesium that are involved in the pulling study are shown in Figure 1. The three pi transitions (in which M does not change) are shown in Figure 1a. They are excited by the component of the microwave magnetic field that is parallel to the C-field. The central one, M=0 to M=0, is the desired clock transition. The other two are the main transitions adjacent to the clock transition in the microwave spectrum. They are essentially independent of the clock transition since they involve neither of the M=0 clock states. Any unbalance in the skirts of the these pi transitions at the center of the clock transition leads to a background slope that is responsible for the Rabi pulling.

The four important microwave sigma transitions in which M changes by ±1 are shown in Figure 1b. They are excited only if there is a component of the microwave magnetic field perpendicular to the C-field. In beam tubes designed for frequency standards, this component is normally kept as small as possible and the transition intensity is usually small compared to the pi transitions. The frequency spacing of these transitions from the central clock transition is half that of the pi transitions. It is evident from Figure 1b that the sigma transitions always involve one of the clock states. The sigma transitions are responsible for the Ramsey pulling.

Calculation Technique

Since the three pi transitions involve separate pairs of states, they may be treated independently. The technique for the complete approximate solution to the problem is then fairly straightforward but tedious. The six level system is solved exactly for the three pi transitions and the sigma transitions are then considered as small perturbations. Second order time dependent perturbation theory is then used to solve for their effects using the exact solutions for the pi transitions as basis states. The calculations must be done in each of the interaction regions of the beam tube where the microwave fields are present as well as in the drift space between them. In addition, the velocity distributions in the beam and the deflection characteristics of the magnets for the various states and atomic velocities need to be taken into account in the case of magnetic deflection beam tubes. The final result is an analytic expression for the current at the detector. This is a function of the microwave frequency and amplitude as well as all the tube parameters including the C-field which determines the Zeeman frequency.

Once the current has been calculated, its dependence on frequency departure from the center of the unperturbed clock transition is examined for various values of the parameters such as Zeeman frequency. The point of zero slope is considered to be the perturbed or pulled clock frequency. This is where a servo with small, slow modulation would lock. While this is not fully equivalent to the usual modulation techniques, it can't be far in error. Analytic expressions have been obtained for the pulling. Only the contributions to the current with odd symmetry in frequency departure from the line center contribute to frequency pulling; the even symmetric part only changes the line shape slightly and does not affect its symmetry or cause pulling.

The analytic expressions obtained by the second order perturbation treatment were checked by exact numerical solution of the six level problem for some specific cases. The agreement is excellent leading to confidence in the analytic expressions.

Results

The first calculations were done for a single sigma transition to show the nature and size of the effect. Figure 2 shows the central part of the pure Ramsey clock transition, F=3, M=0 - F=4, M=0. The tube parameters used are: L (drift length) = 16.7 cm, l (interaction length) = 1.0 cm, Zeeman frequency = 49 KHz, triangular velocity distribution with Vmin = 1x10E4 cm/sec, Vmax = 2x10E4 cm/sec, and Vpeak = 1.3x10E4 cm/sec. The peak of the transition probability is just slightly less than one.

Figure 3 shows the even part of the contribution to the clock transition from the F=3, M=1 - F=4, M=0 sigma transition. The angle between the C-field and the microwave field was chosen to be 5 degrees and all other parameters are the same as for Figure 2. Note that the even part only changes the line width slightly and does not

affect the location of the line center. It looks very similar to a Ramsey pattern with 180 degrees phase difference between the two excitation regions. The effect is small since the contributions to the transition probability are only a few parts in 10E5. The solid curve is the perturbation calculation and the dashed curve is the exact numeric calculation of the six level system.

Figure 4 shows the odd part of the contribution to the clock transition from the same F=3, M=1 - F=4, M=0 sigma transition. All conditions are the same as for Figure 3. This contribution looks similar to a Ramsey pattern with 90 degrees phase difference between the two excitation regions. It causes a shift in the point of zero slope of the clock transition of about 1x10E-13 and is the origin of the Ramsey pulling. The magnitude and sign of the shift vary strongly with zeeman frequency and velocity distribution. The dashed curve again is the result of the exact numerical calculation.

Not shown is a similar calculation for the contribution due to a pi transition such as F=3, M=1 - F=4, M=1. At the clock transition the contribution would be just the skirt of the pi transition since it is independent. The even part at the clock transition corresponding to Figure 3 would be approximately a constant value with essentially no structure. The odd part corresponding to Figure 4 would have approximately a constant slope with essentially no structure. This is the origin of the Rabi pulling. The slope can be measured or compensated for by any of several techniques; for example, sinusoidal modulation with third harmonic detection essentially ignores the background slope and greatly reduces the Rabi pulling.

Note that if the transition intensity for the M=+1 - M=+1 transition is equal to that for the M=-1 - M=-1 transition, the Rabi pulling essentially vanishes since the slopes of the skirts cancel at the clock transition except for a small difference in the Zeeman frequencies due to the second order C-field correction. The situation is more complicated for Ramsey pulling where the pulling will essentially vanish if the initial state populations at the entrance to the microwave region and the final state populations at the detector are symmetric. These conditions can be met well with certain types of optical pumping so that very small Rabi and Ramsey pulling can be expected in properly designed optically pumped standards.

Calculations of the Rabi pulling for a model of an actual double beam tube are shown in Figure 5. The microwave structure is the same as for Figures 2, 3, and 4. The velocity distributions for all the states were calculated by a very accurate ray tracing program. The maximum pulling is about 1.6E-12.

Figure 6 shows the calculated Ramsey pulling for the same tube. The maximum value is about 3 times larger than the maximum Rabi pulling. Note that the period of the pulling is about twice that of the Rabi pulling as might be expected since the sigma transitions are only half as far away in frequency as the pi transitions.

The calculated total combined pulling is shown in Figure 7. The maximum value is about 7x10E-12 and is primarily Ramsey pulling. Both Rabi and Ramsey pulling depend on the microwave power. The calculations were done with the theoretical power set to maximize the dc beam current at the center of the clock transition, the same way that is usually used in actual standards.

Figure 8 shows the measured pulling (experimental points and dashed line) of a double beam tube of the type modeled in Figures 5, 6, and 7 along with the calculated total pulling (solid line) of Figure 7. The calculations and measurements agree fairly well at low Zeeman frequencies. The more rapid damping of the calculated pulling at higher Zeeman frequencies is probably due to the velocity distributions in the actual tube being narrower than those calculated. The calculations assumed ideal magnetic deflection fields which are not achieved in the tubes and this could account for the discrepancy.

It is clear that Ramsey pulling can degrade the accuracy and long term stability of standards and, unlike Rabi pulling, cannot be reduced by slope detection techniques. Therefore it is important to make the C-field and microwave magnetic field as parallel as possible and to reduce state population asymmetries. Figure 9 shows the magnetic field distribution in one end of a conventional implementation of a Ramsey cavity. The C-field is in the z direction and the beam goes through the

cavity in the y direction at the cavity end. It is evident that the microwave magnetic field is parallel to the C-field only on the cavity symmetry plane. The tubes studied here have double beams which are placed symmetrically on each side of the cavity symmetry plane where they see a definite component of microwave field orthogonal to the C-field. Even tubes with a single beam cannot have perfectly parallel fields because of the non-zero size of the beam but they can be much better than the double beam tubes considered here. To get the best results, tubes with a single beam should have the cavity structure aligned so that the centroid of the beam passes through the symmetry planes.

Figures 10 and 11 respectively show the full microwave spectrum of a double beam tube of the type discussed here showing significant sigma transitions, and a new single beam tube having greatly reduced sigma transitions. The new tube also exhibits better symmetry of the pi transitions adjacent to the clock transition. As a consequence of these improvements, this new tube should show considerable improvement with regards to both Rabi and Ramsey pulling.

Summary

Results of a theoretical investigation of the frequency pulling by sigma transitions and some comparisons with an actual cesium dual-beam tube have been presented. It was shown that the sigma transitions adjacent to the clock transition always involve one of the clock levels so their frequency pulling (Ramsey pulling) differs fundamentally from the pulling by pi transitions (Rabi pulling). Rabi pulling is caused by the residual slope due to unbalanced pi transitions and can be significantly reduced by background slope detection techniques. Ramsey pulling results from a fundamental shift in the frequency of the clock transition rather than a background slope. It can only be decreased by: 1) reducing the sigma transitions which are excited when the microwave magnetic field and the C-field are not parallel or 2) making all the state populations prior to microwave interaction and detection symmetric about M=0. A new single-beam tube has been designed that has greatly reduced sigma transitions compared with the dual beam tubes. Properly applied optical pumping could give excellent results.

Acknowledgements

We thank Lou Mueller, Doug Dull, and Jack Kusters of the Santa Clara Division of Hewlett-Packard for their interest in our work and assistance in design and construction of the new experimental tube.

References

[1] A. De Marchi, in Proceedings of the 41st Annual Symposium on Frequency Control 1987, (IEEE, New York, 1987),p. 53.

[2] L. S. Cutler, C. A. Flory, R. P. Giffard, and A. De Marchi, "Frequency pulling by hyperfine sigma transitions in cesium beam atomic frequency standards," J. Appl. Phys. 69 (5), pp. 2780-2792, March 1991.

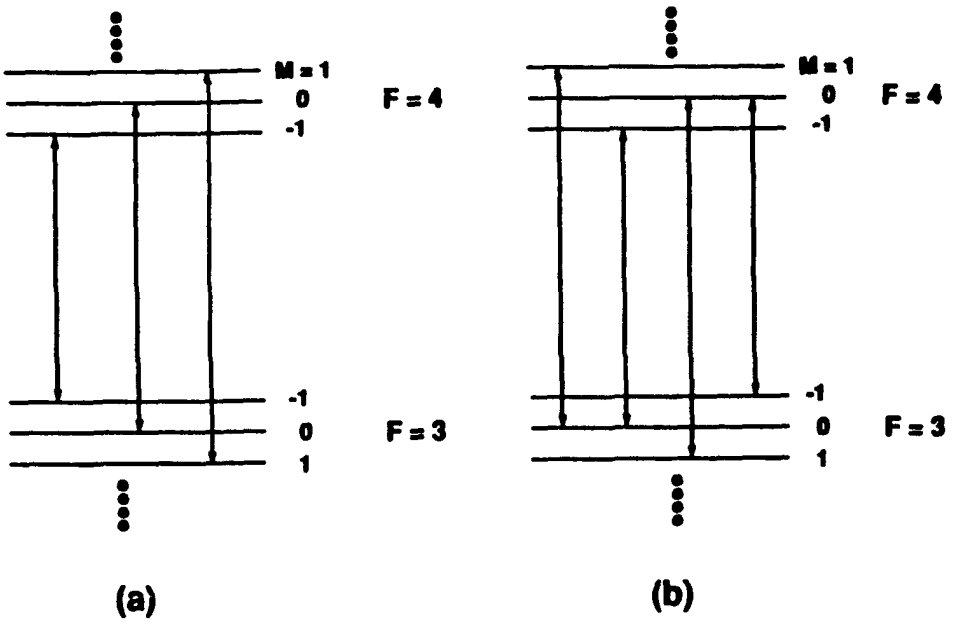


Figure 1. (a) Allowed Delta M=0 (pi) transitions for the subset of six hyperfine levels in cesium 133. (b) Allowed Delta M=±1 (sigma) transitions for the same level subset.

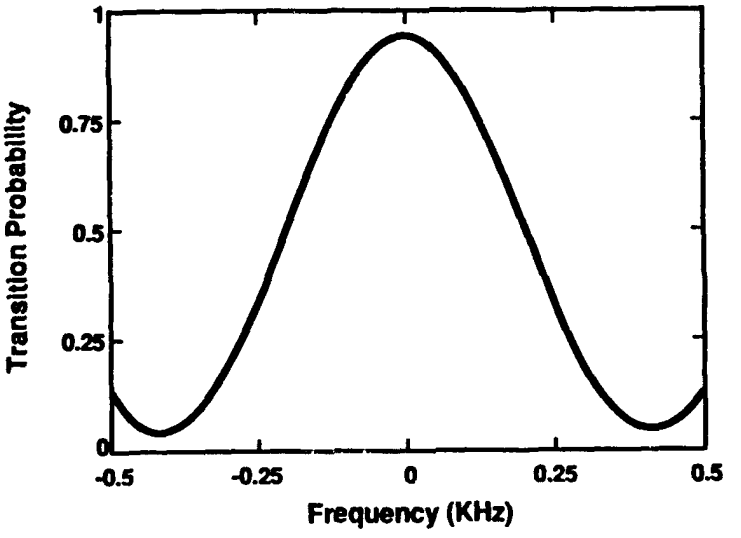


Figure 2. Calculated transition probability for the (3,0) - (4,0) clock transition as a function of microwave frequency difference from the center of the line. Assumed tube parameters are given in the text.

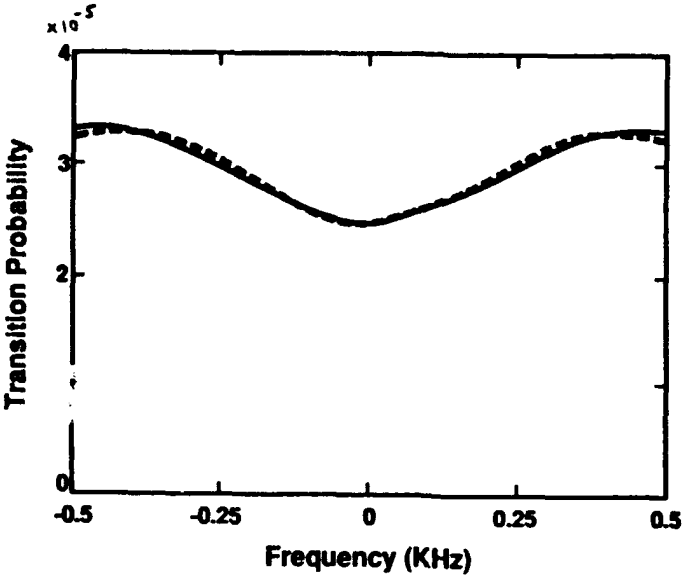


Figure 3. Even part of the calculated contribution to the clock transition probability from the (3,1) - (4,0) sigma transition. The solid curve is the approximate analytic result and the dashed curve is the result of an exact numerical calculation. The assumed tube parameters are given in the text.

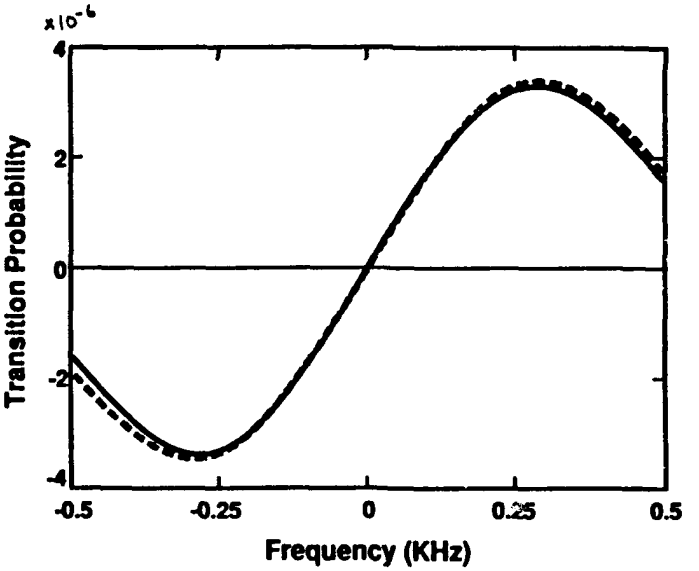


Figure 4. Odd part of the calculated contribution to the clock transition probability from the (3,1) - (4,0) sigma transition. This produces the Ramsey pulling. All parameters are the same as for Figure 3. Again, the dashed curve is the exact calculation.

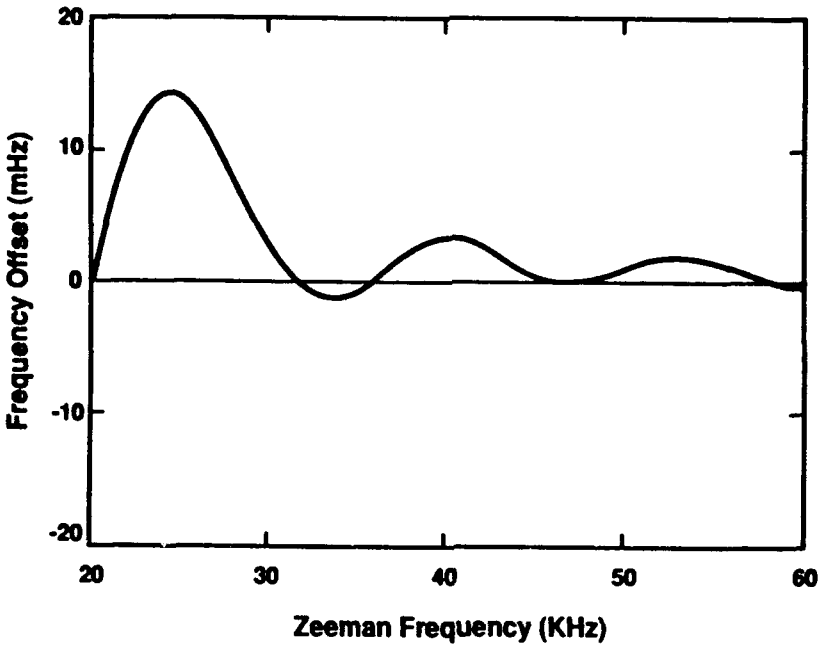


Figure 5. Calculated frequency offset as a function of Zeeman frequency due to Rabi pulling only for a model of an actual double beam tube. The parameters are given in the text. The approximate analytic results were used.

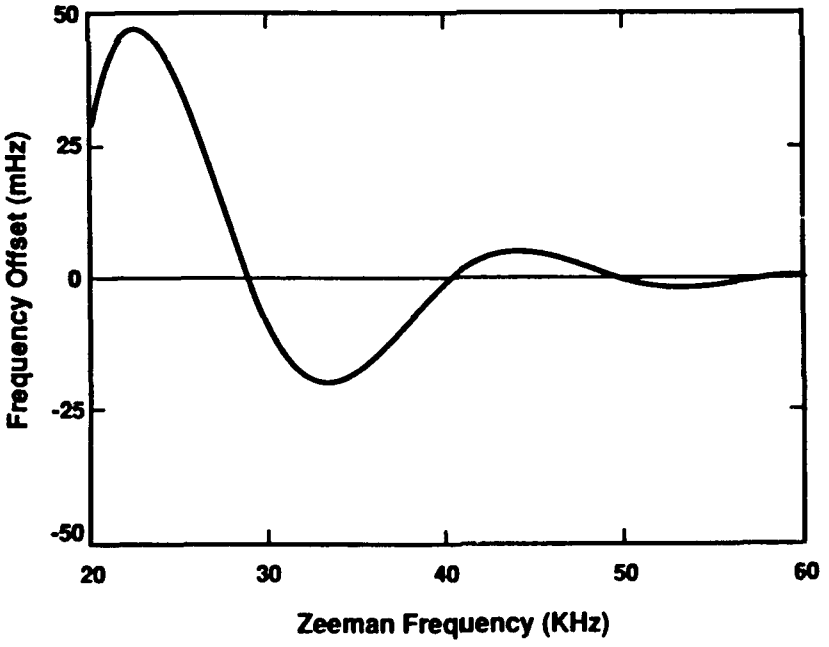


Figure 6. Calculated frequency offset for Ramsey pulling only for the same model as in Figure 5. Maximum pulling is about three times as large as the Rabi pulling.

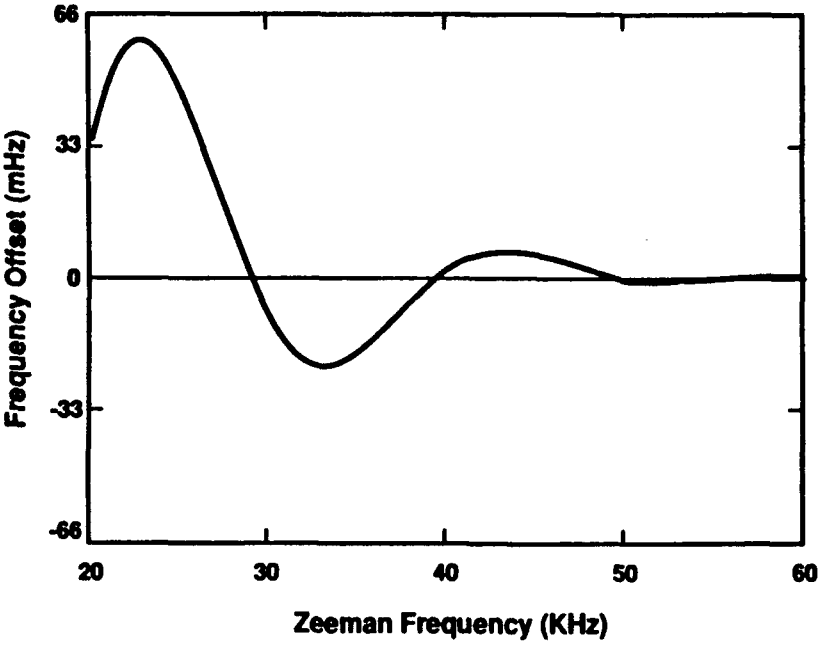


Figure 7. Total calculated frequency offset from both Rabi and Ramsey pulling from Figures 5 and 6.

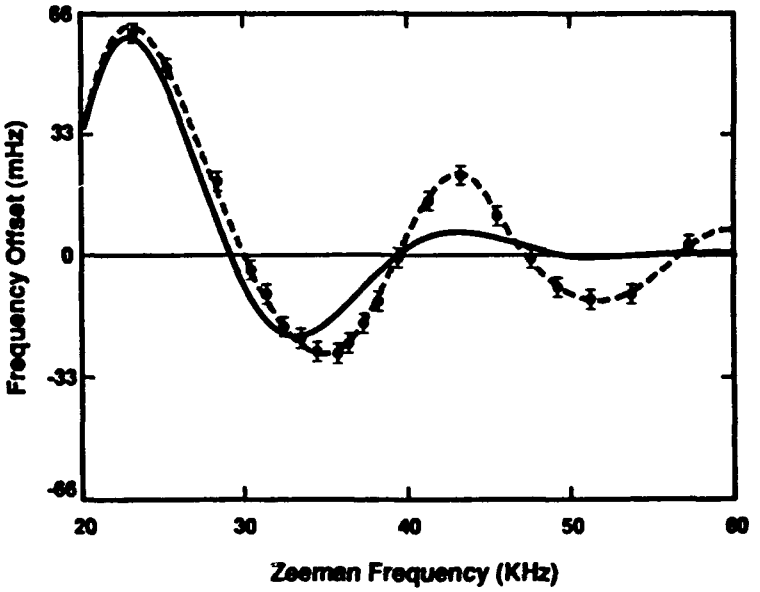


Figure 8. Measured total frequency offset in a double beam tube (Points and dashed line) and the calculated total offset (solid line) from Figure 7. The agreement between theory and experiment is quite good at low Zeeman frequencies. The more rapid damping of the oscillations in the calculated results at higher Zeeman frequencies is most likely due to difference in the calculated and actual velocity distributions because of non-ideal deflection fields in the actual tube.

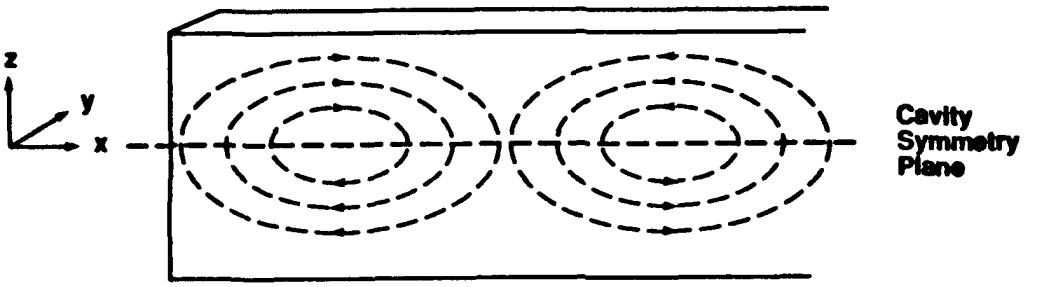


Figure 9. Microwave magnetic field lines in a conventional Ramsey cavity end section. The C-field is in the z direction and the cesium beam propagates in the y direction. Clearly, the microwave field is parallel to the C-field only at the symmetry plane.

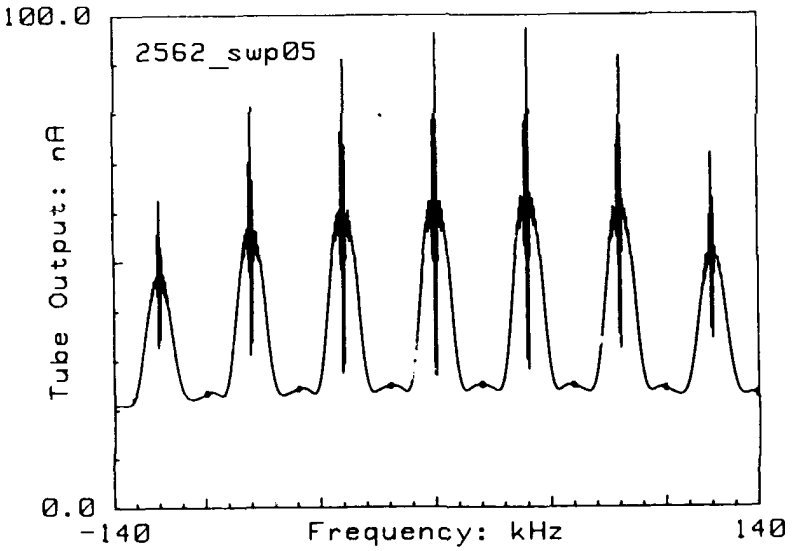


Figure 10. Full microwave spectrum of a double beam tube. Sigma transitions are clearly present between the pi transitions.

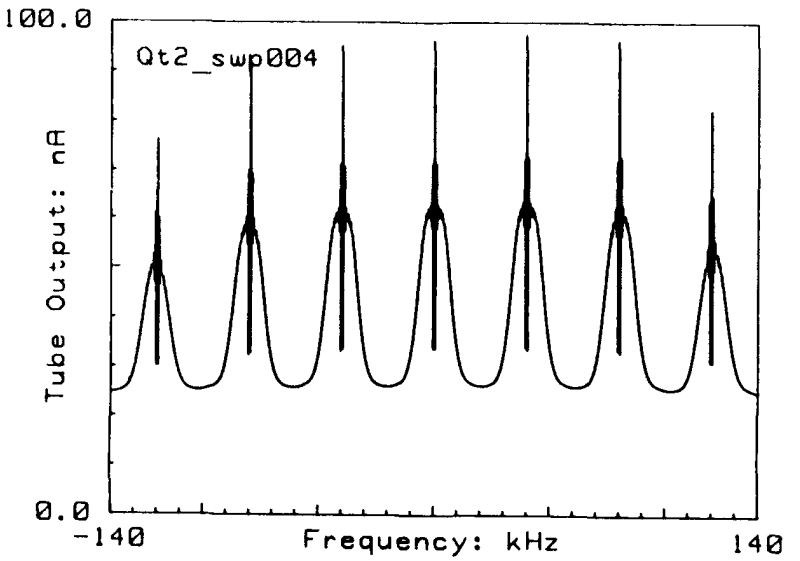


Figure 11. Full microwave spectrum of a new experimental beam tube. Sigma transitions are not visible on this scale. This tube should exhibit much smaller Ramsey and Rabi pulling.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A NEW HIGH-PERFORMANCE CESIUM BEAM TUBE COMPENSATED FOR RAMSEY PULLING

Louis F. Mueller
Douglas B. Dull
John A. Kusters

Hewlett-Packard Company
Santa Clara, CA 95052-8059

Abstract

Previous papers have discussed problems observed with high-performance cesium beam tubes which result from undesirable transitions of the cesium atom. The theory of why the problem occurs and possible corrective actions have been discussed in a companion paper. In particular, $\Delta m = \pm 1$ (σ) transitions have been shown to be responsible for the effect, now known as Ramsey Pulling.

This paper discusses the design and performance of cesium beam tubes which incorporate those corrections and which preserves the benefits of the high-performance beam tube, but with negligible Ramsey pulling. The design results in a tube which is smaller, lighter, and potentially longer-lived than its predecessor.

Introduction

In previous work [1-4], A. DeMarchi, *et.al.*, showed that long-term instability in low-performance cesium standards was related in many cases to a frequency shift associated with a change in rf power input to the beam tube. Further theoretical work by Cutler, *et.al.*, [5-6] demonstrated that this is due in part to $\Delta m = \pm 1$ (σ) transitions in the beam tube response.

Following the theoretical model, Cutler and Flory developed new atomic beam optics and magnetic state selector apparatus which theory had shown to reduce Ramsey pulling to a negligible level.

The design was coupled with an improved mechanical design which incorporates improvements in virtually all other areas of the beam tube. These are the result of nearly 30 years experience in building cesium beam tubes, and our continuing emphasis on Total Process Control [7-8].

An Improved High-Performance Tube

The results of the work described above is a new high-performance cesium beam tube designed to replace the existing Hewlett-Packard High-Performance tube. A photograph of the new and old tube is shown in Figure 1. Figure 2 shows the new tube installed in the HP5061B. Noticeable in the photographs are the adapters attached to the new tube. This became necessary when the decision was made to go to a smaller diameter tube. The adapters allow the tube to be mounted in any HP5060A or HP5061A/B.

Similarities

The new tube has the same performance specifications and the same part number as its predecessor. It meets all of the same specifications including figure of merit, acceleration compensation, beam current, electron multiplier gain, and magnetic field sensitivity. When installed in the 5061A/B, the set-up, alignment, and front panel readings are identical to that experienced with the older tube. Mechanically, with the spacer adapters, the new tube fits in the same tube cradle as the previous tube. When

CH2965-2/91/0000-554 $1.00 © 1991 IEEE

installed in the 5061, the spacer adapters permit the harmonic generator assembly to be located in the same region as before.

Differences

The new high-performance tube has negligible Ramsey pulling. As described above, it is smaller than the old tube. The new tube has the same length but is about 1.5 cm. smaller in diameter. The reduction in diameter results in a weight reduction of about 1.8 kg. without compromising performance. The smaller diameter also reduces the amount of torque that can be applied to the tube, thus ensuring that its alignment can be better maintained throughout the life of the tube.

As described in [6], the tube has a single atomic beam, in contrast to its predecessor's dual atomic beam. The cesium oven holds 20% more cesium. Additional gettering material is used throughout the tube. Improved electron multipliers and ion pumps are incorporated into the design.

To maintain the same overall beam flux, the oven is operated at a higher temperature, but with reduced power consumption. In addition, optimized beam optics and magnetic state selectors enable more efficient utilization of the cesium supply. The result of the changes is that the total life expectancy of the tube is expected to increase approximately 25%.

Performance

Figure 3 is the Ramsey response curve for the older high-performance beam tube. Particularly noticeable are the responses seen between the Rabi pedestals. The responses at half the Zeeman frequency are those identified as the unwanted transitions which are the primary cause of Ramsey pulling.

Figure 4 is the Ramsey response curve for the new high-performance tube. Of interest is the smooth continuous response in the regions between the Rabi pedestals.

Figure 5 shows the (0,0) and (1,1) transitions for a 40 kHz Zeeman frequency, also demonstrating the visible absence of $\Delta m = \pm 1$ responses.

Doubling the Zeeman frequency to 80 KHz. and increasing the vertical gain by a factor of 20, one obtains the response shown in Figure 6. From this, a very small amount of structure seems to be present in the 40 KHz region.

Increasing the Zeeman frequency to 200 KHz produces the response curve shown in Figure 7. Here a small response structure is finally evident. In comparison, an older high-performance tube was measured under the same conditions. Its response is shown in Figure 8. Evident is a definite Ramsey response, indicative of the non-parallelism of the microwave and static magnetic fields in the interaction region.

Comparison with Older Tubes

Figure 9 is a bar graph showing the measured line width of approximately 600 high performance beam tubes. The gaussian distribution is quite evident, as is the rather small standard deviation. The new tube exhibits an average line width of about 375 Hz. This is a result of choosing a different cut of the cesium velocity distribution to get the maximum reduction in unwanted transitions.

A consequence of this is that the true figure of merit averages around 17. Figure 10 shows the measured figure of merit for the same tubes from Fig. 9. The key question is the possible effect on normal operation with a lower figure of merit. Figure 11 is a collection of data taken on various 5061B instruments [9]. Evident in this data is that there is no correlation between the figure of merit and the measured Allan variance for measurement periods of 1 day and 2 days. Indeed, the data implies that the electronics probably dominates the time-domain stability of the cesium standard. The two major outliers at a 1 day sampling time are from units known to be environmentally sensitive.

Time Dependency Effects

As a part of the extensive testing performed to qualify the new tube for use, a series of special tests were performed looking for any form of latency effect or time-dependent effect which might affect its performance in a standard.

One such experiment used the equipment hookup shown in Figure 12. Here, the tube was exercised through the following sequence:

1. Set the tube for normal operation at the resonance peak, with a 40 kHz Zeeman frequency.
2. Step the microwave frequency 20 kHz.
3. Dwell at the new frequency for a determined period of time continuously recording the beam current.
4. Switch to the frequency used in step 1.
5. Continuously record beam current.
6. Compute the average beam current before and after switching. Compute the change in beam current from the average and plot.

Data was also taken in the reverse direction, starting at the valley 20 kHz away from resonance, then switching to the peak.

Figures 13 and 14 show some of the experimental data obtained. In both cases, the same general form of a time-dependent effect is seen. The peak-to-valley transition was done with beam currents of 500 nA. The other used beam currents of 280 nA.

Plotting the transient response, i.e., the total change in current from the initial value to the steady state value, as a function of residency time yields the plots shown in Figures 15 and 16.

Although the data was taken at different peak beam currents, the overall effect is quite similar.

A major change occurs when the peak beam current is decreased to 134 nA as shown at one point in Fig. 16. As this is repeatable, the non-linear character of the effect is quite noticeable.

All changes in peak beam current were done by changing the electron multiplier voltage. Changes in Zeeman spacing, or in ionizer current had no significant effect on the response.

Conclusion

The latency effect was observed on all cesium tubes examined, including the standard tube, the old high-performance tube, and the mini-tube used in the HP 5062C.

As the effect is primarily dependent on the electron multiplier gain, it appears to be a charging phenomenon either internal to the electron multiplier, or at the entrance orifice to the electron multiplier.

Given the transient behavior of the effect, and the normal operation of a cesium beam tube, the latency effect cannot affect the performance of a properly run cesium standard. However, this effect may be a serious limiting factor in cesium beam switching experiments.

Acknowledgments

The authors wish to acknowledge their indebtedness to Drs. Len Cutler and Curt Flory of HP Laboratories for their development of the beam optics and magnetic pole pieces which made the new tube possible, and to Dr. Robin Giffard, also of HP Laboratories, for making many of the measurements.

We also acknowledge the contribution of David Gottwals and Jerry Curran for their mechanical process expertise that made the physical embodiment of the theory practical, and the members of the Precision Time and Frequency group for their valued association in making the tube a production reality.

We also wish to thank Drs. Joseph White and Edward Powers of the Naval Research Laboratories for allowing us to use their experimental data.

References

1. A. DeMarchi, G.D. Rovera, and A. Premoli, "Pulling by Neighboring Transitions and its Effect on the Performance of Cesium Beam Frequency Standards," *Metrologia*, vol. 20, p. 37 (1984).

2. A. DeMarchi, "Understanding Environmental Sensitivity and Ageing of Cesium Beam Frequency Standards," *Proc. 1st European Frequency and Time Forum*, p. 288, Besançon (1987).

3. A. DeMarchi, "New Insights into Long Term Frequency Instabilities (Drift & Noise) of Cesium Beam Frequency Standards," *Proc 41st Annual Symposium on Frequency Control*, p. 53, (1987).

4. A. DeMarchi, "A First Account of Long Term Stability Results Obtained on Various Cesium Standards by the Power Sensitivity Minimization Technique," *Proc. 20th Annual Precise Time and Time Interval (PTTI) Meeting*, Washington, DC (1988).

5. L.S. Cutler, C. Flory, R.P. Giffard, and A. DeMarchi, "Frequency Pulling by Hyperfine σ-Transitions in Cesium Beam Atomic Frequency Standards," *Journal of Appl. Phys.*, (March 1991).

6. L.S. Cutler, C. Flory, R.P. Giffard and A. DeMarchi, "Frequency Pulling in Cesium Beam Frequency Standards due to $\Delta m = \pm 1$ (σ) Transitions," *Proc. 45th Annual Symposium on Frequency Control*, (1991).

7. J.A. Kusters and C.A. Adams, "Applications of Total Process Conurol Techniques in the Production of High Precision Quartz Resonators," *Proc. 39th Annual Symposium on Frequency Control*, p. 475, (1985).

8. E.R. Straka, R.W. Sarrica, and L. Hansen, "Total Process Control and its Implementation in the Manufacturing of Cesium Beam Tubes," *Proc. 44th Annual Symposium on Frequency Control*, p. 98, (1990).

9. J. White, E. Powers, Naval Research Laboratory, private communication, 1990. Used with permission.

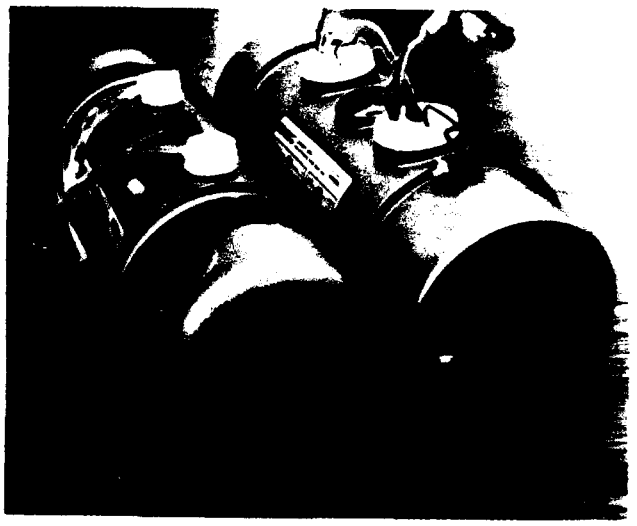


Figure 1 New High-Performance tube in the foreground, old tube in the background

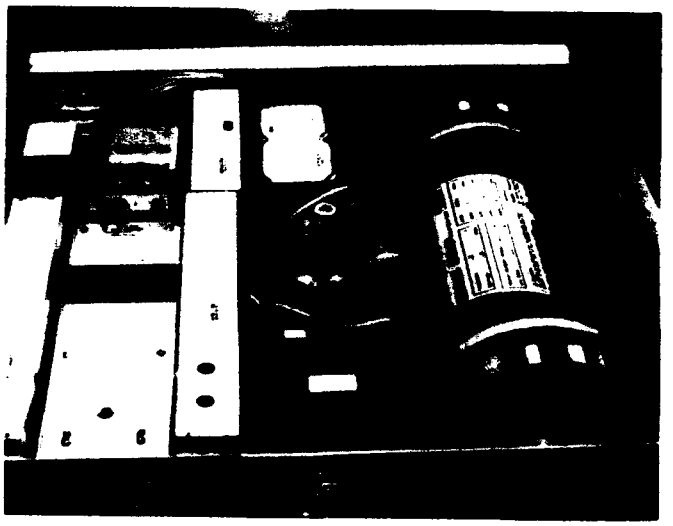


Figure 2 New High-Performance tube installed in Hewlett-Packard 5061B.

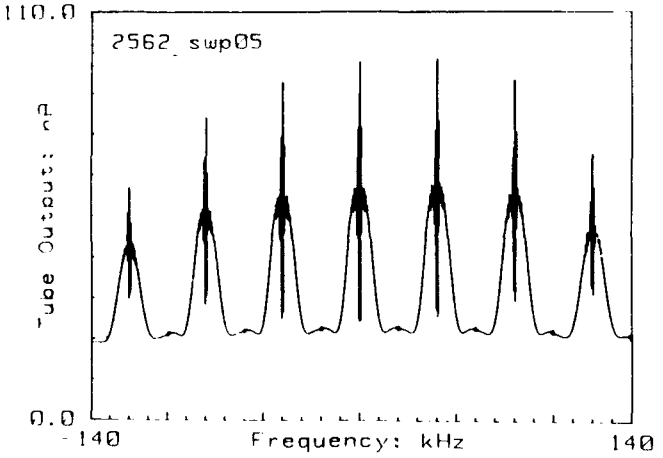


Figure 3 Old High-Performance cesium beam tube - Ramsey spectrum.

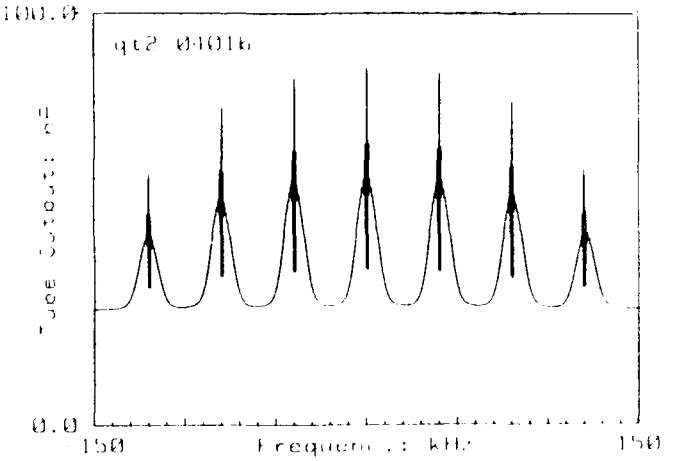


Figure 4 New High-Performance cesium beam tube - Ramsey spectrum.

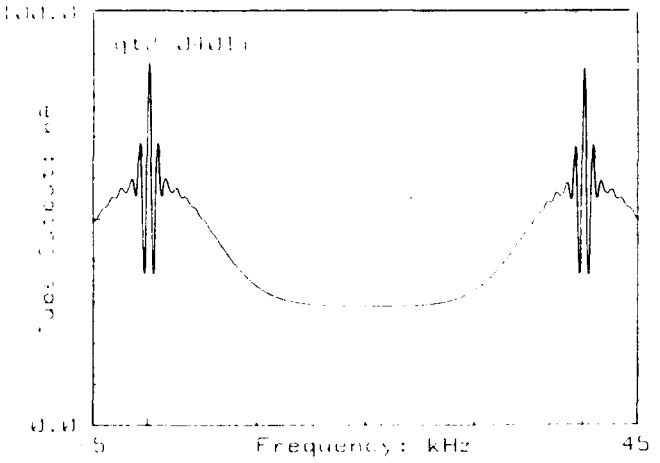


Figure 5 New tube showing the Ramsey response for the (0,0) and (1,1) peaks, 40 kHz Zeeman frequency.

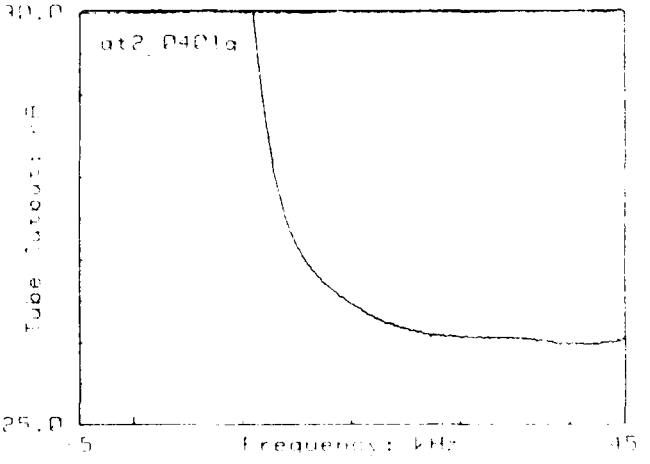


Figure 6 New tube, same as Fig. 5 but with 80 kHz Zeeman, 20 x increase in vertical scale.

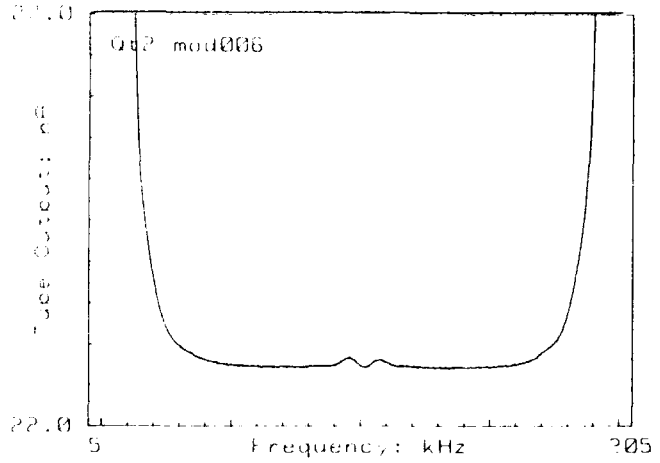


Figure 7 New tube, same as Fig. 5 but with 200 kHz Zeeman separation.

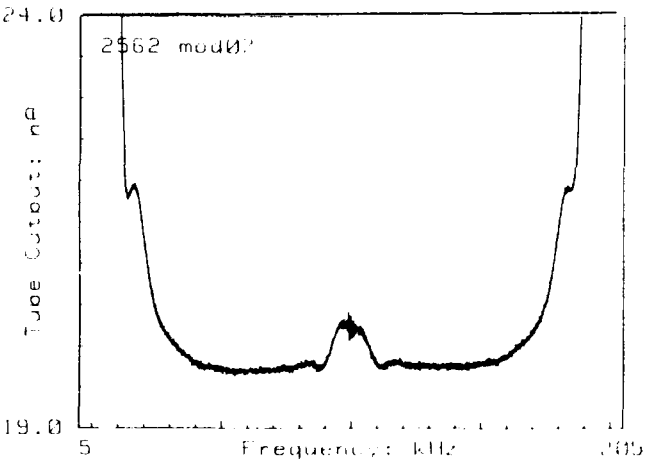


Figure 8 Old tube, same setup as in Fig. 7, 200 kHz Zeeman, showing Ramsey structure indicative of non-parallelism in the interaction region.

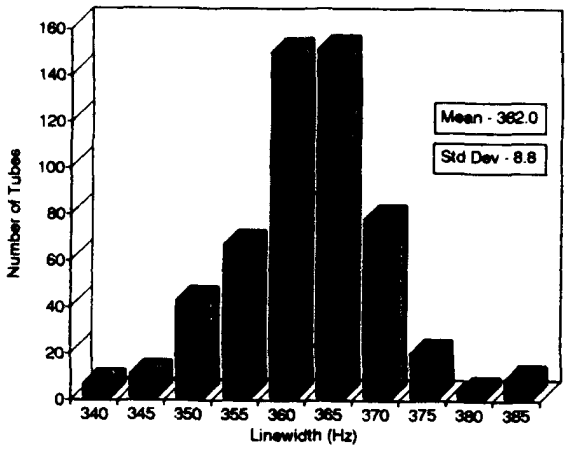


Figure 9 Measured line width of approximately 600 old High-Performance beam tubes.

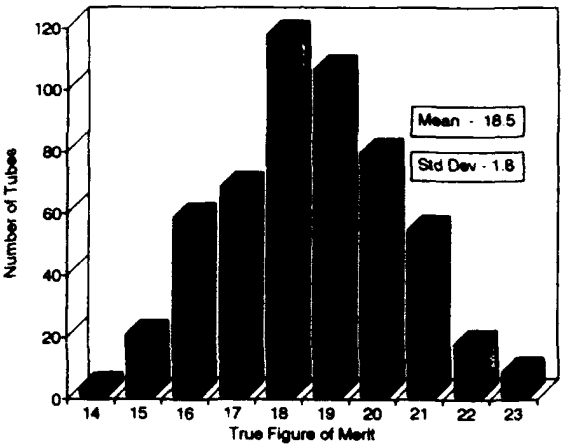


Figure 10 Measured figure-of-merit of approximately 600 old High Performance beam tubes.

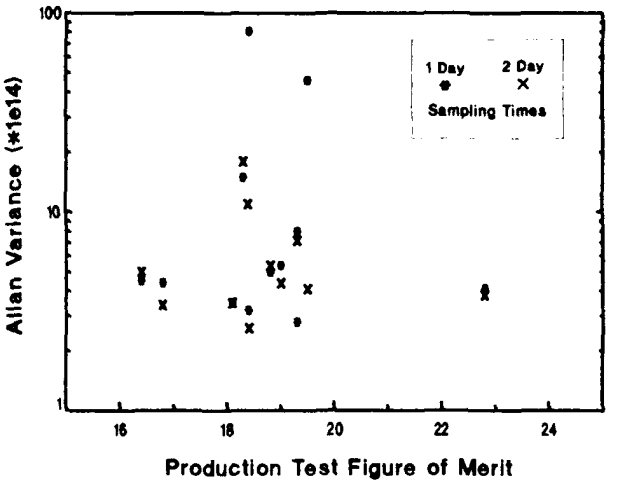


Figure 11 Allan Variance as a function of beam tube figure-of-merit for 12 HP 5061B, Opt. 004 instruments. Data courtesy of Naval Research Laboratories.

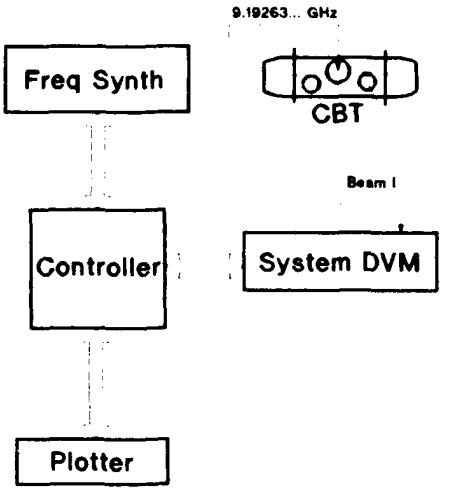


Figure 12 Equipment setup - Time dependent effects.

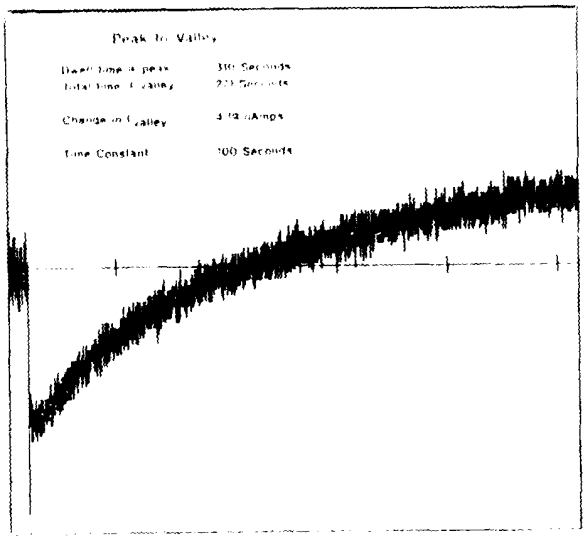


Figure 13 Measured latency effect, transition from resonance peak to valley.

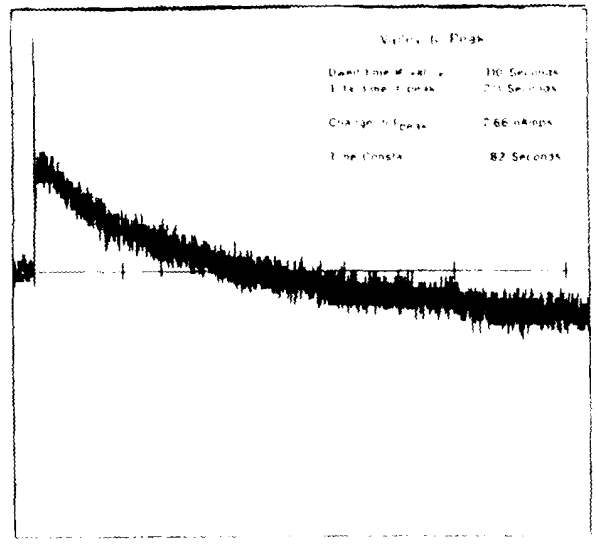


Figure 14 Measured latency effect, transition from resonance valley to peak.

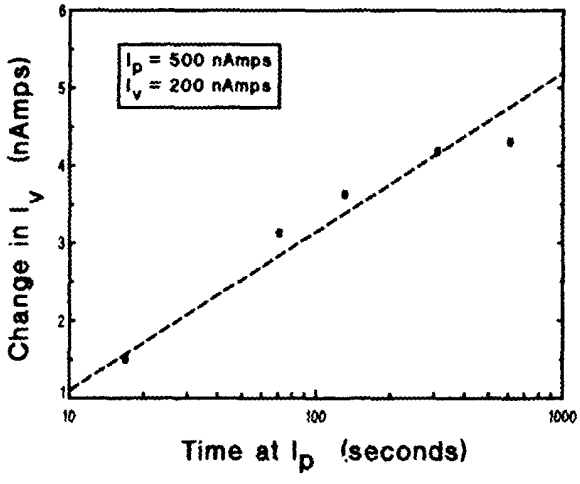


Figure 15 Latency effect observed for transtitions from resonance peak to valley.

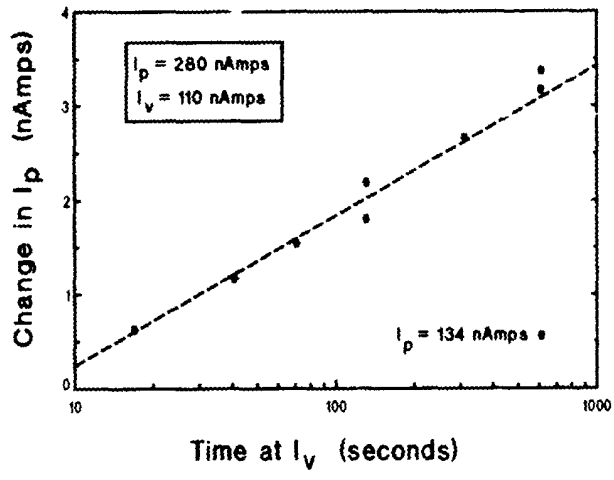


Figure 16 Latency effect observed for transitions from resonance valley to peak.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PROSPECTS IN IMPROVEMENT OF ACCURACY PERFORMANCES OF PASSIVE ATOMIC FREQUENCY STANDARDS

A.G.Gevorkjan , V.S.Zholnerov

Leningrad Scientific Research Radiotechnical Institute (LSRRI), 2, Rastrelli square, Leningrad ,193124 , USSR

Abstract

The results are presented concerning with the LSRRI's experimental and theoretical investigations to be aimed for improvement of accuracy characteristics of passive atomic frequency standards. It is given a brief review of the parametric frequency shifts of the clock atomic transition in a caesium atomic beam tube and those of rubidium gas cell atomic discriminator.The transformation of fluctuations in an atomic beam tube and in passive atomic frequency standand is considered. The limit accuracy characteristics of passive frequency standards are given that are derived on the basis of a multiparametric optimization of atomic beam tube and atomic gas cell discriminator as well as a complex optimization of whole standard.

1.Introduction

Frequency stability of the passive atomic frequency standard (AFS) is governed by fluctuation behaviour of separate blocks of the AFS. The influence of fluctuations on frequency stability is realized either by the direct conversion of the signal, which controls the quartz oscillator (QO) frequency, into the fluctuations or by the conversion of the atomic transition frequency due to parametric shifting of frequency.The parametric frequency shifts of the clock atomic transition in a Cs atomic beam tube (ABT) [1,2] and those of rubidium gas cell atomic discriminator (AD) had been studied in some details. So we shall give only a brief review on the problem of shifts. Our primary attention will be focussed on the conversion of fluctuations in the passive AFS.

2. Review of parametric frequency shifts of the clock atomic transition in the passive atomic frequency standards

Frequency shifts of the atomic transition in the caesium ABT can be devided into three groups taking into account the behaviour of the shifts under external perturbations : the shifts conditioned by (1) direct magnetic field in Ramsey resonator,(2)the temperature of the atomic beam source (or the ambient temperature), (3) microwave excitation power. Temperature variations of the atomic beam source may be responsible for frequency shifts due to thermal detuning of the microwave cavity or to the effect of spray current flowing through the resonator under the action of thermo-emf which arises at intermetal junction in ABT of the applied construction. The slope of the dependence of the transition frequency on the temperature of the beam source is equal $2 \cdot 10^{-13}$ 1/K in production ABT of commercial AFS. This value had been reduced to $2 \cdot 10^{-14}$ 1/K as a result of improvement of ABT by introducing of the dielectric isolation interrupting the thermo-currents.

Some kinds of frequency shifts belong to the category of the microwave power dependent shifts. These are the following : the shifts arising from frequency pulling of the clock atomic transition by ajacent Zeeman transitions and by the microwave resonator, the shift arising from phase detuning of the interaction regions of Ramsey resonator,and the shift resulting from the second order Doppler effect. The shift arising from frequency pulling by ajacent Zeeman transitions is of the most considerable value [1-2]. This shift is an oscillating function (with some zero points) of the direct magnetic field strength and there are several magnitudes of the magnetic field strength at which the frequency shift due to microwave field power variations is equal to zero. For the improved ABT the experimentally measured slope of the atomic transition frequency on the relative change of microwave power was found to be $2 \cdot 10^{-12}$ at the direct magnetic field strength corresponding to the smoothest microwave power dependence of frequency.

There are several kinds of parametric fre-

CH2965-2/91/0000-562 $1.00 © 1991 IEEE

quency shifts of the atomic transition in the AD of the rubidium gas cell: the shifts associated with (1) direct magnetic field, (2) the temperature of the absorbing cell, (3) microwave power and (4) light. Light shifts are divided into two groups: direct shift which is proportional to the pump intensity, and indirect shifts which are proportional to the derivatives of the light shift with respect to such parameters as the temperatures of the spectral lamp, of the isotopical filter, of the absorbing cell etc [3]. Minimization of the slope of the dependence of the atomic transition frequency on the temperature of the absorbing cell is achieved by optimal selection of a buffer gas mixture [4]. The microwave power dependence of the transition frequency is caused by the presence of the light shift and of the absorbing cell temperature shift. When these kinds of shifts are suppressed down to zero the dependence becomes negligible.

3. Conversion of fluctuations in the passive AFS

Experimental investigations of fluctuations in separate assemblies of the AFS have shown that the fluctuations involve flicker components. So the power spectral densities (PSD) of the fluctuations of QO frequency, of phase and power of the frequency multiplier, of the direct magnetic field strength and the temperature of the atomic beam source in a commercial atomic-beam AFS are, respectively :

$$W_{\delta}(\Omega)=\left(23.1+\frac{0.2}{\Omega}+\frac{0.16}{\Omega^2}\right)\cdot 10^{-9}\ \left(\mathrm{Hz^2/Hz}\right)$$

$$W_{\varphi}(\Omega)=\left(7.3+\frac{14.4}{\Omega}+\frac{114}{\Omega^2}\right)\cdot 10^{-6}\ \left(\mathrm{rad^2/Hz}\right)$$

$$W_{P}(\Omega)=\left(1.06/\Omega^{1.24}\right)\cdot 10^{-16}\ \left(\mathrm{W^2/Hz}\right)$$

$$W_{H}(\Omega)=\left(0.067+2.96/\Omega^{1.37}\right)\cdot 10^{-12}\left(\mathrm{Oe^2/Hz}\right)$$

$$W_{T}(\Omega)=\left(2.34/\Omega^{1.46}\right)\cdot 10^{-2}\ \left(\mathrm{K^2/Hz}\right)$$

The PSD are of the same pattern for the fluctuations of the ABT output current and for the fluctuations of current in the photodetector of the gas cell discriminator :

$$W_{I}(\Omega)=\left(A/\Omega^{\gamma}\right)+B \qquad (1<\gamma<2).$$

In order to find the AFS response to fluctuating perturbations one should know the conversion behaviour of the fluctuations in the main assembly of AFS-atomic discriminator. Let us consider the conversion of fluctuations in ABT. We assume the atomic beam intensity $n(t)$, the amplitude $\lambda(t)$ of the magnetic component of microwave field and the microwave phase $\xi(t)$ to be random quantities. PSD of the ABT current fluctuations in the presence of phase-modulated microwave field $\varphi(t)=\Delta\varphi\cos\Omega_m t$ takes the following form with regard to synchronous detection [5] :

$$W_{i}(\Omega)=e^2\left(I_0+0.5K_F\left[W_n(\Omega+\Omega_m)+W_n(\Omega-\Omega_m)\right]+\right.$$
$$+K_n W_n(\Omega)+K_\lambda(\Omega)W_\lambda(\Omega)+K_\xi(\Omega)W_\xi(\Omega)+$$
$$+K_\lambda^{+}(\Omega)W_\lambda(\Omega+\Omega_m)+K_\lambda^{-}(\Omega)W_\lambda(\Omega-\Omega_m)+ \qquad (1)$$
$$\left.+K_\xi^{+}(\Omega)W_\xi(\Omega+\Omega_m)+K_\xi^{-}(\Omega-\Omega_m)\right)$$

where

$$i_0=I_F+I_0\int_0^{\infty}\left[1+J_0(z)\right]\cos\frac{\Omega L}{v}\,X(v)\,dv\ ,$$

$$I_F=I_0\int_0^{\infty}W_F(v)w(v)dv\ ;\quad I_0=eK_d n_0\ ;$$

$$z=2\Delta\varphi\sin(\Omega_m L/2v);$$

$$K_F=e^2K_d^2\left[\int_0^{\infty}W_F(v)w(v)dv\right]^2\ ;$$

$$K_n=e^2K_d^2\left[\int_0^{\infty}U(v)\cos\left(\frac{\Omega_m L}{2v}+\Phi\right)dv\right]^2\ ;$$

$$K_\lambda(\Omega)=I_0^2\left\{\left[\int_0^{\infty}Q(v)\left(1+\cos\frac{\Omega L}{v}\right)\cos\left(\frac{\Omega_m L}{2v}+\Phi\right)dv\right]^2+\right.$$
$$\left.+\left[\int_0^{\infty}Q(v)\sin\frac{\Omega L}{v}\cos\left(\frac{\Omega_m L}{2v}+\Phi\right)dv\right]^2\right\}\ ;$$

$$K_\xi(\Omega)=I_0^2\left\{\left[\int_0^{\infty}R(v)\left(1-\cos\frac{\Omega L}{v}\right)\cos\left(\frac{\Omega_m L}{2v}+\Phi\right)dv\right]^2+\right.$$
$$\left.+\left[\int_0^{\infty}R(v)\sin\frac{\Omega L}{v}\cos\left(\frac{\Omega_m L}{2v}+\Phi\right)dv\right]^2\right\}\ ;$$

$$K_\lambda^{\pm}(\Omega)=I_0^2\left\{\left[\int_0^{\infty}Q(v)\sin\frac{\Omega_m L}{2v}\left[1+\cos\frac{(\Omega\pm 2\Omega_m)L}{v}\right]dv\right]^2+\right.$$

$$+\left[\int_0^\infty Q(v)\cos\frac{\Omega_m L}{2v}\left[1+\cos\frac{(\Omega\pm2\Omega_m)L}{v}\right]dv\right]^2+$$

$$+\left[\int_0^\infty Q(v)\sin\frac{\Omega_m L}{2v}\sin\frac{(\Omega\pm2\Omega_m)L}{v}\,dv\right]^2+$$

$$+\left[\int_0^\infty Q(v)\cos\frac{\Omega_m L}{2v}\sin\frac{(\Omega\pm2\Omega_m)L}{v}\,dv\right]^2\Bigg\};$$

$$K_\xi^2(\Omega)=I_0^2\Bigg\{\left[\int_0^\infty R(v)\sin\frac{\Omega_m L}{2v}\left[1-\cos\frac{(\Omega\pm2\Omega_m)L}{v}\right]dv\right]^2+$$

$$+\left[\int_0^\infty R(v)\cos\frac{\Omega_m L}{2v}\left[1-\cos\frac{(\Omega\pm2\Omega_m)L}{v}\right]dv\right]^2+$$

$$+\left[\int_0^\infty R(v)\sin\frac{\Omega_m L}{2v}\sin\frac{(\Omega\pm2\Omega_m)L}{v}\,dv\right]^2+$$

$$+\left[\int_0^\infty R(v)\cos\frac{\Omega_m L}{2v}\sin\frac{(\Omega\pm2\Omega_m)L}{v}\,dv\right]^2\Bigg\};$$

$$U(v) = J_1(z)X(v)\sin(\langle\delta\rangle L/v)\ ;$$

$$Q(v) = (1/v)J_1(z)Y(v)\sin(\langle\delta\rangle L\ v)\ ;$$

$$R(v) = J_1(z)X(v)\cos(\langle\delta\rangle L/v)\ ;$$

$$\Phi = -\arctan\frac{\int_0^\infty J_1(z)X(v)\sin(\Omega_m L/2v)dv}{\int_0^\infty J_1(z)X(v)\cos(\Omega_m L/2v)dv}\ ;$$

$$X(v) = \sin^2(2\langle\Lambda\rangle l/v)W_S(v)w(v);$$

$$Y(v) = \sin(4\langle\Lambda\rangle l/v)W_S(v)w(v);$$

$$W_P(v) = \sum_{k=1}^{16}[A_{kk}(v)B_{kk}(v)]/16\ ;$$

$$W_S(v) = \frac{[A_{22}(v)-A_{11}(v)][B_{11}(v)-B_{22}(v)]}{16}\ ;$$

$\Lambda=(\hat{\mu}\,\hat{E}_\lambda)\lambda(t)]/(2\hbar)$ - the microwave excitation parameter; $\hat{\mu}$ -the magnetic dipole operator for caesium atom; $\hat{E}_\lambda$ - the polarization vector of the magnetic component of microwave field; l - the length of the interaction region; L -the distance between the interaction regions of Ramsey resonator; δ - the amount off-resonance of microwave field with respect to the atomic transition frequency; A_{mm} , B_{mm} - the coefficients describing the influence of the first and the second magnetic selectors on the population of Zeeman sublevels of Cs atomic ground state; K_d-the efficiency of the atomic beam detector; e - the electronic charge ; $w(v)$ - the atomic velocity distribution of the beam; $W_\lambda(\Omega)$, $W_\xi(\Omega)$,$W_n(\Omega)$- PSD of random processes $\lambda(t)$, $\xi(t)$ and $n(t)$; $J_0(z)$ and $J_1(z)$ - the first kind Bessel's functions of the zero and first order. The suscripts 1 and 2 denote caesium states $|3,0\rangle$ and $|4,0\rangle$, respectively.

As seen from expression (1), the major contribution to the PSD of fluctuations of QO frequency control signal under operation conditions ($\langle\delta\rangle=0$) is provided by several effects: shot noise of the atomic beam , converted phase fluctuations of microwave field , which are transferred from Fourier frequencies being equal to zero or in the vicinity of the second harmonics of phase modulated microwave field. The last fluctuations act as an additional fluctuation source since they are not filtrated by a system of automatic-frequency control. These fluctuations set a limit on the short-term frequency stability of the AFS. Over the lower frequency range the conversion coefficient of microwave field frequency fluctuations $K_\xi(\Omega)$ without Fourier frequency shifting is equal to the slope squared of the ABT amplitude-frequency characteristic in the vicinity of zero detuning of the field frequency. The conversion of fluctuations is of the analogous pattern in the gas-cell AD but the conversion is described by much more complicated expressions.

Now we consider the conversion of fluctuations by the AFS frequency locked loop. The equivalent block circuit of the passive AFS is shown on figure 1,indicating both the transfer functions of separate assemblies of the AFS and random perturbations.As it has been shown above, the PSD of the inherent fluctuations of the QO, those of non-multiple frequency converter and of the AD are as follows :

$$W_0(\Omega) = (A_1/\Omega^2) + (B_1/\Omega) + C_1 \qquad (2a)$$

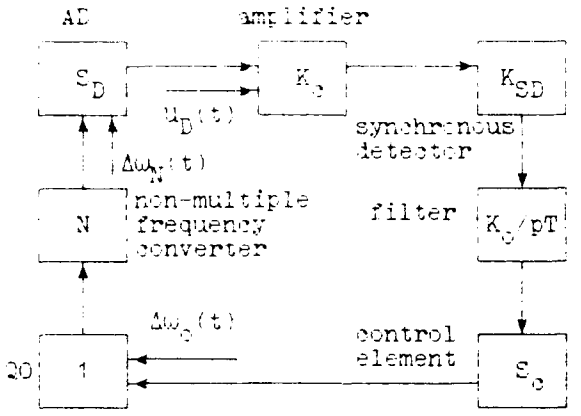


Figure 1.The equivalent block circuit of the passive AFS.

$$W_N(\Omega) = A_2 + B_2\Omega + C_2\Omega^2 \qquad (2b)$$

$$W_D(\Omega) = (A_3 / \Omega^\alpha) + C_3 \qquad (2c)$$

The quantity C_3 includes the PSD of white frequency noise of microwave field,the PSD having been translated by the AD conversion from Fourier frequencies being near the second harmonic of the phase-modulated field. As had been shown in [6], in this case the relative two-sample rms frequency variation of the AFS is

$$\sigma(\tau) = \left[\sigma_0^2(\tau) + \sigma_N^2(\tau) + \sigma_D^2(\tau) \right]^{1/2} \qquad (3)$$

where

$$\sigma_0^2(\tau) = \frac{1}{\omega_0^2}\left[\left(C_1 - \frac{A_1}{\Omega_0^2}\right)\frac{K}{\Omega_0\tau} + \frac{A_1}{\Omega_0\tau^2} + \frac{B_1 E(\Omega_0\tau)}{\pi\,\Omega_0^2\tau^2}\right];$$

$$\sigma_N^2(\tau) = \frac{1}{N^2\omega_0^2}\left[\left(C_2\Omega - \frac{A_2}{\Omega_0}\right)\frac{K}{\tau} + \frac{A_2}{\tau} + \frac{B_2 E(\Omega_0\tau)}{\pi\,\tau^2}\right];$$

$$\sigma_D^2(\tau) = \frac{1}{N^2 S_D^2 \omega_0^2}\left[\frac{A_3 \sec(\alpha\pi/2)}{(\Omega_0)^{\alpha+1}\,\tau^2} F(\alpha,\Omega_0,\tau) + \right.$$

$$\left. + \frac{C_3}{\tau}\left(1 - \frac{K}{\Omega_0}\right)\right]; \quad E(\Omega_0\tau) = 2\ln 2 + 3\ln(\Omega_0\tau\gamma/2) +$$

$$+ 0.5\left[e^{-2\Omega_0\tau}\,\overline{Ei}(2\Omega_0\tau) + e^{2\Omega_0\tau}\,Ei(-2\Omega_0\tau)\right] -$$

$$- 2\left[e^{-\Omega_0\tau}\,\overline{Ei}(\Omega_0\tau) + e^{\Omega_0\tau}\,Ei(-\Omega_0\tau)\right];$$

$$F(\alpha,\Omega_0,\tau) = \frac{(\Omega_0\tau)^{\alpha-1}}{\Gamma(\alpha)}\left\{2^{\alpha-3}\left[{}_1F_1(1,\alpha;2\Omega_0\tau) + \right.\right.$$

$$\left. + {}_1F_1(1,\alpha;-2\Omega_0\tau)\right] - {}_1F_1(1,\alpha;-\Omega_0\tau) -$$

$$\left. - {}_1F_1(1,\alpha;\Omega_0\tau) - (2^{\alpha-2}-2) - \frac{(2^\alpha-2)(\Omega_0\tau)^2}{\alpha(\alpha+1)}\right\} -$$

$$-4\sinh^4(\Omega_0\tau/2);$$

$\gamma = 1.781072$ - Euler constant, $\overline{Ei}(x)$ and $Ei(-x)$ - the exponential integrals, $\Gamma(\alpha)$ - gamma-function , $\Omega_0 = (NS_D K_0 K_{SD} K_0 S_0)/T$ - the out-off frequency of the AFS frequency lock loop.

Fig.2 and 3 show the dependences of the frequency stability $\sigma(\tau)$ on the time measurement interval τ and on the out-off frequency $\Omega_0 = 10^{-3}$ Hz (1), 10^{-2} Hz (2), 10^{-1} Hz (3),1 Hz (4),10 Hz (5),10^2 Hz (6) for AFS with middle- and high-quality AD. In the figures curve A represents the inherent frequency stability of the QO and curve B characterizes the contribution of AD at $\Omega_0 \to \infty$.

4.Frequency stability of the passive AFS and ways to improve it

In order to determine the accessible limit stability of the passive AFS it is essential to choose the AFS parameters in an optimal manner. For this purpose a special program has been worked out of two-criterion multiparametric optimization of the rubidium gas cell AD with respect to a minimum light shift and a maximum figure of merit. Besides, a program of optimization of the geometry-optical Cs ABT system has been developed with respect to a maximum figure of merit as well as a program of complex optimization of the passive AFS with respect to the minimal value of integral quality index Q; the quality index Q is equal to the area between the stability curve $\sigma(\tau)$ and the best frequency stability curve over a chosen measurement time interval $[\tau_1,\tau_2]$ [6]. Fig. 4 shows the frequency stability curve $\sigma(\tau)$ for an atomic-beam frequency standard at the optimal out-off frequency Ω_0 in the case of a compact ABT with the optimal geometry-optical system. The experimental value of $\sigma(\tau)$ is equal to $(3\text{-}5)\cdot10^{-12}$ at τ = 1 s and $(3\text{-}5)\cdot10^{-14}$at τ = 1 day.The analysis of the partial contributions of the fluctuations from the direct magnetic field, the atomic beam source temperature and microwave power into the long-term fre-

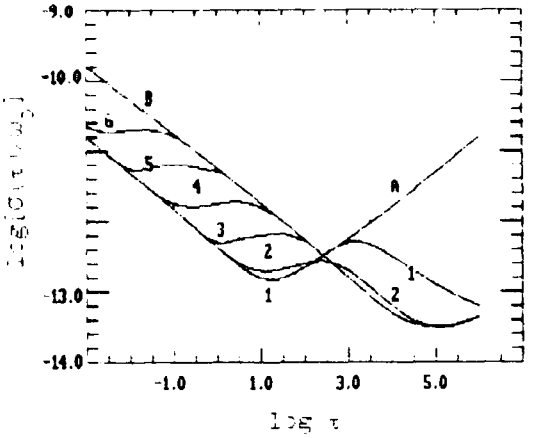


Figure 2.

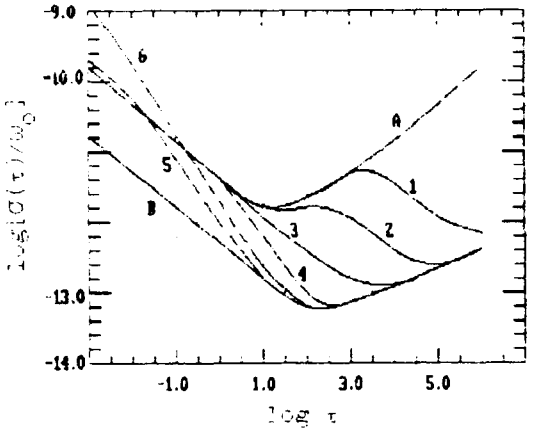


Figure 3.

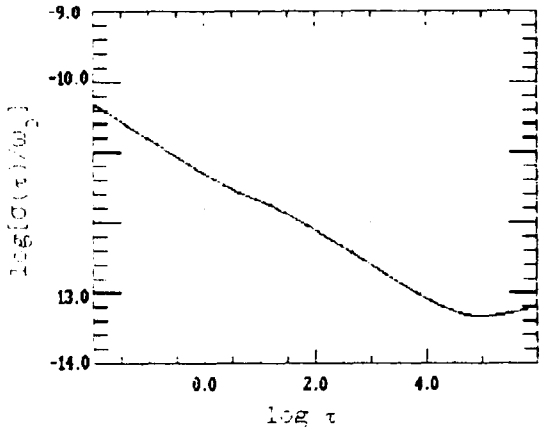


Figure 4.

quency stability shows that for further improvement of the long-term frequency stability one need to reduce fluctuations of the current source, which produces direct magnetic field, to lower the temperature shift of the ABT frequency.

The optimization of the gas cell AFS shows that its accessible limit stability is $\leq 1 \cdot 10^{-12}$ at $\tau = 1$ s and $(3-5) \cdot 10^{-14}$ at $\tau = 1$ day. However, the experimental studies attest that the initially established operation conditions providing zero light shift cease to be optimal under long-term operation due to ageing of the spectral lamp, isotopical filter or absorbing cell. This leads to systematic relative frequency drift with value of $(2-3) \cdot 10^{-12}$ over the month. So further investigations are required to reduce this drift.

Another way of improving the frequency stability of the passive AFS is to apply new physical concepts in designing of the AD. For instance, the application of laser pumping and detection of the atomic beam allows to increase essentially the AD figure of merit and to reduce systematic frequency shifts [7]. Some theoretical and experimental work has been done on the optimal choice of laser pumping and detection processes as well as that of the parameters of injection laser [7]. Using the results of the work as the base, the specialists of Leningrad Scientific Research Radiotechnical Institute (LSRRI) together with their collaborators of STA's "Saljut" and "Polus" complete the development of the pre-production sets of a compact AFS with atomic beam laser pumping and detection. The frequency stability of such AFS will be $\sigma(\tau) = 3 \cdot 10^{-12} \tau^{-1/2}$.

References

[1] A.De Marchi, G.D.Rovera, A.Premoli, Metrologia, v.20, pp. 37-47, 1984.

[2] A.De Marchi, IEEE Trans. UFFC-34, pp.598-601, 1987.

[3] E.B.Alexandrov, et. al., Zhurnal Technicheskoy Fisiki (USSR), v.59, pp.118-124, 1989.

[4] V.S.Zholnerov, Optica i Spectroscopiya (USSR), v.43, pp.957-960, 1977.

[5] V.S.Zholnerov, Radiotechnica i Electronica (USSR), v.35, pp.1702-1708, 1990.

[6] V.S.Zholnerov, Radiotechnica i Electronica, to be published.

[7] A.N.Besedina, et.al., Kvantovaja Electronica (USSR), v.18, No.2, 1991.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

COMPACT RECTANGULAR-CYLINDRICAL CAVITY FOR RUBIDIUM FREQUENCY STANDARD

E. Eltsufin, A. Stern and S. Fel*

TFL-Time & Frequency Ltd.
P.O.B. 1792 Holon 58117, ISRAEL

Abstract - TFL designed a novel compact rectangular-cylindrical dielectric-filled 6.8GHz microwave cavity which incorporates Step-Recovery-Diode (SRD) multiplier for use in a rubidium vapor cell frequency standard. The design comprises an upper rectangular part with dielectric slab and a semi-cylindrical lower part. The cavity operates in a pseudo TE_{101} mode where the field in the upper part are compressed due to the dielectric slab and the field in the lower part are compressed due to the cylindrical shape thus compensating the influence of the glass cells. This increases the field intensity inside the vapor cell and at the same time retains the field uniformity. An even higher efficient use of the magnetic field is achieved due to a high filling factor of cavity's free space with the glass cells. The circuit input impedance has been successfully designed to be a real 50 ohm to ease tuning and preclude reflections and consequent oscillations of the circuit. The actual required input RF power is about 0.1W, which is the known threshold value for the SRD.

Introduction

A microwave module is an essential part of any rubidium frequency standard and its performance is vital for the whole system operation. A microwave module usually incorporates a microwave cavity with rubidium vapor cells and a Step Recovery Diode (SRD) multiplier. When designing this module one needs to solve several problems. Firstly, a relatively high (about 23-27 dBm) RF power which is generally required for proper operation may cause internal EMI and overheating. Another problem is that the SRD bias point is very critical and even a small drift may cause SRD oscillation with spurious frequencies. This instability is caused by mismatching and consequent reflections into the multiplier circuit and into the microwave cavity.

The traditional design for the microwave cavity in a rubidium frequency standard has been a cylindrical resonator operating in a TE_{011} mode. In order to resonate at the 6.8GHz rubidium-clock-transition-frequency this resonator must, however, have a rather large diameter. A rectangular TE_{101} cavity, partially loaded with dielectric, which enables a considerable size reduction was introduced by H.E Williams et al [1].

* Advisor to TFL.

We have found, however, that the inclusion of the rubidium glass cell into the later cavity distorts considerably the field lines, degrades the Q factor and shifts the resonance frequency of the cavity.

The paper describes and analyses a novel design of a rectangular-cylindrical microwave cavity that takes into account the presence of the glass cell. The SRD multiplier circuit which forms an integral part of the design is also discussed.

Rectangular TE_{101} mode cavity

Consider first a rectangular, dielectric loaded, cavity which is operating in a pseudo TE_{101} mode [1]. The cavity is shown in Fig.1.

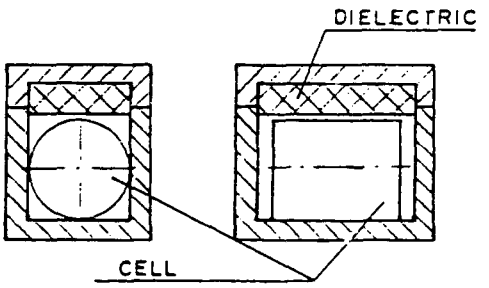


Fig 1. Rectangular dielectric loaded cavity

The propagation velocity of electromagnetic waves in a dielectric filled microwave cavity depends on the dielectric constant and on the dielectric thickness. It may be interesting to evaluate the wavelength elongation caused by the presence of a glass cell in a rectangular microwave cavity. The elongation is defined as the difference between the effective wavelength in the cavity and free-space wavelength. A glass thickness, Tx, is defined as the total thickness of a glass along the X axis and is dependent on the Y position. Consider a cylindrical glass cell located inside the low part of the rectangular cavity as shown in Fig.2.

The glass thickness, T_x, for $y \leq R_1$, as is given in Fig.2, is:

$$T_x = \sqrt{(R_2^2 - y^2)} - \sqrt{(R_1^2 - y^2)} \qquad (1)$$

CH2965-2/91/0000-567 $1.00 © 1991 IEEE

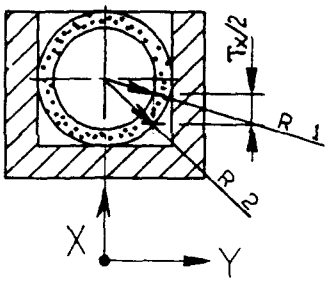


Fig 2. Rectangular cavity with a glass cell

The effective elongation is found from:

$$E_\epsilon = 2T_x(\sqrt{\epsilon} - 1) \qquad (2)$$

where, $\epsilon = 3.78$ is the glass dielectric constant. Then the fractional elongation is:

$$E_F = E_\epsilon / 2R_2 =$$

$$= [\sqrt{(1-(y/R_2)^2)} - \sqrt{((R_1/R_2)^2-(y/R_2)^2)}]*(\sqrt{\epsilon}-1) \qquad (3)$$

The fractional elongation derived from (3) is plotted in Figure 3.

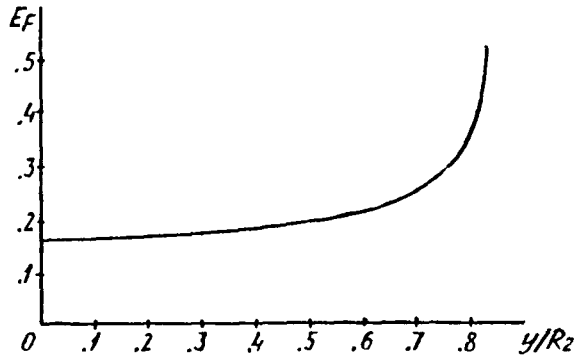


Fig 3.Fractional elongation
in the rectangular cavity

This calculation shows that a glass cell cause a high distortion of magnetic field inside the rectangular cavity.

Rectangular-cylindrical cavity

In order to account for the presence of the cylindrical glass cell a novel rectangular-cylindrical cavity was devised (see Figs.4 and 5). The design comprises an upper rectangular part with dielectric slab and a semi-cylindrical lower part. The cavity operates in a pseudo TE_{101} mode where the field lines in the upper part are compressed due to the dielectric slab. The lines in the lower part are compressed due to the cylindrical shape thus compensating the influence of the glass cell. This increases the field intensity inside the vapor cell and at the same time retains the field uniformity. An even higher efficient use of the magnetic field is achieved due to a high filling factor of the cylindrical cell inside the rectangular-cylindrical cavity.

Microwave cavity design goals

The following requirements have been formulated for the new microwave cavity design:

1. Quality factor of a cavity should be more than 20 with the design goal of 100 to consume minimum RF power.
2. Cavity should be compact to reduce cavity heaters power consumption and to improve magnetic field immunity.
3. Microwave magnetic field should be as homogeneous as possible with its maximum in the rubidium cells center.
4. Thermal field of a cavity should be homogeneous with a small temperature gradient to prevent rubidium condensation on a cell windows.
5. Cavity should be a two part construction to enable the insertion of cells.

Rectangular-cylindrical cavity analysis

The new cavity design is presented in Figures 4 and 5 and is a modification of the dielectric-filled rectangular cavity operating in a pseudo T_{101} [1]. The new cavity is also operating in a pseudo TE_{101} mode but the fields in the cavity center resemble the pure mode except that they are more compressed in the lower part of the cavity. The position of the boundary line, between cavity cover and cavity base, is defined to be at zero current line on the cavity surface so as not to degrade cavity quality factor significantly. The dielectric is used to reduce the cavity size. Magnetic field is induced by a quarter wavelength microstrip-line loop installed in the cavity cover. The cavity is adjusted with a screw positioned at the electric field maximum.

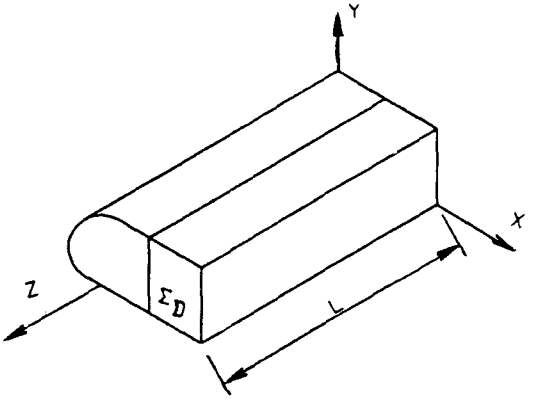


Fig 4. TE_{101} Rectangular-cylindrical cavity

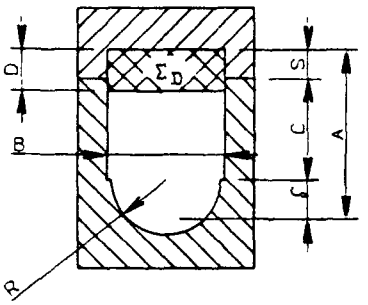


Fig 5. Cross-sectional view

If a dielectric slab is inserted as in Fig. 5, then the boundary conditions for a cavity operating in TE_{101} mode can be expressed as:

$$H_z: \cos[(2\pi/\lambda_{c1})(A-D)] = -M\cos[(2\pi/\lambda_{c2})D] \quad (4)$$

E_y:

$$j(\lambda_{c1}/2\pi)\sin[(2\pi/\lambda_{c1})(A-D)] = jM(\lambda_{c2}/2\pi)\sin[(2\pi/\lambda_{c2}D)] \quad (5)$$

where,

$$M = -\cos[(2\pi/\lambda_{c1})(A-D)]/\cos[(2\pi/\lambda_{c2}D)] \quad (6)$$

λ_{c1} - cavity base cutoff wavelength

λ_{c2} - cavity cover cutoff wavelength

A - cavity effective height

Dividing equation (5) by equation (4) we get an expression which allows to determine A and D for λ_{c1} and λ_{c2} [2]:

$$(\lambda_{c1}/2\pi)/tg[(2\pi/\lambda_{c1})(A-D)] = -(\lambda_{c2}/2\pi)/tg[(2\pi/\lambda_{c2})D] \quad (7)$$

$$\lambda_{c1} = \lambda_o/\sqrt{[1-(\lambda_o/\lambda)^2]} \quad (8)$$

$$\lambda_{c2} = \lambda_o/\sqrt{[\epsilon_D-(\lambda_o/\lambda)^2]} \quad (9)$$

where,
λ_o = 43.86mm (free-space wavelength for 6.84GHz)
ϵ_D = 5.78 in our case

If cells are inserted, then the effective wavelength in the cavity is a function of the cavity length:

$$\lambda = 2[L + T_z(\sqrt{\epsilon}-1)] \quad (10)$$

where,
ϵ = glass dielectric constant of the cells
T_z = glass total thickness along Z axis
The glass thickness, T_x is not taken into account here for simplicity of presentation. Equation (7) has been solved with computer and the resulting graphs are shown in Figures 6,7 and 8.

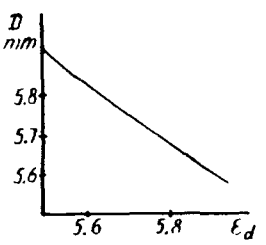


Fig 6. D vs. ϵ_D

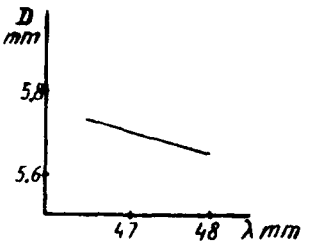


Fig 7. D vs. λ

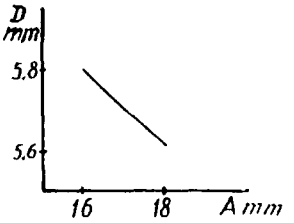


Fig 8. D vs. A

It is seen from these graphs that the only critical parameter is the dielectric constant of the material in the cavity cover.
The fields become:

$$H_{z1} = \cos[(2\pi/\lambda_{c1})x] \quad (11)$$

$$H_{z2} = -M\cos[(2\pi/\lambda_{c2})(A-x)] \quad (11a)$$

$$E_{y1} = (\lambda_{c1}/2\pi)\sin[(2\pi/\lambda_{c1})x] \quad (12)$$

$$E_{y2} = M(\lambda_{c2}/2\pi)\sin[(2\pi/\lambda_{c2})(A-x)] \quad (12a)$$

The fields are plotted in Figures 9 and 10.

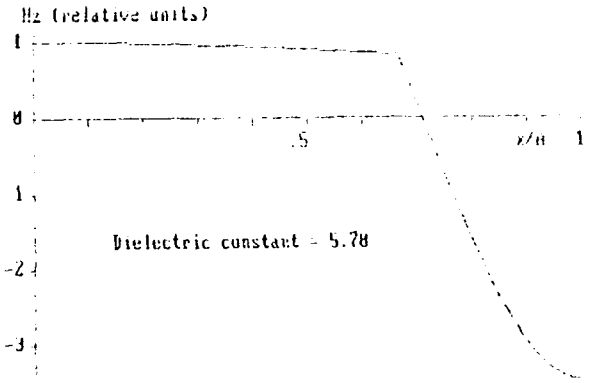


Fig 9. H_z distribution

Figure 9 shows that a cavity may be constructed from two pieces joined together in y-z plane at x/A=0.71 (H_z=0) without significant degradation of the quality factor.

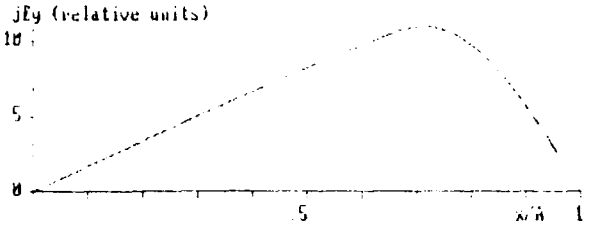


Fig 10. E_y distribution

As mentioned before the calculation is performed neglecting the presence of the glass distributed along the X axis and also neglecting a hole located on one of the X-Y planes also. Actually, the fields plotted in Figures 9,10 have been recalculated to find the exact position of H_z=0. This calculation is not presented here because it gives a small deviation from the presented results.

Microwave cavity performance

The measured Q of the new cavity including glass cells is found to be about 90. This is only two times less than theoretical limit for an empty rectangular cavity without holes. The achieved Q is two times higher than reported [1] for rectangular dielectric loaded cavity.

Another advantage of the new design is an absence of parasitic resonances at other frequencies. The input reflection coefficient S_{11} of the cavity is measured as zero that simplifies its matching with the SRD circuit. The measured characteristics of the cavity are shown in Figures 11 and 12.

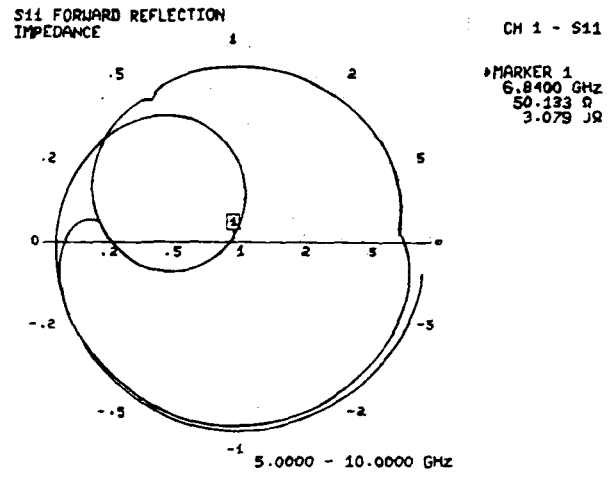


Fig 11. Smith chart of the rectangular-cylindrical cavity

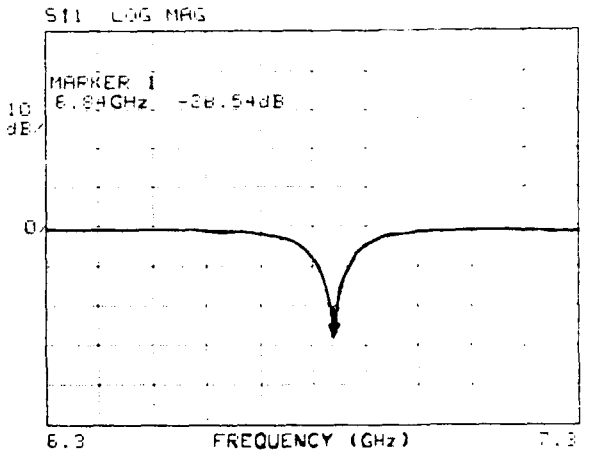


Fig 12. S_{11} of the rectangular-cylindrical cavity

SRD multiplier circuit

The SRD multiplier circuit is chosen to be an integrated part of the cavity to preclude reflections and consequent oscillation of the circuit, to reduce RF power consumption and to ease adjustment.

We can see from the Smith chart, Fig.11, that the cavity with a quarter wavelength microstrip loop has a parallel resonance at 6.84GHz. This leads to the conclusion that a voltage source should feed the cavity i.e. a SRD is to be in parallel with cavity input. The SRD-loop schematic diagram is shown in Figure 13.

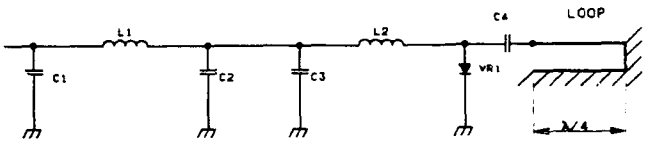


Fig 13. SRD-lood circuit

Capacitor C3 is a microwave single layer capacitor to short microwave current. This circuit can be redrawn as in Figure 14 for 6.84GHz, where C_j is SRD junction capacitance.

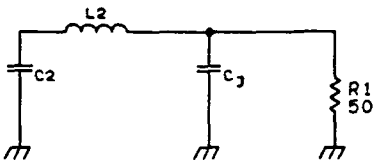


Fig 14. SRD-loop simplified schematic

Tank (L2 and Cj) is to resonate at about 2/3 of 6.84GHz. Inductor L2 is made like a piece of air strip line with variable inductance to compensate a large Cj tolerance. Quality factor of L2,Cj tank is about 1 and L2 trimming is not critical.

The following factors define input impedance of the circuit:

1. Inductance from input terminal to ground through the open SRD.
2. Resistance of the open SRD.
3. Conductive interval of a SRD which depends on:
   - 3.1 Input power.
   - 3.2 SRD bias current.
   - 3.3 SRD lifetime.
   - 3.4 Q of the cavity.

Tank C2,L2 is desirable to resonate at the input frequency that may require a relative high capacitor value. A high value of C2 together with its large size may cause parasitic resonances, so intermediate value and surface mount capacitor is chosen. Matching circuit C1,L1 is used to match 50ohm coaxial cable with the load. The circuit filters high order harmonics of 90MHz input signal and shorts high order harmonics travelling back to the coax. Inductor L1 is adjusted for 90MHz input frequency. Matching and SRD circuits are made implementing surface mount and "chip and wire" technologies.

The actual required input RF power is found to be about 0.1W which is the known threshold value for the SRD.

Summary of the performance

1. Input power: 20dBm.
2. Input impedance: 50ohm.
3. Performance with TF4010A standard:
   - 3.1 Allan variance at 1 sec is 8E-12.
   - 3.2 Phase noise = -133dBc at 100Hz offset (1Hz BW).

Conclusion

The small size of the cavity and its rectangular shape enabled the design of a very compact Physics-Package, only 87cc in volume which is currently used in TFL's compact rubidium standard.

The Physics-Package was successfully tested and is implemented in a new compact model of rubidium frequency standard, TF-4010A, that has been put recently into production by TFL. The new model inherits the main feature of its already modified predecessor TF-4000B and has the following main supplementary features:

1. 500cc volume.
2. Low power consumption and low warm up current.
3. Very low phase noise.

References

[1] H.E. Williams, T.M Kwon and T. Mellelland, Proc. 37th Annual Symposium on Frequency Control p.12, 1983

[2] N. Marcuvitz, Wavegide handbook, p.389, 1986

ULTRA-STABLE Hg^+ TRAPPED ION FREQUENCY STANDARD*

J. D. Prestage, R. L. Tjoelker, G. J. Dick, L. Maleki

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

We report the development of a fieldable frequency standard based on $^{199}Hg^+$ ions confined in a hybrid rf/dc linear ion trap. This trap permits storage of large numbers of ions with reduced susceptibility to the second-order Doppler effect caused by the rf confining fields. A 160 mHz wide atomic resonance line for the 40.5 GHz clock transition is used to steer the output of a 5 MHz crystal oscillator to obtain a stability of $2 \cdot 10^{-15}$ for 24,000 second averaging times. For longer averaging intervals measurements are limited by instabilities in available hydrogen masers frequency standards.

Introduction

Atomic frequency standards with high stability for averaging times τ longer than 1000 seconds are necessary for a variety of astrophysical measurements and long baseline spacecraft ranging experiments. The millisecond pulsar, PSR 1937+27, shows stability in its rotational period that exceeds that of all man-made clocks for averaging times longer than 6 months. Comparison of this pulsar period with an earth based clock of stability 1×10^{-15} over averaging periods of one year may show the effects of very low frequency (≈ 1 cycle per year) gravitational waves[1,2]. Spacecraft ranging measurements across the solar system would be improved with clocks whose stabilities exceed 1×10^{-15} for averaging times of 10^4 to 10^5 seconds. This clock performance would also improve gravitational wave searches in spacecraft ranging data. Another use for long term stable clocks in NASA's Deep Space Network would be in maintaining synchronization with Coordinated Universal Time (UTC).

The performance of microwave frequency standards in use today are summarized in Fig.1 [2,3]. For short term stability ($\tau < 150$ seconds) the Superconducting Cavity Maser in use at JPL shows stability as good as 2×10^{-15} for averaging times up to 2000 seconds[4]. Hydrogen masers are presently the most stable frequency standard for $150 < \tau < 30,000$ seconds and are the primary standard in use in JPL's Deep Space Tracking system. For averaging times greater than 6 months the millisecond Pulsar PSR 1937 + 21 exceeds the stability of international timekeeping abilities at a level of about 1-2 parts in 10^{-14} [3]. For $\tau > 10^6$ seconds the most stable clock yet measured is the Hg^+ ion standard based on 2×10^6 ions confined in an rf Paul trap[2,5]. The subject of the present paper is a linear ion trap based Hg^+ frequency standard now being tested at the Frequency Standards Laboratory at JPL. Stability measurements of the Hg^+ standard in two modes of operation are shown in Fig. 1, the lower curve using an H-maser local oscillator while the upper curve showing the performance of a 5 MHz crystal oscillator steered by the Hg^+ ion resonance. In both cases determination of the long term stability (i.e., $\tau \geq 25,000$ seconds) is

*This work represents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract sponsored by the National Aeronautics and Space Administration.

CH2965-2/91/0000-572 $1.00 © 1991 IEEE

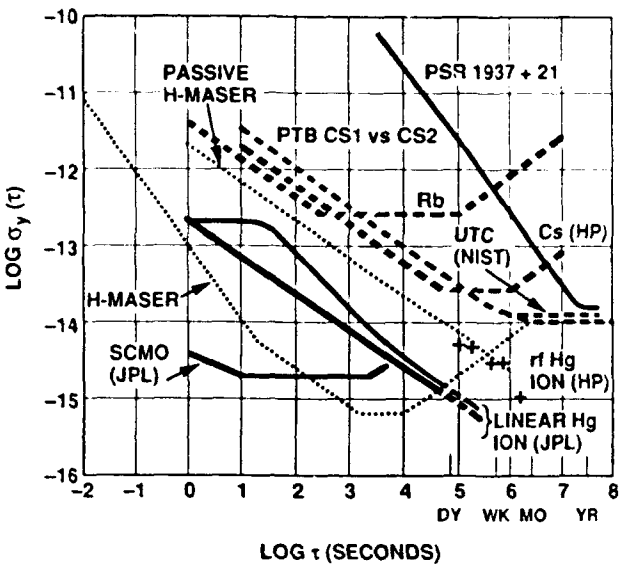


Figure 1: Fractional frequency stability of several precision frequency standards.

limited by reference H-maser instabilities.

Typically the largest source of frequency offset in standards based on ions confined in Paul traps stems from the motion of the ions caused by the trapping fields via the second-order Doppler or relativistic time dilation effect. Though increasing ion number will lead to increasing signal to noise in the measurement of the clock transition, the frequency offset also grows with the number of ions forcing a trade-off situation. Often fewer ions are trapped in order to reduce the (relatively) large offset and frequency instabilities which may result. For practical standards of high *stability* such frequency offsets can be tolerated to the extent they can be stabilized.

Linear Ion Trap

In a conventional hyperbolic Paul trap, ions are trapped around a node of the rf electric field at the center. The strength of the electric field and the resulting micromotion of the trapped particles grows linearly with distance from this node point. As ions are added the size of the ion cloud grows until the second order Doppler shift arising from the micromotion in the trapping field dominates the second order Doppler shift from the ion's thermal motion at room temperature. For typical operating conditions [6] a spherical cloud containing $2{\cdot}10^6$ mercury ions shows a 2nd order Doppler shift of $2{\cdot}10^{-12}$, a value some ten times larger than that due to thermal motion alone. In order to increase the number of stored ions with no corresponding increase in second-order Doppler shift from ion micromotion we have developed a hybrid rf/dc linear ion trap[7]. This trap confines ions along a line of nodes of the rf field effectively providing the same capability as a large number of hyperbolic traps together. The trapping force transverse to the line of nodes is generated by the ponderomotive force as in conventional Paul traps while the axial trapping force is provided by dc electric fields [7-11].

We can compare the second-order Doppler shift, $\Delta f/f$ generated by the trapping fields for a cloud of ions in these two types of traps, assuming that both traps are operated with the same rf trapping force as characterized by the ion secular frequency ω. If the same number of ions N, is held in both traps, the average distance from an ion to the node of the trapping field is greatly reduced in the linear trap. Since the distance from the node determines the magnitude of the rf trapping field, the 2nd order Doppler shift of an ion's clock frequency due to motion in the trapping field is reduced from that of a hyperbolic trap. If R_{sph} is the ion cloud radius in the hyperbolic trap and L is the ion cloud length in the linear trap, the Doppler shift in the two traps are related by[7]

$$\left(\frac{\Delta f}{f}\right)_{lin} = \frac{5}{3}\frac{R_{sph}}{L}\left(\frac{\Delta f}{f}\right)_{sph}. \qquad (1)$$

As more ions are added to the linear trap this shift will increase. It will equal that of the spherical ion cloud in a hyperbolic trap when

$$N_{lin} = \frac{3}{5}\frac{L}{R_{sph}}N_{sph}. \qquad (2)$$

Equations (1) and (2) are valid when the ion cloud radii, R_{lin} and R_{sph}, are much larger than the Debye length which is the characteristic plasma density decay length at the ion cloud edge, and is

about 0.4 mm for typical Hg^+ ion plasmas used in frequency standard work [6].

In addition to its larger ion storage capacity, the dependence of the second-order Doppler shift on trapping parameters in a linear trap is very different from that in a conventional Paul trap. For many ions in a Paul trap this shift is given by[6,7]

$$\left(\frac{\Delta f}{f}\right)_{sph} = -\frac{3}{10c^2}\left(\frac{N\omega q^2}{4\pi\epsilon_o m}\right)^{2/3} \quad (3)$$

where ω is the secular frequency for a spherical ion cloud containing N ions each with charge to mass ratio q/m. c is the speed of light and ϵ_o is the permittivity of free space. Ions in a long linear trap where end effects are negligible show a 2nd order Doppler shift from the motion generated by the rf confining field given by[7]

$$\left(\frac{\Delta f}{f}\right)_{lin} = -\left(\frac{q^2}{8\pi\epsilon_o mc^2}\right)\frac{N}{L} \quad (4)$$

where N/L is the linear number density of ions in the trap.

In contrast to the spherical case as described Eq.(3), this expression contains no dependence on trapping field strength, as characterized by ω, and depends only on the linear ion density N/L. If for example, the rf confining voltage increases and consequently the micromotion at a given point in space increases, the ion cloud radius will decrease so that the second-order Doppler shift (averaged over the cloud) from ion micromotion remains constant. Similar statements can be made about variations in any parameter that affects the radial confinement strength [8].

The sensitivity of the finite length linear trap to variations in radial trapping strength (characterized by ω) is [8]

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = -2\frac{R_t}{L}\frac{\delta\omega}{\omega}, \quad (5)$$

and to variations in endcap voltage is

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = 2\cdot\frac{R_t}{L}\frac{\delta V_e}{V_e} \quad (6)$$

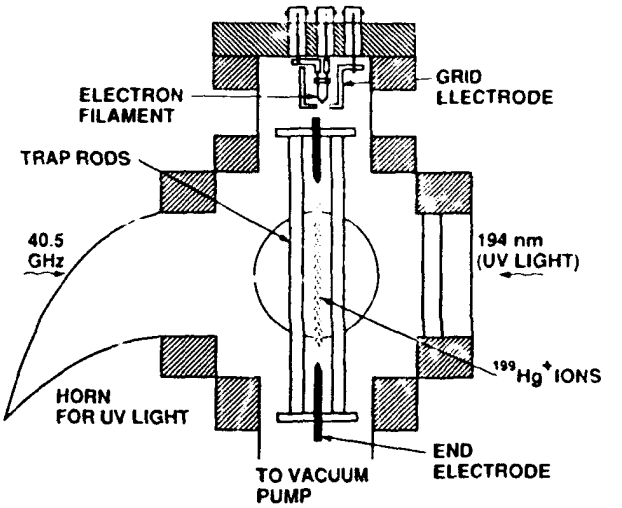


Figure 2: Linear ion trap assembly residing in its high vacuum enclosure. State selection light from the ^{202}Hg discharge lamp enters from the right, is focused onto the central 1/3 of the trap and is collected in the horn. Fluorescence from the trapped ions is collected in a direction normal to the page.

where R_t is the trap radius. The Paul trap shows a corresponding sensitivity to trap field strength variations

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{sph}}{\left(\frac{\Delta f}{f}\right)_{sph}} = -\frac{2}{3}\frac{\delta\omega}{\omega}. \quad (7)$$

A comparison of Eqs. (5) and (7) shows the linear trap based frequency standard to be less sensitive to variations in trapping field strength than the Paul trap by a factor of $3R_t/L$. For the trap described in the next section this factor is about 1/3.

Initial Operation with H-Maser Local Oscillator

Our linear trap is shown in Fig. 2. Ions are created inside the trap by an electron pulse along the trap axis which ionizes a neutral vapor of ^{199}Hg.

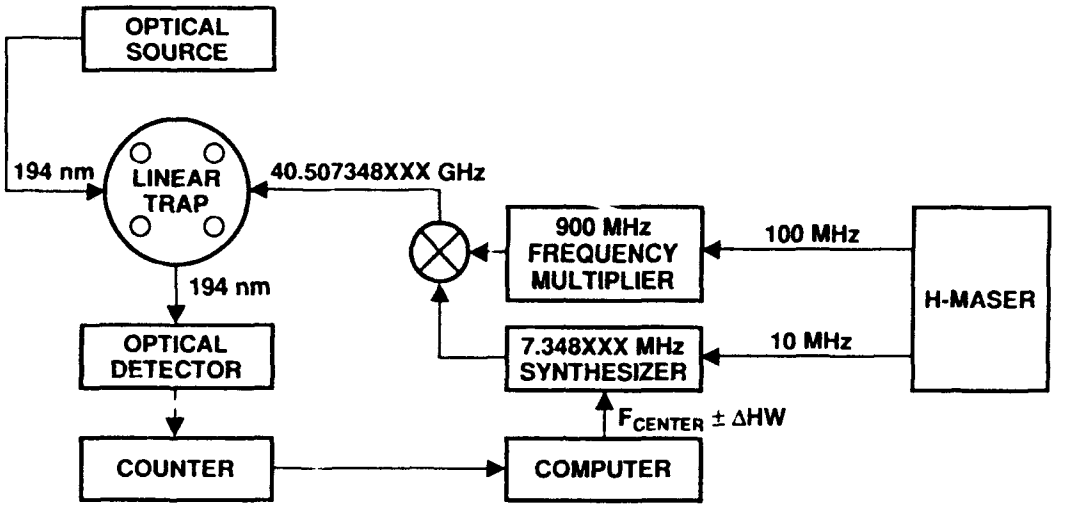


Figure 3: Schematic operation of the mercury ion trap using a hydrogen maser as the local oscillator.

A helium buffer gas (10^{-5} torr) collisionally cools the ions to near room temperature. Resonance radiation (194 nm) from a ^{202}Hg discharge lamp optically pumps the ions into the F=0 hyperfine level of the ground state[12]. This UV light is focused onto the central 1/3 of the 75 mm long ion cloud. The thermal motion of the ions along the length of the trap will carry all the ions through the light field so that pumping is complete in about 1.5 seconds for typical lamp intensities.

To minimize stray light entering the fluorescence collection system this state selection light is collected in a pyrex horn as shown in Fig. 2. Placement of the LaB_6 electron filament is also chosen to prevent light from entering the collection system. Its placement and relatively cool operating temperature together with good filtering of the state selection/interrogation UV light in the input optical system have allowed frequency standard operation without the use of a 194 nm optical bandpass filter in the collection arm. This triples data collection rates since such filters typically have about 30% transmission for 194 nm light.

Microwave radiation (40.5 GHz) propagates through the trap perpendicular to the trap axis thereby satisfying the Lamb-Dicke requirement that the spatial extent of the ion's motion along the direction of propagation of the microwave radiation be less than a wavelength. Radiation enters the trap region through the pyrex horn (see Fig. 2) and propagates in the opposite direction to the UV state selection/interrogation light. This allows collection of atomic fluorescence in both directions perpendicular to the incident pumping light. For the resonance and stability data shown in this paper, fluorescence was collected in only one of these two directions.

At the present time no thermal regulation is incorporated into the ion standard itself. However, for all stability measurements described the ion trap standard together with its support electronics were housed in an environmentally controlled test chamber where temperature variations were regulated to approximately ± 0.05°C. This level of regulation is much less demanding than that required, e.g., by hydrogen masers.

We have used the technique of successive oscillatory fields[13] to probe the approximately 40.5 GHz hyperfine clock transition in $^{199}Hg^+$ ions confined to the linear trap described above. In the

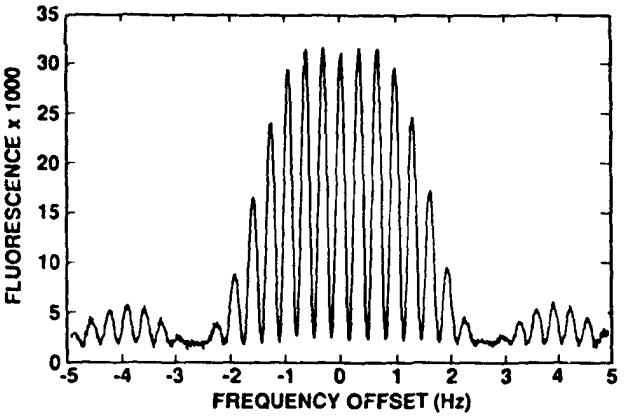


Figure 4: $^{199}Hg^{+}$ clock transition as measured with method of successive oscillatory fields. This line shape results from two 0.35 second microwave pulses separated by a 2.5 second free precession period. The central line is about 160 mHz wide and the data shown is an average of ten scans.

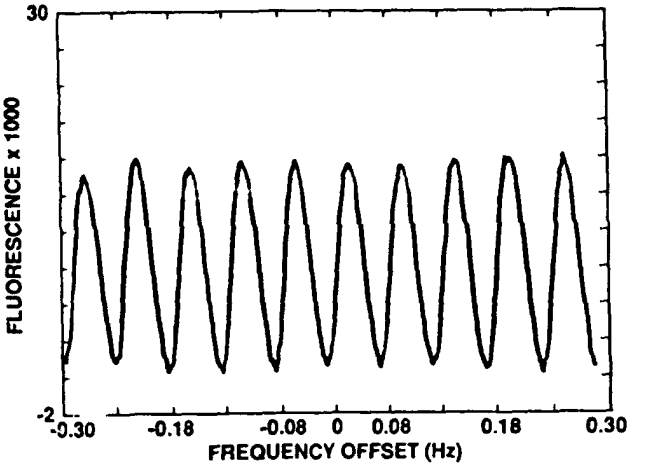


Figure 5: High Q Ramsey fringes resulting from two microwave pulses of 0.275 seconds separated by a 16 second free precession period. The fringes are about 30 mHz wide and the data shown is an average of four scans.

initial measurements the 40.5 GHz signal is derived from an active Hydrogen maser frequency source as shown in Fig. 3. A representative resonance line of the $^{199}Hg^{+}$ clock transition is shown in Fig. 4. State selection and interrogation is accomplished during the 1.5 seconds following the lamp turn on. It is necessary to switch the UV state selection/interrogation light level to near zero during the microwave interrogation period to prevent light shifts and broadening of the clock transition. A background light level of about 300,000 per 1.5 second collection period has been subtracted to generate the resonance shown. The successive oscillatory field pulses consist of two 0.35 second microwave pulses separated by 2.5 second free precession period. The data shown is an average of ten scans with a 10 mHz frequency step size.

The central portion of the narrowest resonance lines yet obtained with this apparatus are shown in Fig. 5. This line is derived from two 0.275 second pulses separated by a 16.5 second free precession period[14]. The linewidth of 30 mHz represents a line Q of over 1×10^{12} on the 40.5 GHz transition. The data displayed is an average of 4 full scans and, to our knowledge, is the highest Q transition ever measured in a microwave atomic transition.

To determine the frequency stability of the overall system of ions, trap, microwave source, etc., we have locked the output frequency of the 40.5 GHz source to the central peak of the 160 mHz resonance in a sequence of 16384 frequency measurements. The time required for each measurement is about 7 seconds and the loop response time was 5 measurement cycles. By averaging the frequencies of 2^N adjacent measurements (N=1,2,..,13) we form the modified Allan variance shown in Fig. 6. Performance of the linear ion trap based Hg^{+} standard measured in this manner is $2 \times 10^{-13}/\sqrt{\tau}$ for averaging times $\tau \leq 20,000$ seconds beyond which H-maser frequency instabilities limit the measurement.

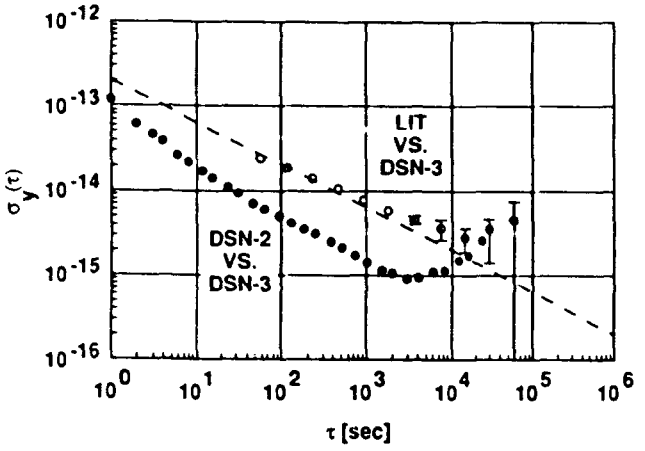


Figure 6: Performance of the Hg^+ system when the local oscillator is a hydrogen maser. The fractional frequency stability of the ion trap system is measured against an LO maser. Also shown is the stability comparison of the LO maser to a another H-maser.

Operation with a Crystal Oscillator

In this section we describe the long term stability achieved when a commercial (Oscilloquartz 8600B no. 104) 5 MHz quartz crystal oscillator of superior performance is servoed to the 160 mHz Hg^+ ion resonance in Fig. 4. The schematic of this system is shown in Fig. 7. Using the JPL Frequency Standards Lab measurement system we have carried out a frequency stability measurement of the Hg^+ steered crystal using two H-masers in the measurement system (designated DSN2 and DSN3) as references.

The results of this 63 hour comparison are shown in Fig. 8 (a) and (b). The first figure shows Allan deviation of the Hg^+ steered crystal as measured against DSN2 (upper trace) and against DSN3 (lower trace). For averaging times less than 10 seconds the Allan deviation is that of the unsteered crystal oscillator since the loop attack time is about 10 seconds. For averaging times shorter than about 13,000 seconds the Hg^+ standard shows the same stability independent of reference maser. For $\tau > 20,000$ seconds the Allan deviation of the Hg^+ vs. DSN2 is the same as that for DSN2 vs. DSN3 (see Fig. 8 (b)) indicating that DSN2 has the limiting performance of the three standards under test for this averaging time period. The lower trace of Fig. 8(a) shows a Hg^+ standard stability of 2×10^{-15} for $\tau = 24,000$ seconds beyond which instabilities in DSN3 probably limit the measurement . A second Hg^+ ion standard is now under construction which will enable stability measurements beyond 24,000 seconds.

Local Oscillator Considerations

Fluctuations in the local oscillator (L.O.) limit performance of a trapped ion standard in two ways. As discussed above, slow variations of the L.O. are compensated by action of the servo system. The effectiveness of this compensation increases with the measuring time, so that for longer measuring times performance approaches the $1/\sqrt{\tau}$ dependence which is characteristic of passive standards. This behavior is clearly shown in Figure 8.

However, fast fluctuations in the L.O. also degrade performance of the standard by an effect that adds to the (white) fluctuation of the photon count from measurement to measurement[15-17]. This limitation continues to the longest times, having the same $1/\sqrt{\tau}$ dependence on measuring time τ as the inherent performance of the standard itself. The cause of this effect can be traced to a time varying sensitivity to L.O. frequency which is inherent in the interrogation process. This limitation was evaluated in a recent calculation for several types of sequentially interrogated passive standards[17].

Using this same methodology, we have calculated the L.O.-induced performance degradation for our particular interrogation scenario. Here, two RF pulses of 0.35 second length are separated by a delay of 2.5 seconds. A dead time of 3.8 seconds follows to give the cycle time of $t_c = 7$

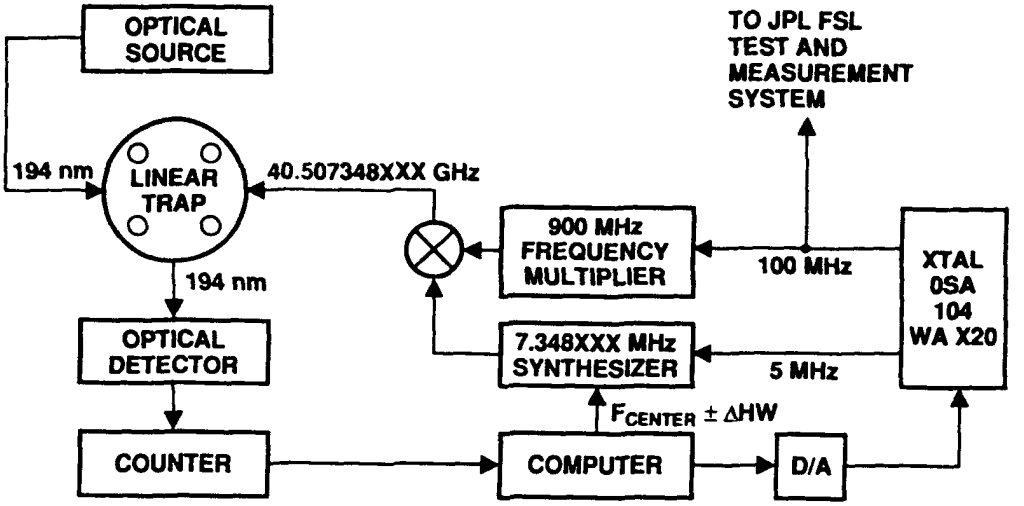


Figure 7: Schematic operation of the mercury ion trap to steer a quartz oscillator.

seconds. Our quartz L.O shows flicker frequency noise with an approximately flat Allan deviation of 1.5×10^{-13}. For this L.O. we calculate a contribution to the Allan deviation of the trapped ion source of $2.6 \times 10^{-13}/\sqrt{\tau}$. This value is slightly larger than the $1.96 \times 10^{-13}/\sqrt{\tau}$ due to random fluctuations in the photon count, again based on actual operating conditions. The two contributions can be combined in to give a limiting stability of $3.25 \times 10^{-13}/\sqrt{\tau}$ which is plotted along with the data in Figure 8. The data confirm this analysis by a very close approach to the line for measuring times $\tau > 10^3$ seconds.

It is worth mentioning that both L.O. and intrinsic statistical performance limitations may be reduced by increasing the interrogation time, as long as the dead time is not increased. The implicitly higher Q and reduced (relative) dead time would result in a comparable improvement for each of the two contributions. For example, an increase in the interrogation time by 4× would reduce the limiting $1/\sqrt{\tau}$ Allan deviation by half.

However it is clear that as performance improves to make possible trapped ion performance in the low-$10^{-14}/\sqrt{\tau}$ range, quartz crystal L.O. performance will not be sufficient in itself to keep pace. In this case a cryogenic L.O. with 10^{-14} type stability for short times could be used, or a quartz L.O. with alternatively interrogated traps to give a uniform servo sensitivity with time[17].

Sources of Frequency Instability

While short term performance of the ion trap standard is determined by signal to noise ratio and line Q of the clock transition resonance, the long term stability is determined by the sensitivity of the atomic system to changes in environmental and operating parameters and on our ability to control and stabilize such parameters. The largest measured offsets of the Hg^+ clock transition frequency under our present operating conditions are shown in Fig. 9. If these offsets are stable the device serves as a practical frequency standard. In addition, when these offsets are quantified, the standard can reach accuracy well below the magnitude

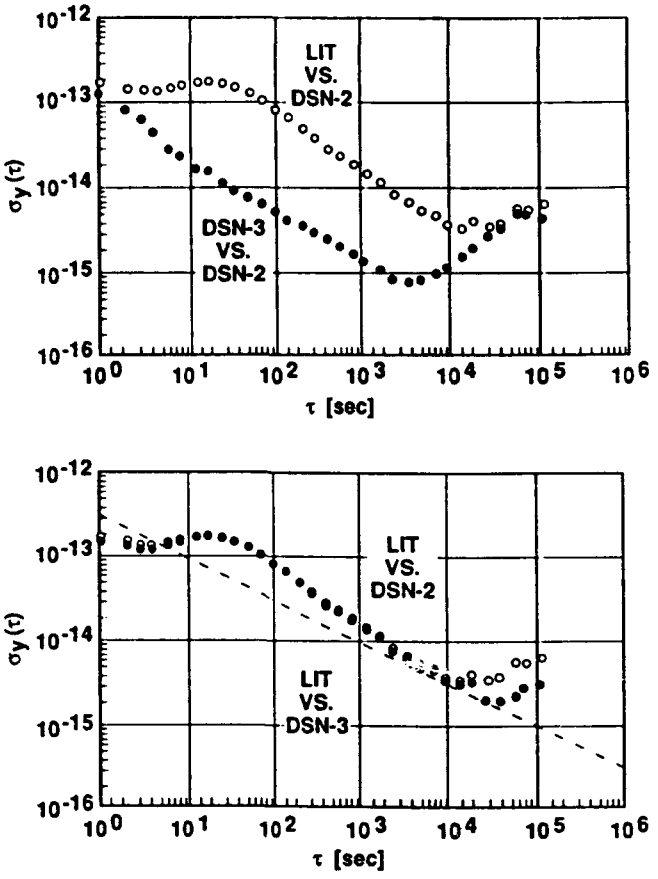


Figure 8: Performance of the Hg^+ system when the LO is a quartz oscillator. (a) Fractional frequency stability compared to the hydrogen maser DSN3 and maser comparison between DSN2 and DSN3. (b) The ion trap standard compared to each maser independently distinguishing the different performance of the two masers for times $\tau \geq 10,000$ seconds. The dashed line represents calculated performance based on actual operating conditions.

The second-order Doppler shift from ion motion driven by the trapping field is determined by measuring clock frequency as ion number, N, decays during the approximately 2000 second ion storage time for our trapping conditions. The frequency offset between our current operation with about 5×10^7 ions and very few ions where trapping field shifts are minimized is $1.5 - 2.0 \times 10^{-12}$. No active ion number stabilization was used in any of the measurements described here though some stabilization is acheived by filling the trap to saturation level for the given DC voltage on the end electrodes. Saturation occurs so long as the net rate of ion generation in the trap is much higher than the ion loss rate. This present procedure is sufficent to maintain ion number stability of at least 0.1% over the 24,000 second averaging time required to reach 2×10^{-15} stability.

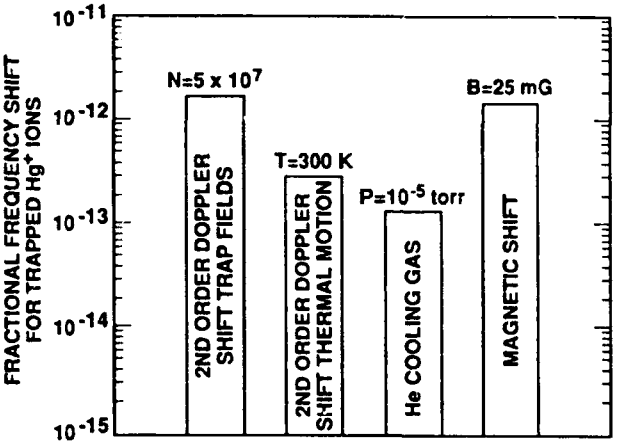


Figure 9: The largest measured perturbations to the $^{199}Hg^+$ clock transition in the linear trap under typical operating conditions.

The fractional sensitivity of the $^{199}Hg^+$ clock transition to magnetic field variations is nearly 1000 times less than that of hydrogen at the same operating field. For the present measurements the field was set at 3.5μT (35 mG). At this operating field the unshielded atomic sensitivity is 1.7×10^{-13} per mG. To reach 1×10^{-16} frequency stability the current in the Helmholtz field bias coils must be stable to 2×10^{-5}. To prevent ambient field disturbances from influencing the ion frequency the trap region is surrounded by a triple layer magnetic shield of shielding factor 10,000. With this shielding factor a 5 mG ambient field change would lead to a 1×10^{-16} shift in the atomic resonance frequency. Magnetic gradients must also be minimized in order to reach the highest frequency stability since gradients over the ion cloud can degrade atomic coherence and limit line Q. A par-

tial solution has been to operate the standard at relatively high field settings (35 mG) but this also increases the atomic sensitivity to fluctuations in ambient field.

The fractional temperature sensitivity of the complete system is measured to be less than $10^{-14}/°C$ which probably comes about via in increase in neutral mercury vapor and a consequent heating and/or ion cloud radius increase. The lamp and its housing must also be temperature controlled as the brightness is highly depend on temperature.

Long term stability requires controlling these variables to high precision. Since the mercury atom is in general less sensitive to environmental changes than other atoms used in frequency standards an order of magnitude improvement may be obtained in long term stability with the same level of control as existing standards.

Conclusions

By steering a 5 MHz crystal oscillator to a 160 mHz atomic resonance ($Q = 3.3 \cdot 10^{11}$) we have measured performance of 2×10^{-15} for $\tau = 24,000$ seconds limited only by the stability of available reference hydrogen masers. Line Q's as high as 1.3×10^{12} have been measured [14], indicating consequent performance for this trap as good as $5 \times 10^{-14}/\sqrt{\tau}$ for $\tau > 150$ seconds.

Acknowledgment

We wish to thank and Randy Berends for making ^{202}Hg lamps and Bill Diener, Albert Kirk and Roland Taylor for assistance with the Frequency Standards Laboratory Test and Measurement System.

References

[1] M. M. Davis, J. H. Taylor, J. M. Weisberg,and D. C. Backer, "High-precision timing observations of the millisecond pulsar PSR1937+21", *Nature* **315**, 547-550, 1985.

[2] D. W. Allan, M. A. Weiss, and T. K. Peppler, "In Search of the Best Clock", *IEEE Trans. Instrum. Meas.* **38** 624-630, 1989.

[3] L. A. Rawley, J. H. Taylor, M. M. Davis, D. W. Allan, "Millisecond Pulsar PSR 1937+21: A Highly Stable Clock",*Science* **238** 761-765, 1987.

[4] Private communication Rabi Wang. See also R. T. Wang and G. J. Dick, "Improved Performance of the Superconducting Cavity Maser at Short Measuring Times", *Proc. 44th Ann. Symp. on Freq. Control* (IEEE Cat. No. 87-654207), 89-93, 1990.

[5] L. S. Cutler, R. P. Giffard, P. J. Wheeler, and G. M. R. Winkler, "Initial Operational Experience with a Mercury Ion Storage Frequency Standard", in *Proc. 41st Ann. Symp. Freq. Control, IEEE Cat. No. 87CH2427-3*, 12-19, 1987.

[6] L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an RF Quadrupole Trap", *Appl. Phys. B* **36**, 137-142, 1985.

[7] J. D. Prestage, G. J. Dick, L. Maleki, "New Ion Trap for Frequency Standard Applications", *J. Appl. Phys.* **66**, No. 3, 1013-1017, August 1989.

[8] J. D. Prestage, G. R. Janik, G. J. Dick, and L. Maleki, "Linear Ion Trap for Second-Order Doppler Shift Reduction in Frequency Standard Applications", *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, **37**, 535-542, 1990.

[9] G. R. Janik, J. D. Prestage, and L. Maleki, "Simple Analytic Potentials for Linear Ion Traps", *J. Appl. Phys.* **67**, No. 10, 6050-6055, May 1990.

[10] H. G. Dehmelt, "Introduction to the Session on Trapped Ions," *Proc. 4th Symp. Frequency Standards and Metrology*, 286, 1989.

[11] D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano, D. J. Heinzen, S. L. Gilbert, C. H. Manney, and M. G. Raizen, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions", *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.* **37**, 515–523, 1990.

[12] F. G. Major and G. Werth, "Magnetic Hyperfine Spectrum of Isolated $^{199}Hg^{+}$ Ions", Appl. Phys. **15**, 201-208, 1981.

[13] N. F. Ramsey, *Molecular Beams* (Oxford University Press, Oxford, 1956).

[14] J. D. Prestage, G. J. Dick, and L. Maleki, "Linear Ion Trap Based Atomic Frequency Standard", *IEEE Trans. Instr. Meas.*, April, 1991.

[15] G. J. Dick, "Calculation of Trapped Ion Local Oscillator Requirements," *Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 133–146 1988.

[16] C. Audoin, V. Candelier, and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards," *IEEE Trans. Instr. Meas.*, April, 1991.

[17] G. J. Dick, J. D. Prestage, C. A. Greenhall, and L. Maleki, "Local Oscillator Induced Degradation of Medium-Term Stability in Passive Atomic Frequency Standards," to be published in *Proceedings, 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, Dec. 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HYDROGEN MASER PERFORMANCE AT THE UNITED STATES NAVAL OBSERVATORY AND THE NAVAL RESEARCH LABORATORY

E. Powers and A. Gifford
Naval Research Laboratory
Washington D.C. 20375-5000
P. Wheeler
United States Naval Observatory
Washington D.C. 20391-5100

Abstract

The United States Naval Observatory(USNO) has been using hydrogen masers for over six years as part of the USNO Master Clock. The Naval Research Laboratory(NRL) has also been using masers references as part of the NRL Clock testing facility since September 1985. The masers reported on in this paper will include the Smithsonian Astrophysical Observatory VLG-11, VLG-12 and masers from the Sigma Tau Standards Corporation. This paper will describe the operation of the masers at USNO and NRL including stability, aging rates, and reliability.

Introduction

The United States Naval Observatory(USNO) is responsible for time keeping within the Department of Defense(DOD). The USNO Master Clock is a system of several hydrogen masers, cesium beam and mercury ion frequency standards. USNO has up-graded its time scale and in recent years has added an NRL designed 48 channel long term clock measurement system[1], a new full Kalman filter based clock ensemble[2], three trapped Mercury Ion devices[3] and fourteen Hydrogen masers[4, 5].

NRL has been in the forefront of advanced clock development for a variety of DOD systems including the USNO Master Clock upgrade program and the Global Positioning System (GPS) atomic clock development program[6]. Hydrogen masers serve as the primary frequency standard in the NRL clock test facility. Starting from the early 1970's NRL purchased four SAO VLG series masers and two Sigma Tau masers to use as references. The two Sigma Tau masers (NRL-N1 and NRL-N2) and one of the SAO VLG-11 (P-12) series masers were chosen to make up NRL's primary frequency reference ensemble with the other three SAO VLG series masers, including the original VLG-10 P-0, are located at other test sites.

Performance of USNO masers

The new USNO long term clock measurement system was designed to take data in a highly automated fashion, and has proved to be ideal for measurement of a large number of clocks over very long periods of time. This 48 channel dual-mixer system records data every hour with a measurement noise of less than five picoseconds. The computers running the measurement software are available over an ethernet type UNIX network, which makes data handling very easy.

At present USNO is operating twelve Hydrogen Masers with two more expected to arrive later in 1991. Of the twelve masers already at USNO, two are SAO VLG-11 series masers (P-18 and P-19), four are SAO VLG-12 (P-22, 23, 24 and P-25) series masers and six are Sigma Tau masers (N1, N2, N3, N4, N5 and N6). Data on two of the Sigma Tau masers, N1 and N6 was not available for this paper because they have been sent to other USNO sites.

The performance of eight of the USNO masers taken over a period of 264 days was measured against the USNO Master Clock. This 264 day span covers MJD 48104 to MJD 48368 (August 1990 - April 1991). The drift removed frequency stability of each combination of clock pairs was then calculated and used in an N-Cornered hat estimation[7, 8], so the true stability of each clock could be found. Table 1 shows the N-Corner Hat estimates of the frequency stability of the USNO masers. Masers denoted as P are built by SAO and masers with a N prefix were built by Sigma Tau. During a one month period starting in March 1991 P-23 exhibited a 8×10^{-15} frequency shift which distorts the stability numbers for this clock.

N2, and P18 are being steered to the USNO Master Clock by use of the high resolution synthesizer built into both types of masers. Because of this, only a shorter span of data was available for the calculation of the stability of masers N2 and

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

N-CORNER HAT FOR A 263 DAY PERIOD COVERING MJD(8104-8367) WITH DRIFT REMOVED.

	1 DAY	2 DAY	4 DAY	8 DAY	16 DAY	32 DAY
N3	1.4	1.5	2.0	2.7	3.8	5.7
N4	1.6	2.1	2.7	3.4	4.9	5.6
N5	1.2	0.85	.75	0.8	0.73	0.9
P19	1.95	2.0	2.05	1.8	1.9	2.2
P22	3.1	3.6	4.4	4.8	6.2	8.7
P23*	5.7	6.9	9.0	11.6	13.0	14.5
P24	0.6	0.97	0.82	0.92	1.05	1.2
P25	1.2	1.4	1.83	3.3	5.49	7.5
N2	1.1	1.36	1.24	2.3		
P18	1.9	2.4	2.7	2.8		

NOTE: All values are times 1 x10 $^{-15}$
*8 x10 $^{-14}$ frequency shift greatly influenced data on P-23

P18. The steer amounts were removed from the data and a 3-Corner hat estimate was calculated. This data is also shown in Table 1.

The drift rate was removed by using a simple linear fit to the frequency data which worked well for most pairs of clocks. P-25 had a logarithmic shape to its frequency drift, with the drift rate slowing by almost 50 %. Drift rates for each clocks are shown below versus BIPM.

N3 - BIPM NONE
N4 - BIPM NONE
N5 - BIPM +0.25 x10^{-15}/DAY
P18 - BIPM +0.29 x10^{-15}/DAY
P23 - BIPM +0.99 x10^{-15}/DAY
P19 - BIPM +1.09 x10^{-15}/DAY
N2 - BIPM +1.26 x10^{-15}/DAY
P24 - BIPM +1.82 x10^{-15}/DAY
P22 - BIPM +2.81 x10^{-15}/DAY
P25 - BIPM +3.37 x10^{-15}/DAY

NOTE: NONE IS < (+/-) 1 x10^{-16}/DAY DRIFT

PERFORMANCE OF NRL MASERS

The Sigma Tau corporation delivered two Hydrogen masers (NRL-N1 and NRL-N2) to NRL in the September 1985. These masers have been in almost continuous operation since the day they arrived having been shut off only once because of dead back-up batteries. No adjustments have been made to any of the masers settings for almost six years.

NRL and USNO installed a cesium clock at a local television station (WTTG channel 5) that is used to produce the 77 Mhz carrier frequency of this station[9]. Both NRL and USNO mix their local references with the received television carrier to produce a 2250 Hz beat frequency. USNO feeds their beat frequency to NRL over a dedicated telephone line and NRL compares the phase of these two 2250 Hz signals. The cesium clock at WTTG will cancel, and NRL is left with the phase difference between NRL-reference (NRL-N1) and the USNO Master Clock. The short term noise in this method is good but signal drop outs can cause phase breaks to occur.

Using this television carrier phase comparison system to measure the long term frequency drift in the NRL reference shows that NRL-N1 had a drift rate of less than 5 x10^{-17}/day for almost six years versus BIPM.

NRL-N2 exhibited a non-linear frequency drift rate during the first two and half years of operation at NRL, with the drift rate slowly increasing. This drift rate changed to a constant rate of +4 x10^{-15}/day for the next two years. Approximately one year ago the drift rate shifted again to a constant rate +3.5 x10^{-15}/day and remained at that rate.

The drift removed frequency stability of NRL-N1 versus NRL-N2 is shown below. Because these two clock are similar one may assume that one of the two clocks are at least the square root of two better than shown in below. NRL-N1 versus BIPM stability estimates for 100 day and 200 day

sample times are also shown below.

TYPICAL FREQUENCY STABILITY NRL-N1 VERSUS NRL-N2 (DRIFT REMOVED)

1 sec	2.5×10^{-13}
10 sec	3.5×10^{-14}
100 sec	6.0×10^{-15}
1 hour	5.0×10^{-15}
2 hour	4.8×10^{-15}
4 hour	4.6×10^{-15}
8 hour	3.6×10^{-15}
1 day	2.8×10^{-15}
2 day	2.1×10^{-15}
5 day	1.2×10^{-15}
10 day	1.2×10^{-15}

FREQUENCY STABILITY ESTIMATE OF NRL-N1 VERSUS BIPM (NOT DRIFT REMOVED)

100 day	4.7×10^{-15}
200 day	5.6×10^{-15}

RELIABILITY

The reliability of both the VLG-12 and Sigma Tau masers has been excellent. The only systematic problem was in the older SAO VLG-11 series maser's vacuum system. The VLG-11 masers require replacement of the ion pump every two or three years. The SAO VLG-12 series masers high vacuum ion pump system is improved and should last much longer. Also several VLG-11 masers have suffered receivers failures due to an amplifier chip running hot and burning out. This design error has been corrected and all of USNO masers have been modified to correct this error.

All of the USNO Sigma Tau masers needed their quartz oscillators replaced. After consulting the company that manufactures this quartz oscillator, a design flaw was discovered. The company changed their design and future problems are not expected. It is expected that the useful lifetimes of both the VLG-12 and Sigma Tau masers will be greater than 12 years.

CONCLUSION

The USNO clocks with the best frequency stability over the time period reported on in this paper were N5, P19 and P24. N5 and P24 long term drift removed frequency stability at 32 day sample times were less than 1.2×10^{-15} and the stability of P-19 was 2.2×10^{-15}.

The clocks with the least drift were N3, N4, N5 and P18. These drift rates were less than 3.0×10^{-16}/DAY versus BIPM. Over a period of less than one year measuring drift rates smaller than this is very difficult because of measurement uncertainties between BIPM and USNO.

NRL master frequency reference NRL-N1 has shown a net long term frequency drift versus BIPM of less than 5×10^{-17}/day over a six year period.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Gernot Winkler and Anthony Kubik of USNO for their contributions to this paper. We would like also to thank, WTTG channel 5. And finally the authors would like to thank Mike Golding from NRL for his help in the N-corner calculations.

REFERENCES

1. S.R. Stein and G.A. Gifford, "Software for Two Measurement Systems", Proceedings of the 38th Annual Symposium on Frequency Control, June 1984, pp. 483-486.

2. G.A. Gifford and P. Wheeler, "Report on the Master Clock Upgrade Program at USNO", Proceedings of the 43th Annual Symposium on Frequency Control, June 1989, pp. 158-161.

3. L.S. Cutler, R.P. Giffard, P.J. Wheeler, and G.R. Winkler, "Initial Operation Experience with a Mercury Ion Storage Frequency Standard", Proceedings of the 41h Annual Symposium on Frequency Control, June 1987 pp. 12-15

4. G.A. Gifford, J.D. White and H.E. Peters, "Hydrogen Maser Research and Development at Sigma Tau Standards and Test of Sigma Tau Standards Masers at the Naval Research Laboratory", Proceedings of the 17th Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, December 1985, pp. 105-128

5. R.F.C. Vessot, E.M. Mattison, W.J. Klepczynski, I.F. Silvera, H.P. Godfried and R.L. Walsworth, Jr. "Results of Two Years of Hydrogen Maser Clock Operation at the U.S. Naval Observatory and Ongoing Research at the Harvard-Smithsonian center for Astrophysics", Proceedings of the 17th Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting, December 1985,

pp. 413-431.

6. R.L. Beard, F. Danzy, E.D. Powers and J.D. White, "Test and Evaluation Methods at the NRL Clock Test Facility", Proceedings of the 43th Annual Symposium on Frequency Control, June 1989, pp. 275-288.

7. J.E. Gray and D.W. Allan, A Method for "Estimating the Frequency Stability of an Oscillator", Proceedings of the 28th Annual Symposium on Frequency Control, 1974, pp. 243-246.

8. M.E. Weiss and D.W. Allan, "Using Multiple Reference Stations to Separate the Variances for Noise Components in the Global Positioning System", Proceedings of the 40th Annual Symposium on Frequency Control, June 1986, pp. 394-404.

9. A. Gabry, G. Faucheron, B. Dubouis and P. Petit, "Distant Comparison of Stable Frequency Standards by Means of the Transmission of a Beat Note Between the Carrier of a TV Broadcast Signal and a Frequency Synthesized from the Frequency Standards", Proceedings of the 31th Annual Symposium on Frequency Control, 1977, pp. 499-502.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

BAROMETRIC PRESSURE–INDUCED FREQUENCY OFFSETS IN HYDROGEN MASERS

Richard A. Dragonette and Joseph J. Suter
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723

Abstract

Barometric pressure variations have been found to induce significant frequency shifts in atomic hydrogen masers by changing the compressive forces exerted on the maser's electromagnetic cavity. The observed frequency shifts, which can be as large as 1 part in 10^{-13}, track rising and falling pressures symmetrically. The barometric pressure fluctuations control the maser's aging rate for periods as short as a few hours and as long as 2 weeks.

Introduction

Atomic hydrogen masers, which exhibit a frequency stability of parts in 10^{-15}, have proven to be the most stable frequency standards available for sampling periods up to 10^5 seconds [1, 2]. The hydrogen maser derives its stable output from electronic observation of the hyperfine transition of atomic hydrogen, occurring at a frequency of 1.4204057518 GHz [3]. The narrow microwave resonance line generated by the hyperfine transition is observed using an electromagnetic resonant cavity operating in the TE_{011} mode. Using superheterodyne techniques, a 5 MHz crystal oscillator is phase locked to the signal coupled from the resonant cavity. The crystal oscillator output reflects the long-term stability (greater than 100 s) of the maser as determined by its large atomic line Q (1.33×10^9).

Changing environmental conditions are expected to have an adverse effect on the performance of the hydrogen maser. The temperature and humidity affect the electronic circuitry used to tune the microwave resonant cavity to the hyperfine transition frequency of atomic hydrogen. Barometric pressure changes compress the vacuum enclosure surrounding the electromagnetic resonant cavity, which distorts the electromagnetic field inside the cavity and changes the maser's resonance frequency. The barometric pressure–induced effects are generally difficult to calibrate because they are often masked by frequency shifts caused by ambient temperature variations. This complication was eliminated in the current series of atomic hydrogen masers by enclosing the resonant cavity within a thermally isolating quartz sleeve. Constructing the outer wall of the resonant cavity from quartz has reduced the temperature sensitivity of the maser more than an order of magnitude over the same design constructed out of aluminum.

Experimental Procedure

The data presented in this paper were recorded by a data acquisition system since 1 March 1990. The system continuously monitors the performance of three hydrogen masers as well as the environmental conditions of the laboratory in which the masers are maintained. Every 5 minutes the system records the values of all the environmental conditions and calculates an updated aging rate for each monitored maser. The 5-minute interval raw data are averaged by the system controller to produce daily and weekly plots of the masers' aging rates and the environmental conditions as functions of time.

The data acquisition system is composed of the environmental monitoring subsystem, the maser aging rate measurement subsystem, and the system controller computer. The system is nearly redundant, for it incorporates two independent computer controllers, an entirely redundant maser aging rate measuring system, and redundant environmental sensors common to both system controllers.

The environmental monitoring subsystem consists of an array of traceable thermometers meeting National Institute of Standards and Technology (NIST) standards, relative humidity sensors, dew point sensors, and barometers with single measurement accuracies of 0.1 °C, 2%, 0.5 °C, and 0.01 inch Hg, respectively. The DC output from each of the environmental sensors is fed into a multiplexer. At 5-minute intervals, the system controller uses a digital voltmeter to scan the multiplexer inputs and record the measured environmental conditions.

The maser aging rate measurement subsystem measures the aging rate of one maser relative to another with a resolution of a few parts in 10^{-15}. The two hydrogen masers monitored in this investigation have been designated NRB1 and NRX. The aging rate of NRB1 relative

CH2965-2/91/0000-586 $1.00 © 1991 IEEE

to NRX is measured by calculating the time rate of change of the phase of the 5 MHz output of NRB1 relative to that of NRX. The measurement system makes this calculation as follows.

As shown in Figure 1, 10 Hz phase signals are derived by multiplying the 5 MHz output from each maser to 200 MHz and mixing this signal with a −10 Hz offset 200 MHz (199,999,990 Hz) reference signal derived from a third maser. The −10 Hz offset 200 MHz signal is generated by using a frequency synthesizer to produce a 5 MHz −0.25 Hz (4,999,999.75 Hz) signal from the 5 MHz output of a third maser. This signal is then multiplied by 40, producing the −10 Hz offset 200 MHz reference. The 200 MHz signals from both NRB1 and NRX are separately mixed with the common −10 Hz offset 200 MHz signal to produce two 10 Hz square waves.

These two 10 Hz square wave signals are inputted to a time interval counter, which continuously measures the time interval separating the rising edges of the two 10 Hz square waves. The measurement system computer controller continuously reads the counter and calculates the time rate of change of the counter's reading. The rate of change of the time interval counter reading is proportional to the time rate of change of the phase of NRB1 relative to NRX, which is the aging rate of NRB1 relative to NRX.

The time interval counter output also drives a digital-to-analog converter for analog display on a strip chart recorder. The strip chart recorder allows independent verification of the value calculated by the computer for the aging rate, since the slope of the chart recorder's trace is the maser aging rate.

At 5-minute intervals, the measurement system computer records and averages 10 samples from the output of the NRB1 versus NRX time interval counter. This 10-sample average is used with 6 similar time interval measurements recorded over the previous 30 minutes to calculate an updated value for the aging rate of NRB1 relative to NRX. The computer then records 10 samples from the output of each environmental sensor, which are averaged, time tagged, and placed in permanent storage with the updated value of the aging rate. The 5-minute interval raw data are later processed by the system controller to produce daily and weekly plots of all monitored parameters as functions of time. Figures 2 through 10 are examples of the system output.

Experimental Results

Figures 2 through 8 illustrate the hydrogen masers' susceptibility to barometric pressure variations over a 1-year period. The left vertical axis of each curve shows the maser aging rate as a function of time, and the right

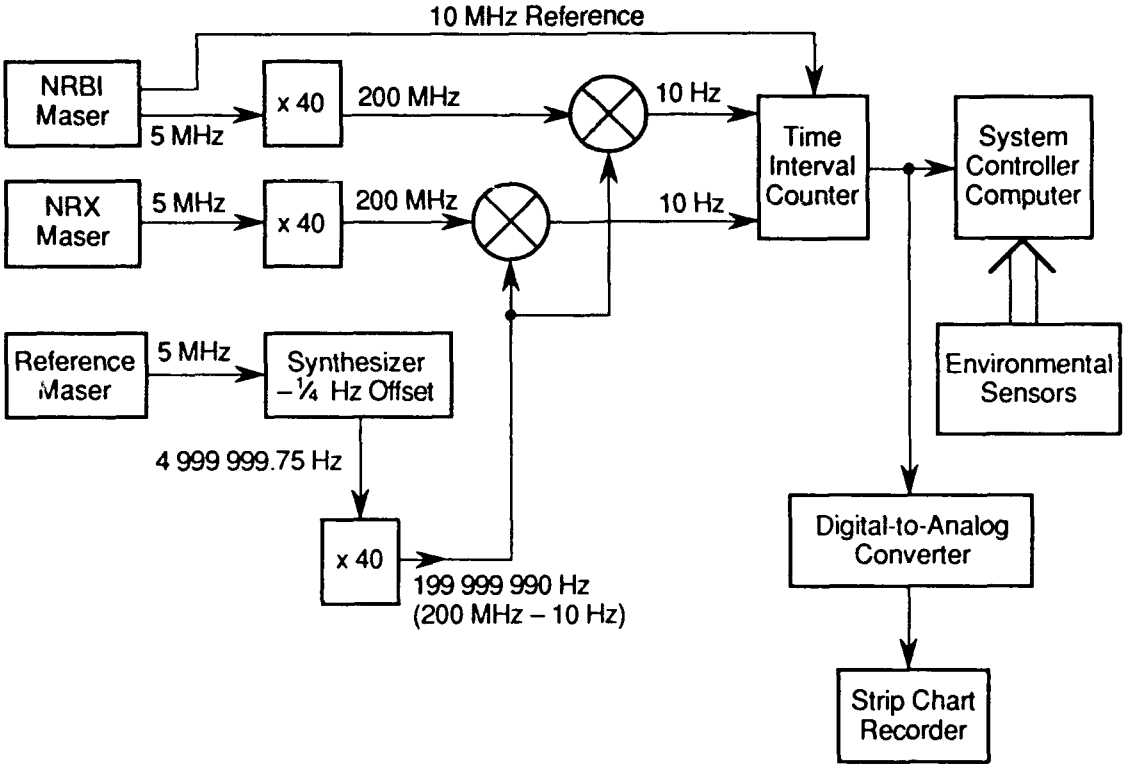


Figure 1. Maser Aging Rate Measurement System.

vertical axis shows the inverted barometric pressure versus time. The inverted barometric pressure curves were generated by subtracting the value of the barometric pressure, as measured in inches of mercury, from 30. The horizontal axis of each curve shows the time in hours or days.

As can be seen in Figures 2 through 8, a strong inverse relationship between changes in the barometric pressure and changes in maser aging rate exists, meaning a decrease in the pressure induces a positive increase in the maser aging rate and vice versa. The similarity of the curves in each figure is an indication of the strength and linearity of the barometer–aging rate relationship. Figure 5 demonstrates the ability of the barometric pressure to control the maser aging rate for periods as long as 2 weeks.

Figures 2 and 3 show the unprocessed data recorded at 5-minute intervals during 23 and 26 October 1990. The curves are presented to demonstrate that the aging rate responds instantaneously to an increase or decrease in the barometric pressure. There is no time lag in the response by the maser to changes in barometric pressure because pressure changes alter the amount of compression exerted on the resonant cavity. A change in compression affects the resonant frequency of the cavity, which in turn changes the maser's output frequency.

Figure 4 demonstrates that maser aging rate equally tracks both increases and decreases in the barometric pressure; that is, the magnitude of the positive or negative change in aging rate is equal for a fixed size increase or decrease in the barometric pressure. A crystal oscillator, in contrast, will typically demonstrate an asymmetry in the magnitude of the induced frequency offset in response to an equal magnitude increase or decrease in barometric pressure.

Figures 5 and 6 clearly illustrate the most significant characteristic of the influence of barometric pressure on hydrogen maser performance. Both figures demonstrate that pressure increases can induce an aging rate as large as 1 part in 10^{-13}. In a laboratory environment, when the barometric pressure is reasonably constant, the typical maser aging rate is a few parts in 10^{-15}. A strong weather front can temporarily degrade maser performance by more than an order of magnitude.

Figures 9 and 10 present the humidity and maser aging rate as functions of time for the periods considered in Figures 5 and 7 and show that humidity is not the source of the phenomena demonstrated in this work. There is some similarity between the humidity and aging rate curves in Figures 9 and 10, but it is much weaker than the correlation between the barometric pressure and the aging rate for the same time period (see Figures 5 and 7); there is also a 24-hour delay between a change in humidity and the change in the aging rate that may have been induced by the humidity change. Figures 9 and 10 are typical of the many plots generated during the past 12 months.

Since the barometric pressure often exerts some control over the humidity (there was no specific effort to control the humidity in the maser laboratory), any similarities between the aging rate and humidity curves are more likely due to the fact that both variables are influenced by the barometric pressure. Because the barometric pressure as a function of time shows a much stronger correlation to the maser aging rate versus time, and since changes in pressure and aging rate occur nearly simultaneously, it would seem that pressure induced the aging rate characteristics seen in Figures 5 and 7 and not humidity.

Discussion

The mechanism by which the barometric pressure influences hydrogen maser aging rate is a mechanical one. The resonant frequency of the microwave cavity used by the maser to observe the hyperfine transition is a function of the size and shape of the cavity. A change in the barometric pressure changes the compressive force exerted on the cylindrical resonant cavity, which in turn changes the cavity's size. The resulting minute change in the cavity's resonant frequency will be reflected in the 5 MHz output by a shift in maser aging rate.

A significant feature of the resonant cavity's design is a thermally insulating quartz sleeve within its outer wall. The addition of the quartz sleeve has reduced the temperature coefficient of the cavity by an order of magnitude over the same cavity design using aluminum. Consequently, there was no noticeable correlation between room temperature and the maser aging rate; the ambient temperature of the laboratory in which the masers are being operated has been controlled to within ±1.5°C.

Conclusion

Based on the evaluation of nearly a year's worth of data, it is empirically evident that there is a strong inverse relationship between changes in the barometric pressure and changes in hydrogen maser aging rate. Barometric pressure fluctuations were found to induce an aging rate as large as 1 part in 10^{-13} in hydrogen masers whose typical performance over the same period is a few parts in 10^{-15}. Induced changes in the maser aging rate were found to occur simultaneously with the corresponding changes in ambient pressure, and the aging rate was found to change equally with both increasing and decreasing barometric pressure. Barometric pressure was found to impact maser aging rate for periods as short as a few hours and as long as 2 weeks.

The effects of barometric pressure were found to mask the effects of the temperature and humidity. Surrounding the maser's microwave resonant cavity with an insulating quartz sleeve prevented temperature fluctuations from degrading hydrogen maser performance in a laboratory in which the ambient temperature was maintained to 23 ± 1.5°C.

References

[1] H. Helwig, in *Precision Frequency Control* (edited by E. A. Gerber and A. Ballato). Orlando: Academic Press, 1985, vol. 2, p. 159.

[2] J. Vanier and C. Audoin, *The Quantum Physics of Atomic Frequency Standards*. Philadelphia: Adam Hilger, 1989, p. 1115.

[3] J. Vanier, "The Active Hydrogen Maser: State of the Art and Forecast," *Metrologia*, vol. 18, p. 173, 1982.

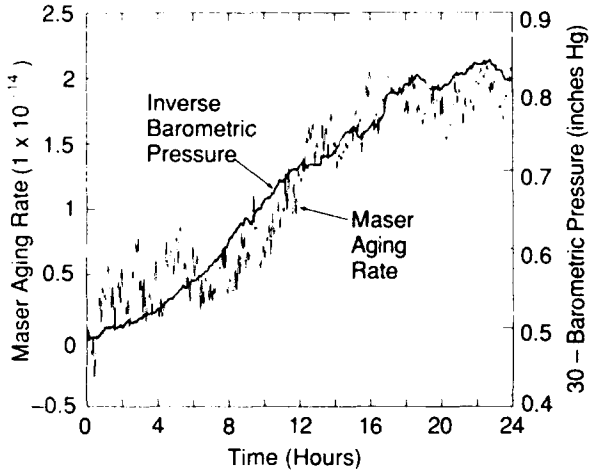


Figure 2. Maser Aging Rate and Inverse Barometric Pressure Versus Time (23 October 1990).

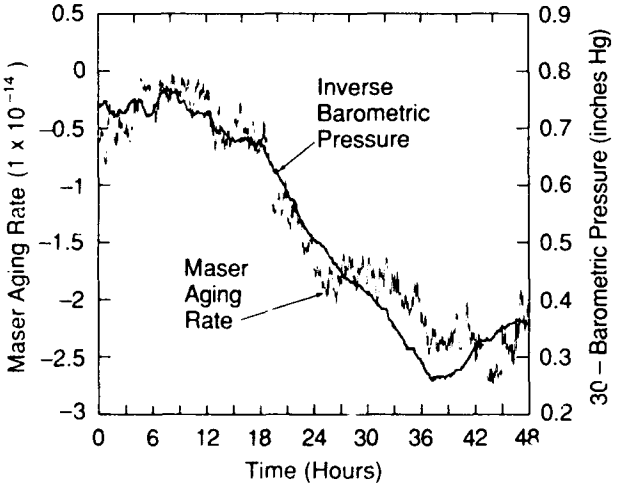


Figure 3. Maser Aging Rate and Inverse Barometric Pressure Versus Time (26 October 1990).

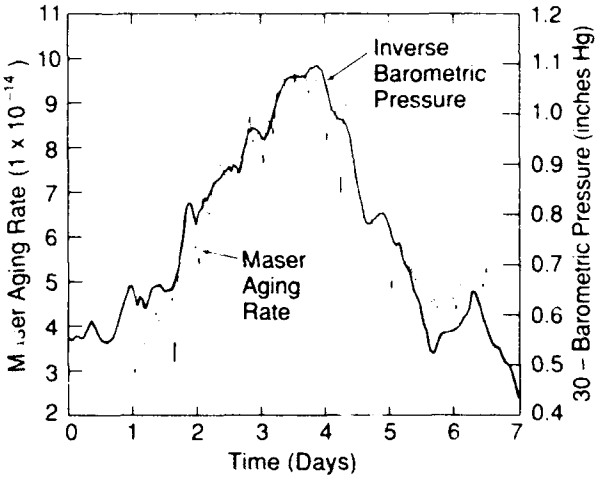


Figure 4. Maser Aging Rate and Inverse Barometric Pressure Versus Time (31 March 1990).

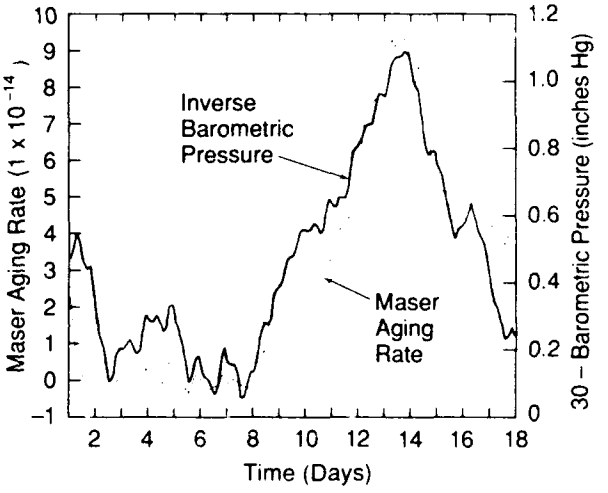


Figure 5. Maser Aging Rate and Inverse Barometric Pressure Versus Time (25 March 1990).

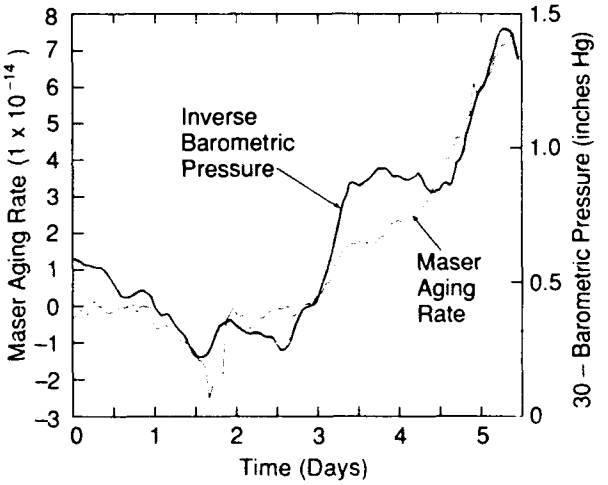


Figure 6. Maser Aging Rate and Inverse Barometric Pressure Versus Time (27 February 1991).

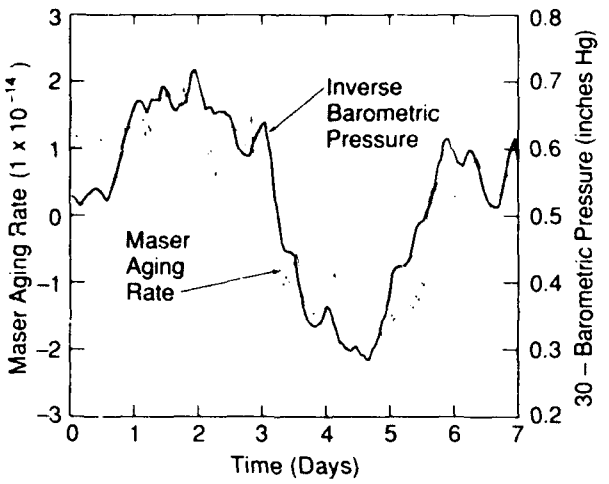


Figure 7. Maser Aging Rate and Inverse Barometric Pressure Versus Time (4 July 1990).

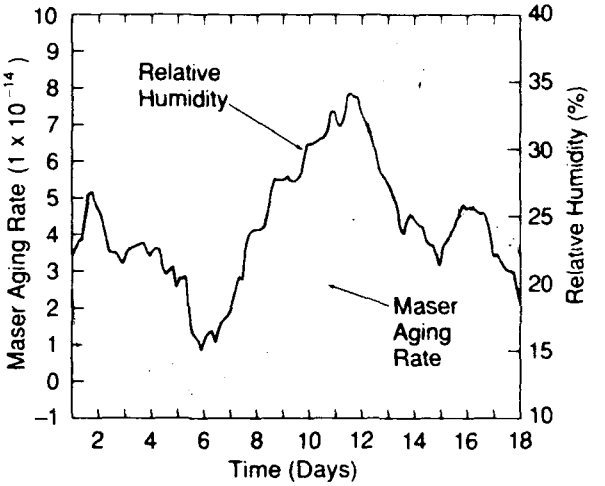


Figure 9. Maser Aging Rate and Relative Humidity Versus Time (25 March 1990).

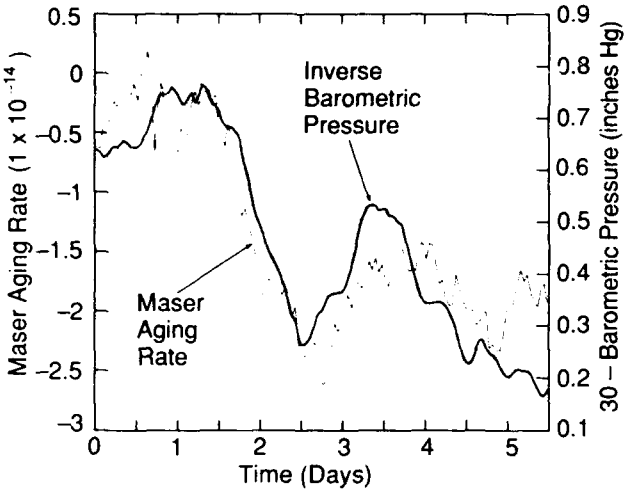


Figure 8. Maser Aging Rate and Inverse Barometric Pressure Versus Time (25 October 1990).

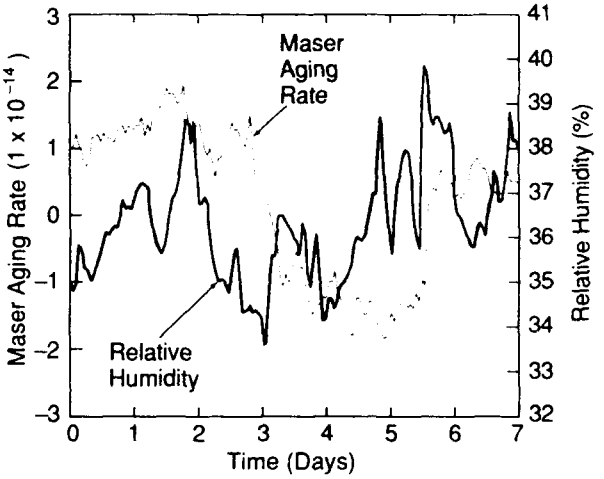


Figure 10. Maser Aging Rate and Relative Humidity Versus Time (4 July 1990).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ATOMIC FREQUENCY STANDARDS FOR SATELLITE RADIONAVIGATION SYSTEMS

Y.G.Gouzhva, A.G.Gevorkyan, V.V.Korniyenko

Leningrad Scientific Research Radiotechnical Institute (LSRRI), 2, Rastrelli square, Leningrad, 193124, USSR

Abstract: Data are given concerning with the LSRRI's developments of atomic frequency standards (AFS) for various applications, including AFS with atomic beam tube (AFS - ABT), with rubidium gas cell and with hydrogen maser. The onboard AFS - ABT to be intended for GLONASS space vehicles is described. The possible promising directions of international cooperation in the field of the AFS creation are considered.

Leningrad Scientific Research Radiotechnical Institute (LSRRI) of USSR Ministry of Radio Industry is conducting researches and developments in the field of creation of different systems and aids for Positioning and Timing Service (PTS) for more than three decades of years, including PTS' main frequency-determining devices, i.e. atomic frequency standards (AFS). LSRRI is a leading organization in designing, production and implementation of a series of PTS systems and aids on different physical principles for various purposes. Among those items, it is necessary, first of all, to commend synchronization and common time systems and aids using space-based navigation-timing GLONASS systems, ground-based long-range pulse-phase "Chayka" radionavigation system and also State System of Common Time and Standard Frequencies of the USSR. This system foresees the use of above mentioned systems and aids, as well as television, meteor tails and other aids. A great number of high qualified specialists in LSRRI's staff including prominent scientists, designers, technologists, provides a great scientific and technical potential of the institute. Its realization makes it possible, to create on the basis of principle new technical solutions the PTS systems and aids in accordance with continuously growing user requirements including the signal generators with corresponding performances on the basis of AFS. In 1991, the AFS of all of the most frequent types being intended for the use in subsystems of different ground- and space-based complexes have designed, developed and, as a rule, industrially producted. Those AFS include:

- AFS on the basis of Cs atomic beam tube (AFS - ABT) for GLONASS space vehicles and AFS - ABT for land stationary equipment with frequency instability at the level of 10^{-13} and long life time;
- AFS on the basis of Rb gas cell (AFS - RGC) for onboard equipment of different vehicles with accuracy performances at the level of 10^{-11} functioning over the wide temperature range in presence of different environmental disturbances;
- AFS on the basis of hydrogen generator (AFS - HG) for land stationary equipment.

The devices to have been developed used in Tsikada navigation system, KOSPAS - SARSAT system for rescuing those who meet disaster and GLONASS navigation-time system. The metrological performances of AFS to be used are over the range from 1.10^{-10} to 1.10^{-13}.

G E M

Cesium beam frequency standard
for on-board space equipment

Specifications

Output frequency	5 MHz
Relative frequency accuracy over the life time	$\pm 1.10^{-11}$
Stability, averaging time:	
1 s	5.10^{-11}
100 s	1.10^{-11}
1 hour	$2,5.10^{-12}$
1 day	5.10^{-13}
Temperature change of frequency, 1/°C	5.10^{-13}
Temperature range, °C	0...+40
Power supply, DC	27 V
Power consumption, not greater than	80 W
Dimensions, mm	370x450x500
Weight, kg	39,6
Operational life, hours	17500

CH2965-2/91/0000-591 $1.00 © 1991 IEEE



Fig. 1 Gem

B E R Y L

Rubidium cell frequency standard
for space vehicles

Specifications

Output frequency	5 MHz
Frequency ajustment step	$2{,}2.10^{-11}$
Relative frequency accuracy	$(2\text{-}3).10^{-11}$
Stability, averaging time:	
1-100 s	$2.10^{-11}\cdot\tau^{-1/2}$
Frequency drift, not greater than	1.10^{-11}/month
Temperature range, °C	-5...+45
Temperature change of frequency, not greater than, 1/°C	$\pm 3.10^{-12}$
Non-harmonic distortion in 100 kHz bandwidth(5MHz), not less than, dB	95
Warm-up (from 25°C) up to $\pm 5.10^{-11}$, min	30
Power supply, DC	23-24 V
Power consumption, in warm-up (30 min)	40 W
in operation mode at 25°C	20 W
Operational life, not less than hours	33,500
Dimensions, mm	300x268x120
Weight, not greater than, kg	8

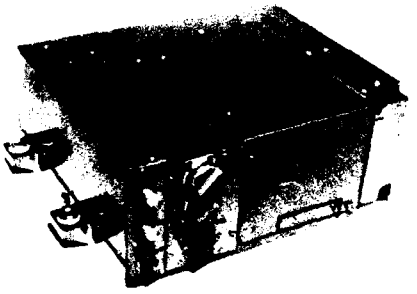


Fig. 2 Beryl

M A L A C H I T E

Cesium beam frequency standard
for on-board space equipment

Specifications

Output frequency		5 MHz
Relative frequency accuracy over the life time		$\pm 1.10^{-11}$
Stability, averaging time:		
	1 s	2.10^{-11}
	100 s	5.10^{-12}
	1 hour	5.10^{-13}
	1 day	1.10^{-13}
Temperature change of frequency, 1/°C		2.10^{-13}
Temperature range, °C		0...+40
Power supply:	DC	27 V
Power consumption, W		90
Operational life, hours		28,500
Dimensions, mm		421x414x655
Weight, kg		52

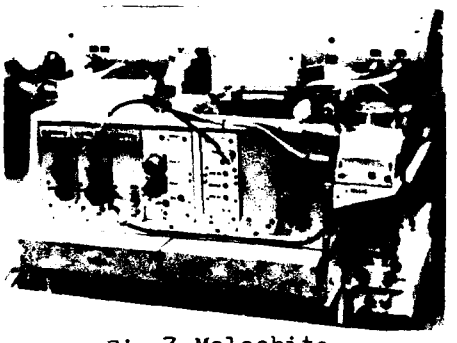


Fig.3 Malachite

S A P P H I R E

Cesium beam frequency standard
for the ground stationary equipment

Specifications

Output frequency		5 MHz
Relative frequency accuracy over the life time		$\pm 1.10^{-11}$
Reproducibility, from turn on to turn on, not greater than		2.10^{-12}
Stability, averaging time:		
	1 s	1.10^{-11}
	100 s	1.10^{-11}
	1000 s	5.10^{-12}
	1 hour	2.10^{-12}
	1 day	5.10^{-13}
Temperature range, °C		+5...+40
Power supply:	AC	220+ 22 V, 50 Hz
	DC	27 V
Power consumption:	AC	180 VA
	DC	110 W
Dimensions, mm		506x413x623
Weight, kg		65
Life time, years		10

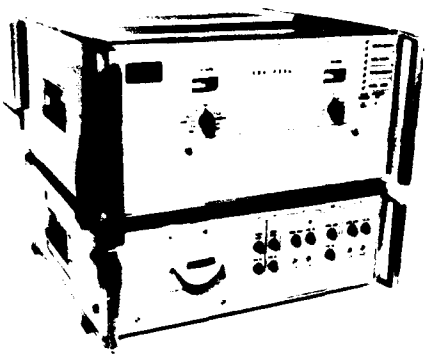


Fig. 4 Sapphire

The supporting of AFS high maintenance performances, technological suitability, reliability and long operational life isn't a trivial task. It is necessary to have a significant scientific margin to solve a series of tasks as follows:

- investigation of destabilizing factors influencing on AFS mentrological performances, researches of monitoring and accounting methods of those factors in order to exclude its influence over the specified limits for different types of AFS in operating conditions taking into account temperature, vibration, shock and other disturbances;
- development of vibration- and shock-proof design;
- development of vacuum technology for radiospectroscope production, methods of filling and frequency ajustment of gas cells and manufacturing of spectral lamps.

The most accurate state-of-the art satellite system is GLONASS navigation system which obviously demonstrates LSRRI's achievments in the field of AFS development. The basis of system frequency-time support is the synchronization system comprising the main synchronizer, onboard satellite frequency and time standard, facilities for satellite orbit monitoring and software for prediction on onboard clock rate. The latest achievment in the field of AFS development for the space equipment is an AFS - ABT with the daily instability of $\sim 10^{-13}$. For the main synchronizer, one uses AFS - ABT with the instability of $\sim 2.10^{-13}$ over 1 day and AFS-HG with daily instability $\sim 5.10^{-14}$.

The further improvement of radiosystems, particularly GLONASS, means the development of onboard space AFS with frequency instability $< 5.10^{-14}$. The researches are being performed in two fields:

1) development of compact HG with a regeneration of resonator quality; the daily instability of $\sim 2.10^{-14}$ is realized;
2) development of AFS - ABT with optical pumping of the atomic beam. The figure of merit $F \geq 20$ and $\sigma(\tau) = 5.10^{-12} \cdot \tau^{-1/2}$ are achieved with beam pumping through one laser. The predicted value are $F \geq 500$ and $\sigma(\tau) = 2.10^{-13} \cdot \tau^{-1/2}$.

The improvement of performances is realized on the basis of utilization of two lasers' pumping, two photodiodes for detection, and increasing of photon collection angle and beam intensity.

As much as the problems relating to improvements of AFS performances seem to be solved by many U.S. companies and agencies, it would be advisable, to evaluate possibilities and prospects of international cooperation within this field. On our opinion, the most promising directions of cooperation are as follows:

- as applied to AFS - ABT - scientific and technical cooperation including:

a) joint researches and developments for the realization of onboard AFS - for space vehicles with daily frequency instability of the order of 10^{-14};

b) joint works on creation of a metrological AFS - ABT with accuracy performances at the level of 10^{-13} for land equipment functioning in real conditions at user platforms;

c) joint realization of predicted accuracy performances of on-board AFS - ABT with optical (laser) pumping of an atomic beam;

- as applied to AFS - RGC - scientific and technical as well as industrial cooperation for designing and industrial production of miniaturized AFS - RGC with accuracy performances at the level of $10^{-12} - 10^{-13}$ and increased operational life;
- as applied to AFS - HG - scientific and technical cooperation on creation of high reliable compact AFS - HG with increased operational life and accuracy performances at the level of $10^{-14} - 10^{-15}$ for on-board GLONASS/GPS space vehicles' equipment as well as for space vehicles of a new generation.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DEEP SPACE TRACKING AND FREQUENCY STANDARDS

James S. Border and E. Robert Kursinski

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr. / MS 238-600
Pasadena, California 91109

ABSTRACT

Radio metric measurements are used for determining the trajectory of a spacecraft and also for scientific investigation of phenomena which affect wave propagation or spacecraft motion. Both the verity of trajectory determination and the quality of radio science depend largely on measurement accuracy. Requirements imposed by tracking are identified for the station frequency reference in terms of timing, stability and spectral purity, requirements are identified for the spacecraft on-board oscillator, and requirements are identified for inter-station time and frequency synchronization. The functional form of the effects of clock epoch offset, clock rate offset, and clock instability on doppler, range and interferometric tracking observables is presented. Examples are given of tracking techniques which are currently limited by the performance of frequency standards.

INTRODUCTION

The use of highly stable frequency standards within the NASA Deep Space Network (DSN) for dissemination of timing and frequency has made possible the precise measurement of the propagation time of electromagnetic radiation between a ground antenna and a spacecraft. These measurements are the basis for the doppler and range observables which are used for spacecraft tracking. Radio metric measurements are used for determining the spacecraft trajectory and also for scientific investigation of phenomena which affect wave propagation or spacecraft motion. Both the verity of trajectory determination and the quality of radio science depend largely on measurement accuracy. A number of applications of spacecraft tracking measurements will be presented, along with discussions of how these measurements depend on frequency and timing.

Observables which may be extracted from the radio signal emitted by a spacecraft and received at a tracking station include signal phase, group delay, and amplitude. The phase of the incoming carrier signal ("doppler") is measured relative to the phase of a reference signal generated from the station frequency standard. Group delay ("range") is extracted by measuring the differential phase between sidetones which have been modulated onto the carrier signal. These measurements provide information about the geometric delay between transmitter and receiver, and also about propagation media. The signal emitted by the spacecraft may be derived from the spacecraft on-board oscillator, or the spacecraft may transpond a signal which was transmitted to the spacecraft from the tracking antenna. In the former case, interpretation of the observables depends on time and frequency synchronization between the transmitter and receiver, while in the latter case it is the stability of the station frequency standard which is important.

The path of a spacecraft is determined by celestial mechanics. But any error in initial conditions, or small perturbing forces acting on the spacecraft, will cause the trajectory to deviate from the planned course. Tracking data are routinely collected to determine the trajectory. These data are used to plan midcourse trajectory correction maneuvers, to plan encounter sequences which include instrument timing and pointing, and to reconstruct the spacecraft orbit for analysis of science data. The spacecraft equations of motion are integrated in an inertial coordinate system; tracking station coordinates must be known in inertial space for analysis of the tracking data. The DSN has adapted the technique of radio interferometry for the purpose of establishing an inertial reference frame defined by the positions of the distant quasars, for measuring the orientation and rotation rate of the Earth in inertial space, and for measuring the baseline vectors joining the tracking stations. This technique, known as very-long-baseline interferometry (VLBI), is also used to directly measure spacecraft angular position.

CH2965-2/91/0000-594 $1.00 © 1991 IEEE

Radio measurements also contain information about media through which the signals pass. Planetary encounter geometries are often selected so that the signal path from spacecraft to Earth will be occulted by the planet's atmosphere and/or ring system. By carefully modeling and removing known geometric factors, details of the intervening media may be gleaned from their effects on the radio signal. In other cases, the motion of the spacecraft itself contains information about the environment. The mass of a small body, or the gravity field distribution of a planet, may be determined from the perturbations induced in the orbit of the spacecraft. If the radius is known from either optical or occultation data, then this provides an estimate of density, which is a quantity of much interest for understanding the evolution of the solar system. If a probe or balloon is placed in the atmosphere of a planet, then wind speed and direction may be directly determined from observations of the motion of the probe or balloon.

One of the predictions of the general theory of relativity is that gravitational radiation may be emitted and propagate. The interaction of these waves with matter is very weak, requiring extraordinary sensitivity for detection of the perturbations these waves create in the motion of the affected matter. The spacecraft doppler link functions as a sensor which may be used to search for low frequency gravitational radiation in the mHz band.

Spacecraft transmit signals in bands which have been allocated for deep space communications. Currently, S-band (2.3 GHz) and X-band (8.4 GHz) are used. The DSN developed S-band capability for uplinks and downlinks in the 1960's. In the mid 1970's, spacecraft were equipped with dual frequency S/X downlinks. Signals at the two downlink bands are coherent with each other, derived from the same source (either the spacecraft on-board oscillator or else a signal uplinked from a tracking station). In 1989, an X-band uplink capability was added. Both Galileo and Magellan may transmit coherent S/X downlinks which are derived from an X-band uplink. The use of Ka-band (32 GHz) is planned for the late 1990's. The move toward higher frequencies is largely driven by the better telecommunications performance which results, but higher frequencies also improve the accuracy of radio metric measurements by reducing effects due to charged particles and by increasing the resolution through the use of shorter wavelengths. An experiment to be flown by the Cassini mission will incorporate Ka-band uplinks and downlinks to search for low frequency gravitational radiation.

The time scales of interest for the various measurements which have been discussed are illustrated in Fig. 1. The information content in observables used for navigation is spread over time scales from about one minute up to

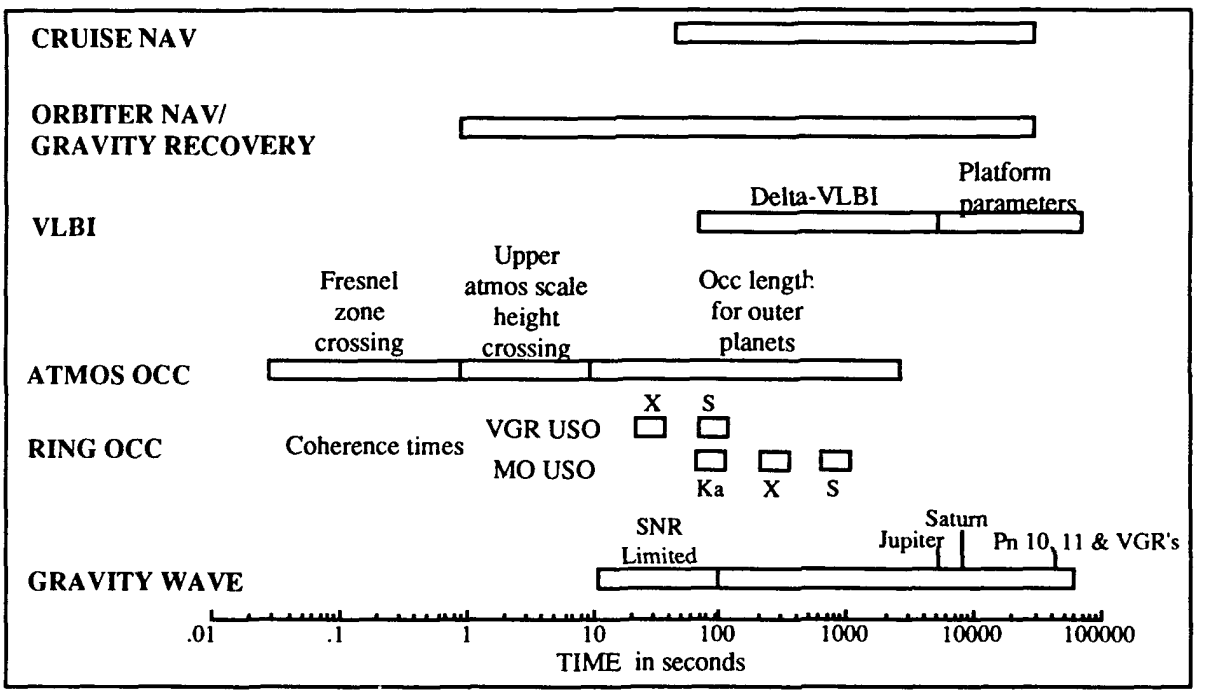


Fig. 1. Time scales of interest for navigation and radio science measurements.

about ten hours (the length of a station pass, where the spacecraft is tracked from rise to set). The higher order components of a planet's gravity field may perturb spacecraft motion over time scales as short as a few seconds. Radio science measurements span time scales from just a few hundredths of a second, for the time for a ray to descend the height of a Fresnel zone in an atmosphere, to the longest available round trip light times for a gravity wave search [Kursinski, 1990].

The next section of the paper defines the several modes which are used for spacecraft tracking, discusses their typical applications, and describes how frequency and timing affect the measurements. The third section develops the functional form of the effects of clock epoch offset, clock rate offset, and frequency instability on doppler, range, and interferometric tracking observables. Error budgets are presented so that frequency and timing effects may be contrasted with other measurement system errors. The final section of the paper presents examples of tracking techniques which are currently limited by the performance of frequency standards.

TRACKING MODES

Several alternative modes are employed for tracking interplanetary spacecraft. In the simplest mode, known as "1-way", a signal is derived from an oscillator on board the spacecraft and, a one-way light time later, is received at a tracking station and measured relative to the station's frequency standard. Besides offering reliability, due to its simplicity, this mode offers a better signal-to-noise ratio for the reception of spacecraft telemetry data. For a distant spacecraft, the phase noise of the on-board oscillator may be less than the phase noise of an uplinked carrier, and, further, a lower receiver system temperature results when the tracking station is configured for listen-only operations. The 1-way mode is essential for most occultation measurements, since no strain is placed on the spacecraft receiver's tracking loop and since the interpretation of received data is much simpler for waves passing only one direction through refractive media or through a diffraction grating. But the accuracy of 1-way tracking observables is limited by the performance of the spacecraft oscillator, and hence this mode has seldom been preferred for the purpose of navigation. Media errors are reduced by about a factor of two for this mode, however. This can offer some navigational advantages, especially for geometries where the signal raypath passes close by the sun.

In "2-way" tracking, a signal is uplinked from a tracking antenna, transponded by the spacecraft, and received back at the same station a round trip light time after transmission. This mode provides direct line-of-sight measurements, namely range and range-rate, of the spacecraft state. The 2-way mode is generally available when commands are being uplinked to the spacecraft and provides the most accurate measurements under most observing conditions, due to the inherent better stability of the ground versus spacecraft oscillator. This tracking mode is generally preferred for navigation, it is preferred for making gravity field measurements, and it is used in the search for detection of low frequency gravitational radiation. For this case, the accuracy of tracking data is ultimately limited by the stability of the ground station frequency standard.

When a different station is used to receive the signal, than was used to supply the uplink, the tracking mode is referred to as "3-way". One use of 3-way tracking is to provide coherent communications with spacecraft so distant that the round trip light time exceeds the visibility window of a single Earth-based station. The range to the most distant spacecraft are measured in this way. This mode can be used to add redundancy and geometric diversity to a tracking schedule, since many stations may be simultaneously tracking in the 3-way mode. Signals acquired at separate antennas may also be combined by arraying to boost the signal-to-noise ratio. This is done routinely with baseband telemetry signals. It may also be employed at microwave frequencies to improve the signal-to-noise ratio of occultation measurements. To enable microwave arraying, the coherence time of the frequency standards used at each station must be such that their relative phase wander by much less than an RF (radio frequency) cycle over the coherent integration time. Tracking accuracy for this case is limited by the performance of inter-station time and frequency synchronization.

In interferometric tracking, two ground stations are used to simultaneously receive signals from one radio source. The difference in time of arrival of the signal at the two stations is measured. This time delay measurement, coupled with knowledge of the baseline vector joining the two antennas, provides a direct geometric determination of the angle between the baseline vector and the direction to the radio source. Two-dimensional angular position is determined by making measurements along two baselines with substantially different orientations. The DSN makes VLBI measurements using the baseline between the Goldstone (California) and Madrid (Spain) complexes, and

using the baseline between the Goldstone and Canberra (Australia) complexes.

Sessions involving many (20 to 200) observations are used to determine the parameters which define the origin and orientation of the coordinate system. Such parameters are referred to as platform parameters. Station and quasar coordinates are estimated by combining data from multiple sessions spanning up to 15 years. The variation in Earth orientation is typically measured using data from just one or two sessions, spaced close together in time. Inter-station time and frequency synchronization is also provided by these measurements.

For a single observation of spacecraft time delay to be meaningful, the interferometer must be calibrated. For this purpose, the DSN has developed the technique of ΔVLBI. A natural radio source (quasar) is observed either just before or just after the spacecraft. The ΔVLBI observable is the spacecraft time delay minus the quasar time delay. This provides spacecraft angular position in the radio reference frame defined by the quasars. The quasar measurement provides calibration for time and frequency offsets between stations, it provides calibration for instrumental phase shifts and group delays, and it also provides reduction in errors due to uncertainties in media delays and platform parameters by an amount which depends on the angular separation between the spacecraft and the quasar.

When spacecraft are sufficiently close in angle that they may be observed in the same beamwidth of an Earth-based antenna, differential VLBI measurements may be made between sources which are simultaneously observed [Counselman *et al.*, 1972; Border *et al.*, 1991]. This technique is referred to as same-beam interferometry (SBI). For example, all spacecraft on the surface of Mars and all spacecraft within areosynchronous orbit could be observed simultaneously at X-band frequencies in the same beam of an Earth-based antenna with 34 meter diameter. The SBI observable provides an extremely precise measurement of the plane-of-sky separation between spacecraft, due to almost complete error cancellation.

Navigation

Line-of-sight components of spacecraft state are directly observed by 2-way tracking, while plane-of-sky components of spacecraft state are directly observed by interferometric tracking. Generally, a spacecraft trajectory is estimated by combining various measurements collected over an arc, ranging from a few hours for a short period planetary orbiter, up to many months for a spacecraft in interplanetary cruise.

During cruise, the diurnal rotation of the Earth imposes a sinusoidal variation on doppler data received at a tracking station from a distant spacecraft. The phase and amplitude of the sinusoidal variation depend on the spacecraft's right ascension and declination, respectively. A single pass of doppler, spanning a sizable fraction of a full 24 hour period, thus provides some information about angular position [Melbourne, 1976]. A longer arc of range and doppler observes the full state of the spacecraft, through sensing the effect of the solar gravitational acceleration on the spacecraft [Curkendall and McReynolds, 1969]. These measurements also begin to sweep out the state space as the orbital motions of the Earth and spacecraft change the geometry. However, these measurements are largely insensitive to declination for spacecraft declinations near zero, and trajectory determination using long arcs of range and doppler depends on dynamic modeling. Solution accuracy can be substantially degraded by mismodeling of non-gravitational accelerations [Ondrasik and Rourke, 1971].

Interferometric measurements naturally complement range and doppler. Angular coordinates may be determined more accurately by a factor of ten or more [Melbourne and Curkendall, 1977; Border *et al.*, 1982]. Inclusion of interferometric measurements greatly reduces sensitivity of the orbit determination process to mismodeled accelerations, since angular components are directly measured. Also, interferometric measurements are sensitive to angular coordinates over the full range of spacecraft declinations.

For a planetary orbiter, the motion of the spacecraft about the planet induces a strong signature in the doppler received at Earth. The planet-relative position of the spacecraft may be recovered from analysis of this signature over one or more revolutions. The orientation of the orbit plane about the line of sight from the Earth to the planet is the least well determined component of state. This component may be directly observed by interferometric delay rate measurements [Poole *et al.*, 1980]. When two orbiters are observed simultaneously, the inclusion of same-beam interferometric measurements dramatically improves orbit accuracy for both spacecraft, by precisely observing components of spacecraft state which are complementary to the component directly observed by doppler [Folkner and Border, 1990].

Radio Science

Radio signals can be used to investigate planetary atmospheres. An atmosphere acts as a lens, bending and inducing phase shifts in signals passing through. The phase shift due to an atmosphere can be computed from the measured total phase by subtracting a model of what the phase would have been if no atmosphere were present. The model depends on a precise reconstruction of the occultation geometry. The atmospheric phase shift is used to recover temperature and pressure profiles through modeling of atmospheric layers [Fjeldbo *et al.*, 1971].

An occultation by a tenuous atmosphere might last only ten seconds, and produce a total phase shift of a few tenths of a cycle at X-band, whereas an occultation by a large outer planet can last more than 1000 seconds producing a phase shift of more than one million X-band cycles. These measurements are normally done in the 1-way mode. Instability of the spacecraft on-board oscillator directly contributes an error. This is especially important for tenuous atmospheres, where the total phase shift is small. Multipath signals are generated when the signal path approaches and crosses an atmospheric layer with a much higher refractive index than neighboring layers. Microwave arraying can significantly improve occultation measurements for which there is a high level of signal attenuation due, for example, to a microwave absorbing ammonia layer within an atmosphere.

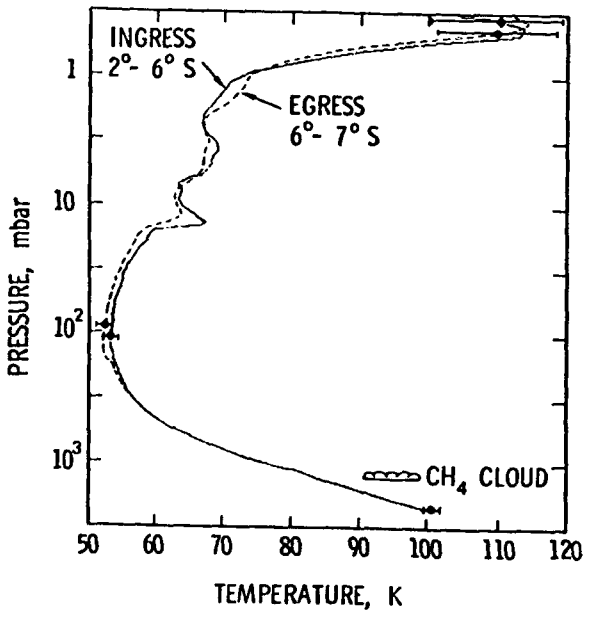


Fig. 2. Vertical temperature vs pressure profile for atmosphere of Uranus recovered from Voyager 2 radio occultation measurements. (Figure from [Lindal *et al.*, 1987].)

Fig. 2 shows the temperature and pressure profile recovered from the Voyager 2 occultation by Uranus. Two curves are shown; the ingress curve is derived from measurements as the spacecraft entered into occultation and the egress curve is derived from measurements as the spacecraft exited out of occultation. Note that the error bars are much larger near the top of the figure, which corresponds to the top of the atmosphere where the propagation effects are smallest. Improved stability of the spacecraft on-board oscillator would provide more accurate determination of these parameters [Kursinski, 1990].

Radio signals are also used to investigate planetary ring systems. When an incident wave strikes a slab of ring material, two signals emerge [Tyler *et al.*, 1983]. The coherent signal is shifted in phase and reduced in amplitude by an amount which depends on the optical depth at the microwave frequency of the signal. An incoherent scattered signal is also transmitted.

The rings act as a diffraction grating. If $T(\rho)$ is the complex transmittance of the rings at radial distance ρ, then a simplified expression for the measured transmittance is given by [Marouf *et al.*, 1986]

$$\hat{T}(\rho_0) = \frac{1-i}{2F}\int_{-\infty}^{+\infty} T(\rho)\, e^{i(\pi/2)[(\rho_0-\rho)/F]^2}\, d\rho$$

where F is the Fresnel scale (square root of product of wavelength times distance from spacecraft to ring plane) and ρ_0 is the radial distance in the ring plane where the direct signal passes. Since the signal is coherent, allowing both the amplitude and phase of $\hat{T}(\rho_0)$ to be recovered, this diffraction integral can be inverted to solve for $T(\rho)$,

$$T(\rho) = \frac{1+i}{2F}\int_{-\infty}^{+\infty} \hat{T}(\rho_0)\, e^{-i(\pi/2)[(\rho-\rho_0)/F]^2}\, d\rho_0$$

The diffraction limited transmittance $\hat{T}$ is measured over a range of values of ρ_0, as spacecraft motion changes the geometry. The coherence time of the oscillator limits the region over which the integral may be inverted, and in fact determines whether inversion is even possible. This, in turn, limits the radial resolution which is obtained for ring opacity. Fig. 3 shows an example of observed opacity from the Voyager 1 encounter with Saturn. The true opacity, derived from inverting the diffraction integral, clearly shows the presence of the narrow F Ring.

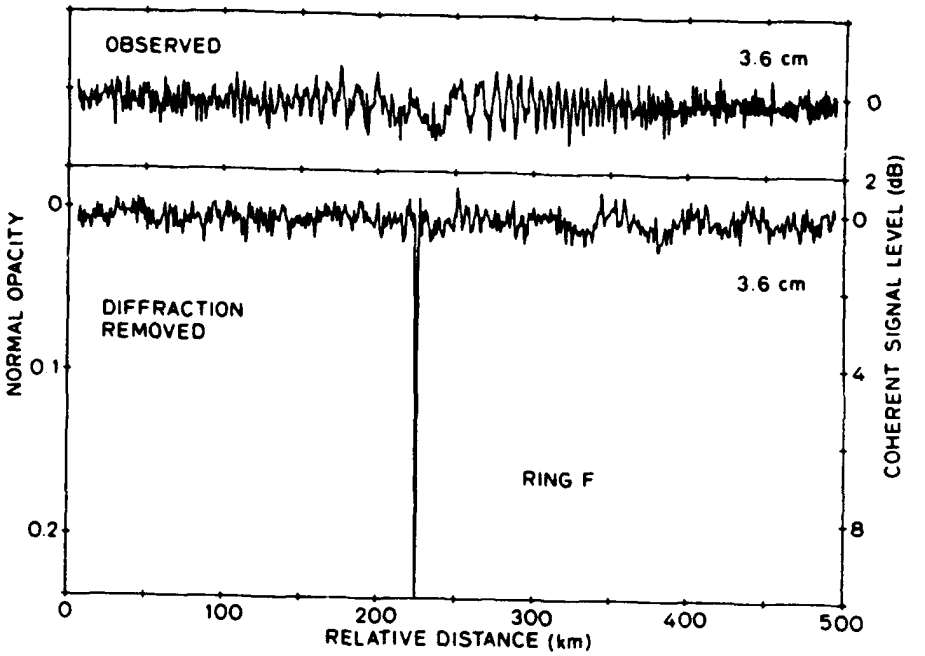


Fig. 3. Opacity, observed and with diffraction removed, for region of Saturn ring plane near the F Ring. Data obtained from Voyager 1 flyby. (Figure from [Tyler *et al.*, 1983].)

OBSERVABLE ERROR BUDGETS

Observables are affected by frequency and timing directly, through the contribution of any phase errors on the reference signals which are used to drive the transmitter and receiver. Since modeling must be used to extract either navigation or radio science parameters from the measured signal, frequency and timing indirectly affect observables through geometric dependence of the models on observable time-tags. Expressions are given here for the functional form of clock epoch offset, clock rate offset, and frequency instability on doppler, range, and VLBI observables. Error budgets are shown so that frequency and timing effects may be compared with other sources of measurement error. First, models must be given for observables.

A 2-way range observable r_k is a measure of round trip light time, as recorded by the station clock,

$$r_k = ST(t_k) - ST(t_k - RTLT),$$

where RTLT is the round trip light time (sec) and $ST(t_k)$ is the time (sec) as kept by the station clock at true time t_k. A 2-way doppler measurement d_k is defined in terms of differenced range, by the expression

$$d_k = \frac{r_{k+1} - r_k}{T_c},$$

where $T_c = t_{k+1} - t_k$ is the doppler count time. The doppler observable is a more precise measure of range change than would be obtained by literally differencing range observables, since doppler is extracted from RF carrier phase measurements, while range is extracted from side tone modulations with a bandwidth of a few megaHertz. But frequency and timing affect doppler and differenced range in the same way. A range measurement is modeled as

$$r_k = \Delta t(t_k) + \rho_R(t_k) - \Delta t(t_k - RTLT) + \varepsilon_k, \quad (1)$$

where Δt is the station clock error, ρ_R is the round trip range, and ε_k is the measurement noise.

Clock Epoch and Rate Offset: Range

A constant epoch offset δt (sec) of station clock time with respect to true time has no direct effect, since clock time errors enter eq. (1) with the opposite sign at the times of transmission and reception. A rate offset $\dot{\delta t}$ (sec/sec) directly affects the measurement by the amount of clock drift over the round trip light time. The range error due to

clock epoch and rate offset, obtained by taking partial derivatives of eq. (1), is

$$\delta r_k = -\dot{\rho}_R(t_k)\,(\delta t) + \mathrm{RTLT}\,(\delta \dot{t})\,.$$

The epoch offset affects the measurement through geometric dependence of the range on the time of signal reception.

The model for 1-way range is written as

$$r_{1,k} = \Delta t_{STN}(t_k) + \rho_1(t_k) - \Delta t_{SC}(t_k - \mathrm{OWLT}) + \varepsilon_k\,,$$

where Δt_{STN} is the station clock error, Δt_{SC} is the spacecraft clock error, ρ_1 is the 1-way range, and OWLT is the one way light time. The model for 3-way range is written as

$$r_{3,k} = \Delta t_{RCVR}(t_k) + \rho_R(t_k) - \Delta t_{TRAN}(t_k - \mathrm{RTLT}) + \varepsilon_k\,,$$

where Δt_{RCVR} is the downlink station clock error and Δt_{TRAN} is the uplink station clock error. Clock epoch and rate offset errors in 1-way range, 3-way range, and 2-way minus 3-way range are given, respectively, by the following three expressions:

$$\delta r_{1,k} = -\dot{\rho}_1(t_k)\,(\delta t_{STN}) + \delta t_{STN\text{-}SC}\,,$$

$$\delta r_{3,k} = -\dot{\rho}_R(t_k)\,(\delta t_{RCVR}) + \delta t_{RCVR\text{-}TRAN}\,,$$

$$\delta r_{2\text{-}3,k} = -\dot{\rho}_{R,STN1}(t_k)(\delta t_{STN1}) + \dot{\rho}_{R,STN2}(t_k)(\delta t_{STN2}) + \mathrm{RTLT}\,(\delta \dot{t}_{STN1}) - \delta t_{STN2\text{-}STN1}\,.$$

Clock synchronization dominates the error budgets for all range data types other than 2-way range. For 1-way range the synchronization between spacecraft and station is important; for 3-way range the synchronization between the uplink and downlink stations is important; for the interferometric data type obtained by differencing 2-way and 3-way range the synchronization between the two downlink stations is important. The utility of 1-way range for deep space tracking is limited, since it is generally impossible to know the clock synchronization to better than the level for which the range can be determined by other means. However, 1-way doppler is useful in some situations due to its high precision. For planetary orbiters, where the dynamics provide some data strength to estimate a range-rate bias, 1-way doppler can be competitive with 2-way doppler.

Clock Epoch and Rate Offset: VLBI

Typically, a ΔVLBI measurement consists of an observation of a spacecraft, followed by an observation of a quasar, followed by another observation of the spacecraft. The direct effects of clock epoch and rate synchronization between stations are eliminated by differencing delays between sources and by interpolation. Timing effects cause indirect errors through geometric model dependence. The error in a ΔVLBI measurement due to an uncalibrated station instrumental group delay of τ_I and due to an uncalibrated clock synchronization error of $\delta t_{STN2\text{-}STN1}$ is given by [Border *et al.*, 1991]

$$\varepsilon_{\Delta VLBI} = [\Delta\dot{\rho}_{1,SRC1}(t) - \Delta\dot{\rho}_{1,SRC2}(t)]\,\tau_I + [\dot{\rho}_{1,SRC1}(t) - \dot{\rho}_{1,SRC2}(t)]\,\delta t_{STN2\text{-}STN1} \qquad (2)$$

where $\dot{\rho}_1$ is the 1-way range-rate from the radio source to station 1 and $\Delta\dot{\rho}_1$ is the station-differenced 1-way range-rate (i.e. interferometric delay rate). The spacecraft is referred to as "SRC1", while "SRC2" denotes either the quasar or, for the case of same-beam interferometry, a second spacecraft. Even though quasar observations provide a sub-nanosecond precision determination of the sum of instrumental group delays and station clock offsets, additional calibrations are required to provide separate estimates of each. The first term in eq. (2) is generally small for the case of radio sources with small angular separations. But, for same-beam interferometry involving planetary orbiters, the second term in eq. (2) can become large if the two spacecraft are in substantially different orbits.

Frequency Instability: Doppler and Range

For a 2-way doppler measurement there are two time scales involved: the averaging interval $(t_{k+1} - t_k)$ over which the measurement is made and the round trip light time (RTLT). Suppose RTLT = $M(t_{k+1} - t_k)$, and let $\tau = t_{k+1} - t_k$. Then, if $\phi(t_k)$ is the station reference signal error (sec) at time t_k, the resulting error in the doppler observable at time t_k is given by

$$\varepsilon_{dk} = \overline{y}_k - \overline{y}_{k\text{-}M}\,,$$

where the fractional frequency deviation is defined by $\overline{y}_k = [\phi(t_{k+1}) - \phi(t_k)]/\tau$. The expected value of the squared error is given by a general form of the Allan variance [Barnes *et al.*, 1971],

$$\langle \varepsilon_{dk}^2 \rangle = 2\,\sigma_y^2(2, M\tau, \tau)\,.$$

Usually, the important quantity in navigation and radio science is how an error changes over a data arc. If $\varepsilon_{rk} = \phi(t_k) - \phi(t_k - RTLT)$ is the error in a range measurement, then the error in the difference of two range measurements separated by $L\tau$ is given by

$$\varepsilon_{r,k+L} - \varepsilon_{rk} = [\phi(t_{k+L}) - \phi(t_{k+L} - RTLT)] - [\phi(t_k) - \phi(t_k - RTLT)]$$

There are again two time scales involved: the time separation between the range measurements and the round trip light time. The expected value of the squared error is given by

$$< (\varepsilon_{r,k+L} - \varepsilon_{rk})^2 > = 2 \tau^2 \sigma_y^2(2, M\tau, \tau) ,$$

where $\tau = RTLT$ and $M\tau = t_{k+L} - t_k$ if $t_{k+L} - t_k \geq RTLT$, or $\tau = t_{k+L} - t_k$ and $M\tau = RTLT$ if $t_{k+L} - t_k \leq RTLT$.

Both range difference and doppler errors are related to the general form of the Allan variance, $\sigma_y^2(2, M\tau, \tau)$. This form may be approximated in terms of the regular form of the Allan variance by successive segmentation [Wu, 1980],

$$2 \sigma_y^2(2, T, \tau) \approx 2 \sigma_y^2(T) + \sigma_y^2(T/2) + \ldots + \sigma_y^2(T/2^N),$$

where the number N of segmentations is defined by $T/2^{(N+1)} < \tau \leq T/2^N$. This expression is convenient for calculations, since Allan variance measurements of frequency standard fluctuations (and other stochastic error sources) are often available.

Frequency Instability: VLBI

The error in a ΔVLBI observable due to instability of the station frequency standards is estimated by

$$\varepsilon_{\Delta VLBI} = \sqrt{2}\, T\, \sigma_y(T) ,$$

where T is the elapsed time over which the measurement is made and $\sigma_y(T)$ is the root square sum of the Allan standard deviations of the frequency standards at the two stations. Same-beam interferometric measurements are not affected by instability of the receiver frequency standards, since these effects cancel perfectly when differencing simultaneous measurements.

Error Budget: Doppler

Clock epoch and rate offsets produce systematic errors in spacecraft tracking observables, while frequency instability causes a random error. To put frequency and timing errors into perspective, it is necessary to look at complete measurement error budgets. Systematic errors are caused by uncertainty in tracking station locations, by uncertainty in the orientation of the Earth in inertial space, and by uncertainty in the static tropospheric delay. These errors are described by multiplying the parameter uncertainty times a geometric mapping function. Random errors are caused by fluctuating tropospheric delay, by fluctuating ionospheric delay after removal of a nominal calibration, and by solar plasma fluctuations. These errors may be characterized by a power spectral density function in the frequency domain, or by an Allan variance in the time domain, or by a temporal or spatial structure function. Thermal noise, which depends on the ratio of signal power to system noise power, contributes white phase noise to each measurement. Instrumentation can also contribute significant errors to tracking observables. Instrumental effects are generally treated in the same way as frequency and timing effects.

Error budget assumptions used in this analysis are summarized in Table 1. These assumptions are based on state-of-the-art media calibrations and platform parameter determinations. The 1 μsec clock epoch offset which is

Table 1. Error budget assumptions.

Thermal noise		0.04 cm (carrier phase)
Instrument bias		500 cm (range)
Station locations and Earth orientation		3 cm per component
Troposphere		0.5 cm zenith bias 0.5 cm line-of-sight fluctuation
Ionosphere		3 cm cal accuracy at X-band 0.2 cm at Ka-band eliminated by dual frequency
Solar plasma		300 cm total, 32 cm drift/8 hr at 90° SEP, X-band 8 cm total, 1 cm drift/8 hr at 180° SEP, Ka-band eliminated by dual frequency
Clock	epoch	1 μsec
	rate	10^{-13}
	stability	10^{-15} at 1000 sec

assumed represents the offset between the observable time-tags and true time. The effect of frequency instability is estimated from an Allan variance curve, rather than from the single representative specification given in the table. Fig. 4 is a plot of the Allan standard deviations of current and proposed frequency standards that are used in spacecraft and at the deep space tracking stations. The curve labeled "H2 Maser - 1991" is an estimate of the current performance of the hydrogen masers in use at the DSN ground stations. Improvements to the hydrogen masers have been proposed to support the Cassini gravity wave experiment in 1998 [Riley *et al.*, 1990]; the expected performance after the improvements are in place is given by the curve labeled "H2 Maser - 1998". The stability of the Mars Observer flight oscillator, which is referred to as an ultra-stable oscillator (USO), is also shown. The final curve shows the estimated performance of a proposed flight oscillator which consists of a crystal that is frequency locked to a linear trapped ion resonator (LIT).

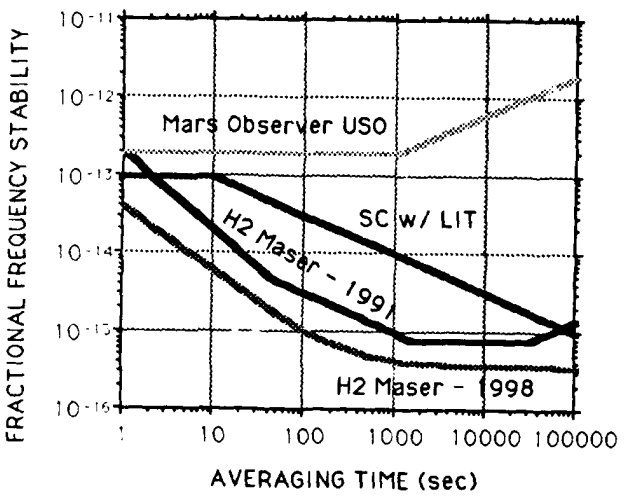


Fig. 4. Allan standard deviations of selected frequency standards, current and proposed, used on-board spacecraft and at deep space tracking stations.

The components of the doppler measurement error budget are shown in Fig. 5 for a spacecraft in interplanetary cruise at a distance of 5 AU (5000 sec RTLT). Errors, expressed in units of range, are shown as a function of time past the beginning of a tracking pass. For the effects of frequency and timing, the accumulated errors past the beginning of the pass are shown; the doppler observable is not sensitive to an initial phase bias. The curve labeled "clock epoch" shows the effect of an offset in the observable time-tags. The curve labeled "frequency stability" applies to 2-way observables, and is based on the current stability of the DSN hydrogen masers. For 1-way observables, the effect of frequency instability is dominated by the spacecraft oscillator. The curves labeled "Mars Observer USO" and "SC w/ LIT" give estimates, respectively, of the errors for the Mars Observer flight oscillator and for the proposed linear ion trap flight oscillator.

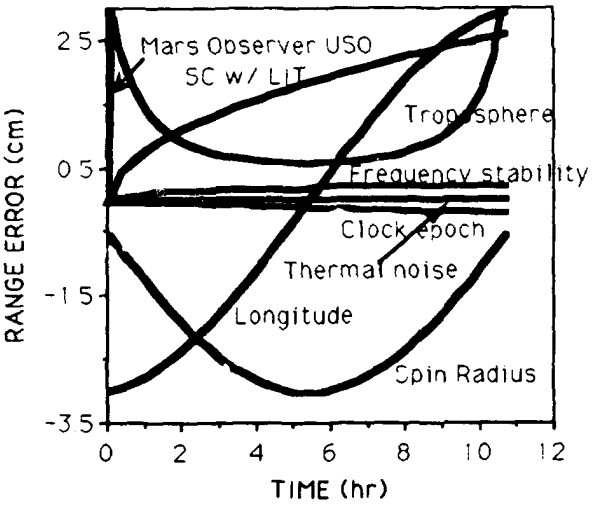


Fig. 5. Measurement error budget for the spacecraft to ground station doppler link. The effects of frequency instability are shown for both 1-way and 2-way observables. Charged particle effects are not shown.

The effects of uncertainty in station coordinates and static tropospheric delay appear quite systematic, and are about an order of magnitude larger than effects due to frequency and timing for 2-way observables. For navigation purposes these signatures are significant, since it is generally not possible to separate errors of this form from errors in spacecraft state. So media and platform effects dominate the 2-way doppler navigation error budget. For 1-way doppler using a spacecraft oscillator comparable to the Mars Observer flight unit, the error budget is completely dominated by spacecraft oscillator instability. Errors grow to the meter level within one hour. But if an oscillator is used which is comparable to the proposed linear ion trap flight unit, then 1-way tracking becomes very competitive with 2-way tracking for the purpose of navigation.

The error due to clock epoch offset is more significant for a planetary orbiter. This error is a function of the range-rate, which can be much more variable for an orbiter than

for a spacecraft in cruise. If only one tracking station were involved, then a clock offset could be absorbed into the spacecraft state. But, typically, orbits are determined by combining data from several tracking passes at different stations. For the Magellan orbit, an epoch offset between stations of just one μsec will introduce cm level residuals in the doppler.

For radio science purposes, systematic effects due to platform and static media errors can usually be filtered out, so that frequency and timing effects become much more important. The 2-way mode provides the most accurate measurements. Frequency and timing effects should be compared with fluctuations in media delays.

Charged particle effects are not shown in Fig. 5. For tracking at X-band only, ionospheric and solar plasma errors completely dominate the 2-way doppler error budget (see Table 1). But for tracking at Ka-band, the effects of charged particles are comparable to the other error sources plotted in Fig. 5. For dual frequency X/Ka downlinks, charged particle errors are reduced further, while they are virtually eliminated for dual frequency X/Ka uplinks and downlinks.

Error Budget: VLBI

The error budget for spacecraft-quasar ΔVLBI measurements is shown in Fig. 6. The current operational system, which uses an instantaneous bandwidth of 250 KHz, is limited in accuracy by system noise on the quasar measurements. Assuming media and platform calibrations as in Table 1, the expected measurement accuracy is 7 cm. This corresponds to about 10 nrad angular position accuracy using the DSN intercontinental baselines. More sensitive instrumentation could improve accuracy to the 1 cm level [Treuhaft, 1988]. The quasar observations reduce the effects of clock offset and instability to an insignificant level. If interferometric spacecraft measurements were to be made without the benefit of a nearly simultaneous quasar calibration, then clock offset would completely dominate the error budget. Fig. 6 also shows the estimated accuracy for spacecraft-only measurements, assuming externally supplied clock synchronization at the 1 nsec level. Though errors due to media delay and platform parameter uncertainty have grown by as much as a factor of five, clock synchronization is by far the largest error source.

The effects of frequency instability completely cancel for same-beam interferometry, since the observable is generated by differencing simultaneous phase measurements from two spacecraft. The SBI observable is sensitive, however, to instrumental group delays and clock epoch offsets, as shown in eq. (2). Instrumental phase delays depend on the signal doppler shift. The net effect is small for two spacecraft with nearly the same doppler, such as a lander and rover, but a large effect may remain for two spacecraft in different orbits. An uncalibrated station clock offset of just 100 nsec would be the dominant error source in X-band SBI measurements obtained from the Magellan and Pioneer Venus orbiters [Border *et al.*, 1991]. To reduce this error to an insignificant amount, absolute station instrumental delays must be calibrated to the 10 nsec level, so that clock offsets and instrumental delays can be separately extracted from either quasar VLBI or external clock synchronization reports.

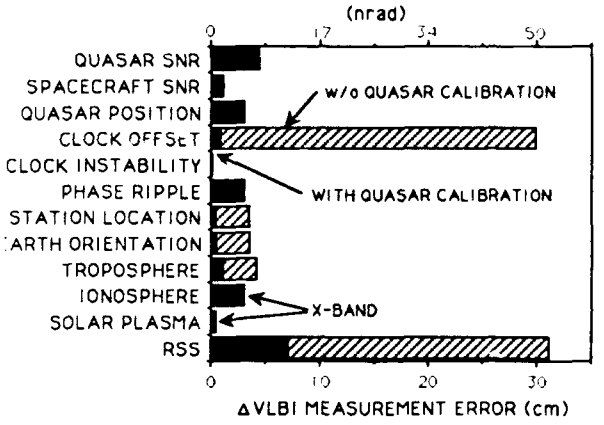


Fig. 6. Spacecraft VLBI measurement system errors both with and without nearly simultaneous quasar calibrations.

FREQUENCY AND TIMING: DESIRED IMPROVEMENTS

The performance of either ground station or spacecraft frequency standards is currently the limiting factor for some applications of deep space tracking for both navigation and radio science purposes. This section presents examples of tracking techniques which would be improved by advances in frequency and timing systems.

Simultaneous Tracking of Multiple Spacecraft

The initiative to explore the planet Mars will likely result in the deployment of several spacecraft all operating simultaneously in the vicinity of Mars. Included will be landers and rovers on the surface, communications and mapping orbiters, rendezvous craft, and possibly balloons, remote weather stations, and navigation beacons. All signals from spacecraft within areosynchronous orbit fall within the same beamwidth of an Earth-based 34 meter diameter antenna at X band. Operations are greatly simplified if all tracking is done in the 1-way mode, because no uplinks are required, and all telemetry and doppler data from all spacecraft may be acquired simultaneously at just one station. More accurate navigation is possible if doppler is supplemented with same-beam interferometry, by the use of antennas at two DSN complexes during periods when Mars is mutually visible.

Tracking accuracy in this case is limited by the performance of the spacecraft oscillators. For either doppler-only data sets or for doppler plus SBI, 2-way tracking provides more accurate orbit determination than 1-way tracking, if spacecraft are equipped with an oscillator comparable in stability to the Mars Observer flight unit [Folkner and Border, 1990]. Improving spacecraft oscillators to have stability comparable to the proposed linear trapped ion flight unit illustrated in Fig. 4 would make 1-way tracking just as accurate as 2-way tracking, for the purposes of navigation, for either doppler-only or doppler plus SBI data sets.

Spacecraft VLBI without Quasars

Inter-station time and frequency synchronization is necessary to support spacecraft interferometric observations. To be useful for navigation, the minimum requirement is that station clock offsets must be known to 1 nsec and clock rates must be known to 10^{-13}. In fact, clock epoch and rate offsets will be the dominant error sources unless they are determined at least ten times more accurately than this requirement. Synchronization is currently achieved by observing natural radio sources either just before or just after the spacecraft. Operations would be simplified if synchronization could be maintained by other means.

If station instrumental stability is sufficiently good, then synchronization can be maintained by infrequent quasar VLBI observations. For example, given stability of 10^{-15} for a time interval of one week, the synchronization error is estimated to be $\sqrt{2}$x(1 week)x(604800 sec/week)x 10^{-15} = 0.9 nsec, which is adequate to meet the minimum requirement. Alternatively, synchronization could be maintained by making more frequent observations of Earth orbiters, such as satellites belonging to the Global Positioning System. In either case, stability and synchronization must apply to the full instrumental path, not just to the primary frequency standards. All instrumental phase shifts and group delays must be calibrated. The prospects for maintaining knowledge of instrumental calibrations will greatly improve as upgrades within the Deep Space Network result in replacing analog receivers and electronic components with digital receivers.

Planetary Gravity Field Determination

Currently there is a conflict when tracking a spacecraft during a planetary flyby, or when tracking a planetary orbiter: 2-way tracking is preferred for gravity field determination, while 1-way tracking is preferred for ring and atmospheric occultation measurements. Improved stability of spacecraft on-board oscillators would reduce the conflict, perhaps making 1-way measurements sufficiently accurate to meet both requirements. However, improvements to ground station oscillators, coupled with improvements in media delay and platform parameter calibrations, could make 2-way tracking more accurate still, thus extending the conflict.

Recovery of Temperature and Pressure Profiles

The stability of the spacecraft on-board oscillator limits the accuracy with which temperature and pressure profiles may be recovered from measured phase shifts of 1-way signals which pass through an atmosphere. Improved stability extends the altitude range to enable measurements of more tenuous atmospheres. An order of magnitude stability improvement raises the maximum altitude by $\log_e 10$ atmospheric scale heights. As an example, consider the case of Mars. A Voyager-class oscillator (10^{-12} stability) provides useful information for an altitude range of from 0 to 37 km, while a Mars Observer-class oscillator (10^{-13} stability) provides useful information for an altitude range of from 0 to 60 km. Further, at a given altitude, an order of magnitude improvement in oscillator stability leads to an order of magnitude improvement in accuracy of the inferred temperature and pressure [Kursinski, 1990].

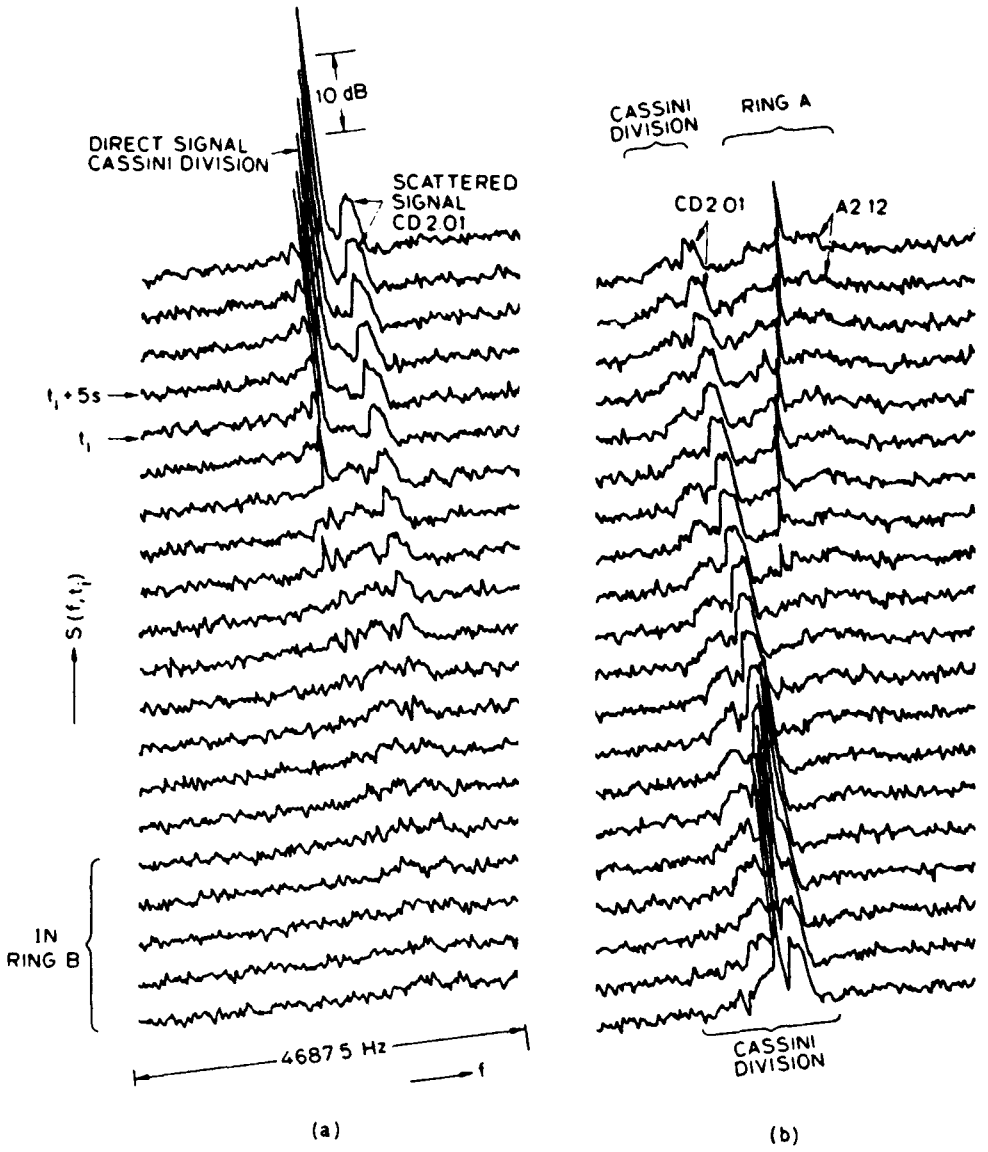


Fig. 7. Power spectra of Voyager 1 signal during Saturn ring occultation. Time sequence begins at lower left and progresses to upper right. (Figure from [Marouf *et al.*, 1983].)

Recovery of Size Distribution of Ring Particles

Forward scattering of a radio signal from ring features results in multiple paths for the signal propagation from spacecraft to tracking station. The power spectrum of the received signal contains information about each path. Information about particle size distribution is provided by recovering the scattered signal spectrum as a function of scattering angle, which is a known function of the geometry. The doppler shift of the scattered signal changes as the geometry changes, so that the scattered signal moves in frequency relative to the direct signal. Fig. 7 shows a sequence of power spectra of the Voyager 2 signal during the occultation by Saturn's rings. The scattered signal from the Cassini Division first appears higher in frequency than the direct signal, and then moves across the direct signal to a lower frequency. To extract maximum information from the scattered signal, it is necessary to keep oscillator phase noise below the level set by the carrier signal to noise ratio.

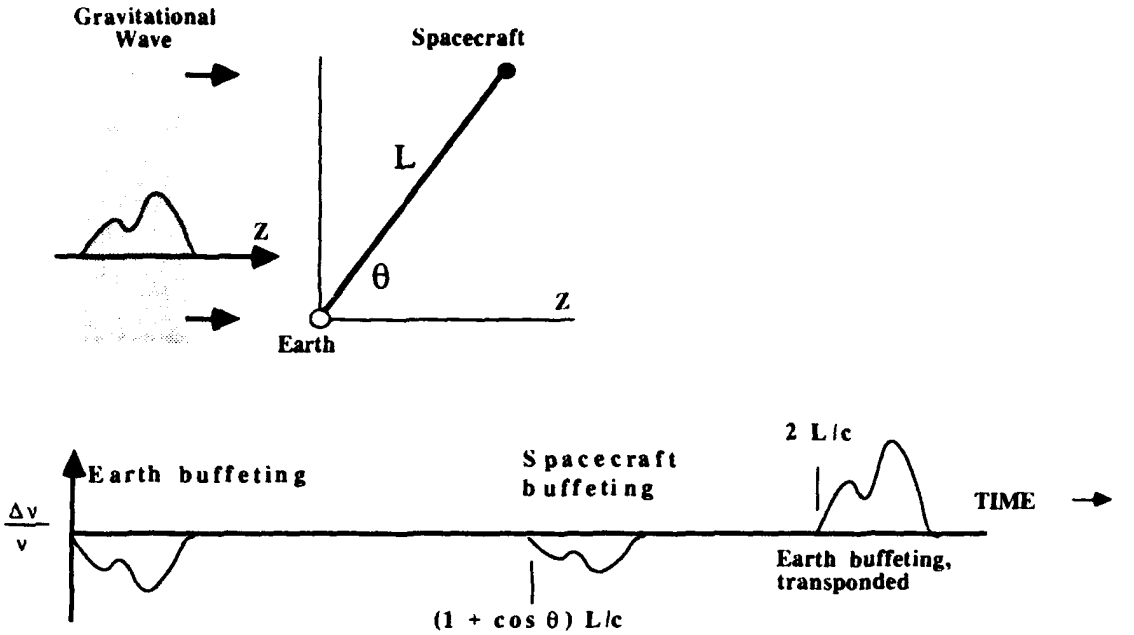


Fig. 8. Three pulse response to burst gravitational waves: time-resolved excitation produces a characteristic signature in the time domain.

Gravitational Wave Search

The DSN doppler navigation system has been extended to search for low frequency gravitational radiation, using a spacecraft in cruise in deep space. The onset of a gravitational wave causes a three pulse signature [Estabrook and Wahlquist, 1975] in the 2-way doppler data which is used to aid in detecting the signal and in determining a cone of direction from which the signal emanated. The sensitivity of this "detector" is characterized by its fractional frequency stability, $\sigma_{\Delta f/f}$. This instrument may be used to search for incident gravitational waves with dimensionless amplitude which is comparable to $\sigma_{\Delta f/f}$.

Fig. 8 illustrates the three pulse response to burst gravitational waves. The effect is first seen in the doppler record when the wave buffets the tracking station. The wave reaches the spacecraft at a later time which depends on the Earth-spacecraft distance and also on the angle between the Earth-spacecraft vector and the direction of propagation of the gravity wave. A one-way light time after the wave buffets the spacecraft, the effect is seen in the received doppler. The original Earth-buffeting effect is again seen in the doppler record a round-trip light time after the wave passed the tracking station.

The range of Fourier frequencies to which the instrument is sensitive is usually taken to be 0.1 to 1/RTLT Hz. At the highest frequencies the error is dominated by the signal-to-noise ratio. At the lowest frequencies there is partial cancellation of the three pulse signature resulting in reduced sensitivity. The goal for end-to-end system stability to support the Cassini gravity wave experiment is 10^{-15} at 1000 sec [Riley *et al.*, 1990]. To achieve this stability, improvements will be made to the DSN hydrogen masers, Ka-band uplinks and downlinks will be utilized, data will be acquired near solar opposition, and careful attention will be paid to calibration of atmospheric, electronic, and mechanical effects.

ACKNOWLEDGMENTS

P. W. Kinman and S. C. Wu of the Jet Propulsion Laboratory provided useful discussions of the effects of frequency instability on doppler measurements. The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

J. A. Barnes *et al.*, "Characterization of Frequency Stability," *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-20, No. 2, May 1971, pp 105-120.

J. S. Border *et al.*, "Determining Spacecraft Angular Position with Delta VLBI: The Voyager Demonstration," paper AIAA-82-1471, AIAA/AAS Astrodynamics Conference, San Diego, California, Aug. 1982.

J. S. Border, W. M. Folkner, R. D. Kahn, and K. S. Zukor, "Precise Tracking of the Magellan and Pioneer Venus Orbiters by Same-Beam Interferometry," paper AAS-91-191, AAS/AIAA Spaceflight Mechanics Meeting, Houston, Texas, February 11-13, 1991.

C. C. Counselman, III, H. F. Hinteregger, and I. I. Shapiro, "Astronomical Applications of Differential Interferometry," *Science*, Vol. 178, Nov. 10, 1972, pp 607-608.

D. W. Curkendall and S. R. McReynolds, "A Simplified Approach for Determining the Information Content of Radio Tracking Data," *Journal of Spacecraft and Rockets*, Vol. 6, No. 5, May 1969, pp 520-525.

F. B. Estabrook and H. D. Wahlquist, "Response of Doppler Spacecraft Tracking to Gravitational Radiation," *General Relativity and Gravitation*, Vol. 6, No. 5, 1975, pp 439-447.

G. Fjeldbo, A. J. Kliore and V. R. Eshleman, "The Neutral Atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments," *The Astronomical Journal*, Vol. 76, No. 2, March 1971, pp 123-140.

W. M. Folkner and J. S. Border, "Orbiter-Orbiter and Orbiter-Lander Tracking using Same-Beam Interferometry," paper AIAA-90-2906, AIAA/AAS Astrodynamics Conference, Portland, Oregon, Aug. 1990.

E. R. Kursinski, "Application of High Stability Oscillators to Radio Science Experiments Using Deep Space Probes," *Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, NASA Conference Publication 3116, Vienna, Virginia, Dec. 1990, pp. 253-268.

G. F. Lindal *et al.*, "The Atmosphere of Uranus: Results of Radio Occultation Measurements with Voyager 2," *Journal of Geophysical Research*, Vol. 92, No. A13, December 30, 1987, pp 14987-15001.

E. A. Marouf *et al.*, "Particle Size Distributions in Saturn's Rings from Voyager 1 Radio Occultation," *Icarus*, Vol. 54, 1983, pp 189-211.

E. A. Marouf, G. L. Tyler, and P. A. Rosen, "Profiling Saturn's Rings by Radio Occultation," *Icarus*, Vol. 68, 1986, pp 120-166.

W. G. Melbourne, "Navigation Between the Planets," *Scientific American*, Vol. 234, No. 6, June 1976, pp 58-74.

W. G. Melbourne and D. W. Curkendall, "Radio Metric Direction Finding: A New Approach to Deep Space Navigation," AAS/AIAA Astrodynamics Specialist Conference, Jackson Hole, Wyoming, Sept. 1977.

V. J. Ondrasik and K. H. Rourke, "Applications of Quasi-VLBI Tracking Data Types to the Zero Declination and Process Noise Problems," paper 71-399, AAS/AIAA Astrodynamics Specialist Conference, Fort Lauderdale, Florida, Aug. 1971.

S. R. Poole, M. P. Ananda, and C. E. Hildebrand, "Radio Interferometric Measurements for Accurate Planetary Orbiter Navigation," AAS Advances in the Astronautical Sciences, Vol. 46, Part 1, Univelt Inc., San Diego, 1982, pp 163-179.

A. L. Riley *et al.*, "Cassini Ka-Band Precision Doppler and Enhanced Telecommunications System Study," report jointly sponsored by the National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California and by Agenzia Spaziale Italiana, Rome, Italia, (internal report), Jan. 22, 1990.

R. N. Treuhaft, "Deep Space Tracking in Local Reference Frames," *The Telecommunications and Data Acquisition Progress Report* 42-94, April-June 1988, pp 1-15.

G. L. Tyler *et al.*, "The Microwave Opacity of Saturn's Rings at Wavelengths of 3.6 and 13 cm from Voyager 1 Radio Occultation," *Icarus*, Vol. 54, 1983, pp 160-188.

S. C. Wu, "On Estimating the Effects of Clock Instability with Flicker Noise Characteristics," *Proceedings of the Twelfth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, NASA Conference Publication 2175, Goddard Space Flight Center, Greenbelt, Maryland, Dec. 1980, pp 681-702.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INTERIM REPORT ON THE GEOGRAPHIC DEPENDENCY AND LATITUDE EFFECTS STUDY

R. BEARD, A. GIFFORD, and S. STEBBINS
U.S. Naval Research Laboratory

S. RASMUSSEN
Department of Defence

T. BARTHOLOMEW
The Analytical Sciences Corporation

Abstract:

The Geographic Dependency and Latitude Effects Study is an effort underway at NRL to investigate the capability of the Global Positioning System (GPS) to disseminate Universal Coordinated Time as maintained by the U.S. Naval Observatory, UTC(USNO). It is assumed that the users will have access to the GPS Precise Positioning Service (PPS) and will use GPS in the passive mode. That is, they will not exchange data with another location to transfer time by the common view method, but rather will obtain UTC(USNO) via the GPS broadcast message. The study addresses the performance expected from this type of user on a global basis. USNO has demonstrated, through operation at their facility as a user, that the system elements can provide time to less than 25 ns. This seems to indicate that the final link with the user is the source of additional errors limiting performance. The other GPS system elements and the worldwide nature of operation for time dissemination has not been well studied and should not be overlooked. Since USNO, the GPS Master Control Station, and the primary GPS test range at Yuma are co-visible, the southern pacific area was chosen as the area of special interest to this study, being the more difficult for operational coverage.

In this study, the effects of the various GPS system error sources are being investigated for their influence on time dissemination. Data is being gathered from a variety of tracking and timing centers to attempt to identify and quantify the errors and their sources. The known influences of GPS satellite position error, propagation effects and the ability to maintain time at remote sites are being investigated. Trends in the error propagation of these different effects will be determined and limits imposed on capability and possible geographic dependency identified. Interim results and tentative conclusions will be presented on these efforts.

INTRODUCTION

This paper will discuss an investigation into the errors associated with recovering Universal Coordinated Time as maintained by the Naval Observatory, UTC(USNO), by authorized users of the Global Positioning System (GPS). The eventual aim of the effort is for users who will have access to the GPS Precise Positioning Service (PPS) and will use GPS in the passive mode. That is, they will not exchange data with another location to transfer time by the common view method, but rather will obtain Universal Coordinated Time and maintained by the U.S. Naval Observatory (UTC(USNO)) via the GPS broadcast message. Consequently, the best accuracy from the system in this mode is desired. The identification of the error sources and their contribution to GPS time dissemination is the objective, with the aim of reducing or eliminating their effect.

GPS SYSTEM OVERVIEW

The GPS system is currently being deployed to provide a general purpose navigation, precise positioning and time dissemination service for military users. The system provides multi-satellite coverage over the entire globe and a combination of navigation signals. GPS is intended to be used as a real-time system for ground and air military platforms requiring accuracies to 10's of meters. The real-time requirement is met by using the coded signals to make near-instantaneous measurements to the GPS satellites. For fixed users, smoothing permits improved accuracies down to less than a meter. The Selective Availability/Anti-Spoofing (SA/AS) capability of the GPS system is

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

the basic difference between the Standard Positioning Service (SPS) and the Precision Positioning Service (PPS). The SPS is meant to be the standard service available to civilian and foreign users.[1] The PPS represents the full military capability of the system which must be accessed through secure cryptological means. These two schemes degrade or deny the system assets or accuracy for the Clear/Acquisition (C/A) only user. One aspect of SA is oscillator dither which dithers the basic clock frequency used to generate the C/A and Protected (P) code signals. The dither is an encrypted sequence which must be decoded by the user to obtain full accuracy and precision from the GPS signals. Anti-Spoofing (AS) provisions change the P code to another code known as the Y code.

GPS RECEIVERS

Receivers developed for use with GPS range from using only L_1 - C/A code only with one to five channels, a channel is designated for each GPS satellite being observed, to the full 5 channel military receiver. Sequential receivers have also been developed which sequence through a number of GPS satellites to provide the number observable for navigation. The most precise receivers developed for geodetic work operate predominately with the carrier frequencies. The coherent frequencies provided by the GPS satellites are used as a fine measure of the satellite range and range rate. This eliminates reliance on the less precise ranging information available from the signal code rates. For the timing user a single channel receiver is normally used since their are usually at a fixed, known location. If the position is not known accurately then GPS can be used to fix, or navigate, the receiver.

To solve for real-time position in the normal mode used with GPS, four satellite simultaneous code data is used. With a single channel receiver multiple measurements would be used to provide the solution. The code data measures the time delay a signal takes to propagate from the satellite to the user. The user can determine the range by, $R = c\,(t_2 - t_1)$, where, R is the range, c is the speed of light, t_2 is the time of reception, and t_1 is the time of transmission. Ranging to three satellites simultaneously allows the determination of the user's three position coordinates. This process presumes that the satellites' and the user's receiver clock are synchronized, which is normally not the case. For this reason, the range determination is called pseudoranging, which denotes unsynchronized clocks between the satellites (normally assumed to be synchronized) and user receiver (offset by some value b_u), and requires an additional satellite measurement to be able to solve for the synchronization error. The user offset is the measurement desired by the timing user. The pseudorange equation between the user, u, and multiple satellites (or measurements), i for i = 1, 4, is then,

$$(b_u + \Delta t_i)\,c = \rho_i + c\,(b_i + \Delta t_n)$$

where, $(b_u + \Delta t_i)c$ = pseudorange,
ρ_i = true range to satellite i,
b_i = satellite clock offset, and
Δt_n = propagation biases and noise.

Alternatively, this equation may be written as a function of the user's position, denoted as x_u, y_u, z_u, and the user clock offset, b_u.

$$(b_u + \Delta t_i)\,c = \sqrt{(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2} + c\,(b_i + \Delta t_n)$$

The solution of this equation is the navigation solution and the basis for the NAVSTAR system performance estimates. If the equation is rewritten in matrix form (cartesian coordinates assumed) as,

$$(R_u \cdot e_i - B_u) = (R_i \cdot e_i + B_i) - P_i + v_i$$

or,

$$\bar{G} \cdot \bar{x} = \bar{A} \cdot \bar{s} - P_i + v_i$$

where; R_u = user's position vector,
e_i = unit vector in direction of satellite i,

$B_u = c\ b_u$, range equivalent user time bias,
$P_i = c\ \Delta t_i$, measured pseudorange vector,
$B_i = c\ b_i$, range equivalent satellite time bias,
$v_i = c\ \Delta t_u$, range equivalent noise term.

and, the noise characteristics are of the form,

$$Cov(v_i) = E[v_i v_i^T] = \sigma^2 I$$

where, σ is the error in the individual ranging link. Then, the estimate of the parameter vector, x_i, is,

$$\hat{x}_i = [G_T Cov^{-1}(v_i) G]^{-1}\ G^T\ [Cov^{-1}(v_i)]\ [\bar{A} \cdot \bar{s} - P_i]$$

and covariance of the estimated solution is,

$$Cov(\delta \bar{x}) = [G^T G]^{-1}\ G^T\ Cov(v_i)\ [(G^T G)^{-1}\ G^T]^T$$

and, from the noise characteristics above,

$$Cov(\delta \bar{x}) = [G^T G]^{-1}\ \sigma^2$$

The navigation accuracy estimate can be then described as the product of the basic ranging error to a satellite, σ, and a dimensionless constant, $[G^T G]^{-1}$, called GDOP for Geometric Dilution of Precision. GDOP is determined by the specific geometry of the GPS satellites to the user at the time of navigation solution. The basic ranging error, which has been called the User Range Error (URE) or User Equivalent Range Error (UERE), is the combination of the errors in the measurement of range between each satellite and the user receiver. URE is a combination of the actual range measurement errors due to instrumentation, residual ionospheric and tropospheric correction, orbital position and clock errors. This is a fundamental link in the time comparison between GPS and the time user. The URE is strongly influenced by both clock and orbit errors.

ERROR SOURCES

The URE and associated error contributions are illustrated in Table 1 (actual error contributions are highly dependent upon the equipment used). The dominant errors appear to be atmospheric delays and multipath. However, the GPS satellite position and clock errors are functions of elapsed time since the last clock update or orbit prediction and represent the primary limiting factors on GPS system performance.

User error statistics are given and discussed in ICD-GPS-200[2] as a measure of system performance. In that document the user error is defined as URA (User Range Accuracy) and the difference between the Control Segment's prediction and the current Kalman estimate.

Table 2. GPS User Range Error

Error Component	1σ(m)
Satellite Orbit Error	1.5
Satellite Clock Error	1.5
Atmospheric Delays	2.4 - 5.2
Satellite Equipment Delays	1.0
Multipath	1.2 - 2.7
Receiver Noise and Resolution	1.5
TOTAL (RSS)	3.6 - 6.3

GPS Orbital Position Errors.

The GPS system relies on an accurate knowledge of the predicted satellite positions. Knowing the satellite position then permits accurate range determination to allow calculation of the user's position or passive time comparison. Since the position must be know in advance, the orbit predictability of the GPS satellites is the key factor in the operation of the system. A precision orbit determination process has been enacted by the Defence Mapping Agency (DMA) to provide the

most accurate satellite ephemerides for surveying and mapping purposes. That non-real-time process provides to most accurate positional information available for comparison and estimation of the predicted positional information broadcast to the users.

Satellite Clock Errors

The GPS satellite atomic clocks time offsets and rates are also predicted for real-time use. The coefficients broadcast in the navigation message are linear predictions of the particular clock's performance based of the observations from the GPS monitor stations and other data computed at the GPS Master Control Station. These calculations are performed from the passively observed pseudorange information collected at the GPS Monitor Stations. In this passive process, the ability to distinguish between a clock error and positional error cannot be done by measurement. Consequently, the computational models are used to separate and estimate the various errors for prediction. This clock/position ambiguity and the inability to measure clock performance independently leads to errors. In the error budgets and operation of the system clock and orbit predictability are roughly equal in effective magnitude.

Propagation Errors

A GPS signals passing through the earth's atmosphere encounter two distinct regions which alter the propagation time and direction, the ionosphere and troposphere. Propagation through the ionosphere experiences an effective change in the refractive index caused by the free electron plasma which makes up the ionosphere. The ray path followed by the radio signal is expressed as,

$$\text{Path Length} = \int_{\text{satellite}}^{\text{user}} n\,ds$$

where, n = refractive index, and ds is along the ray path.

For GPS frequencies the actual bending of the signal path is small and the primary delay is caused by phase or group delay of the signal. The deviation of the refractive index from unity is then $\Delta n = 1 - n$, which may be expressed as,

$$\Delta n = \frac{e^2}{8\pi^2\epsilon_0 m}\frac{N}{f^2} = \frac{40.3}{f^2}N$$

where, f = frequency (Hz), and N = electron content (electrons/m^2).

The difference between the measured path length and the geometric path length is then,

$$\Delta P = \int \Delta n\, ds = -\frac{40.3}{f^2}\int N\, ds$$

The integral is taken along the signal path and represents the integrated electron density acting on the signal. By using two frequencies the integrated electron density function can be measured to first order. Higher order effects are not significant in general since the first order effects residuals are estimated to be on the order of 1-2 meters for high elevation angles.[3] In general, the expression can be used to represent the signal path electron density anywhere in the world against a worldwide model of the ionosphere. This approach is used by the single frequency user of GPS against a standardized model whose coefficients are transmitted in the satellite data message.

Signal errors caused by propagation through the troposphere are independent of the signal frequency. The refractive index of the medium due to troposphere (n_T), is expressed as,

$$n_T = \frac{77.6}{T}\left(B + \frac{4810e}{T}\right)$$

where, T = temperature (°K), B = total barometric pressure (millibars), and e = partial pressure of water vapor (millibars).

These two terms in this expression are the so-called "dry" (n_d) and "wet" (n_w) terms respective-

ly. The excess propagation path is given by,

$$P = 10^{-6} \int n_T \, ds$$

Since these terms are independent of frequency and dependent upon the metrological conditions along the signal path they must be modeled for the GPS user. Measurement of this effect is possible with water vapor radiometers which are large and expensive to be deployed for the modeling residual that results. A model would need to be applied to the GPS observations to account for these errors, which have been measured[4] for elevation angles over 30° to be on the order of 20 cm. For lower angles the effect increases significantly due to the longer integrated path through the troposphere.

Instrumentation Errors

The receiving equipment and its installation effects the performance of navigation and time comparison users. The usually problems of receiving with the appropriate resolution are common. An additional factor affecting the time user is the calibration of the delays in the antenna lines and other distribution networks that may be employed at the receiving site. A delay in the antenna network would not cause a positioning error, for the multi-satellite case as long as the delay were common to all measurements, but could cause a timing error for the time comparison user. It is therefore important that the receiver and its supporting instrumentation be calibrated for delay and such variable timing problems. Comparisons in reception time made with timing pulses, such as a 1 pulse per second time mark, should ensure rise time and pulse integrity for proper resolution and comparison.

Multipath into the receiving antenna is a substantial problem, especially near large reflecting surfaces. These effects are common and can be large error contributors.

GPS TIME TRANSFER

The time comparison user has another complication in that the offset from GPS Time to UTC(USNO) must likewise be known. The procedures for determining and providing the correction term to the GPS Control Segment for inclusion in the Navigation Broadcast message and earth orientation terms necessary for orbit determination are contained in ICD-GPS-202[5].

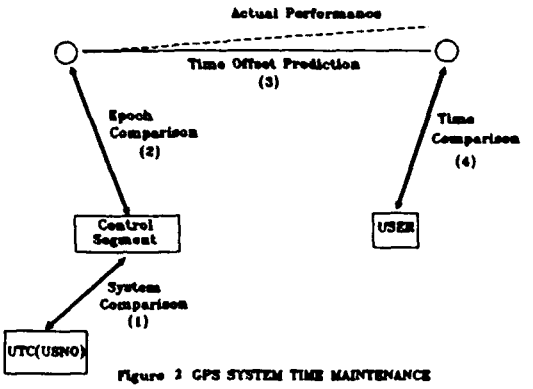


Figure 2 GPS SYSTEM TIME MAINTENANCE

There is an additional factor that limit UTC(USNO) time recovery from GPS. This factor is operational philosophy, GPS has no requirement for time dissemination. To maintain time recovery better than the 100 nanoseconds level is not an overall system requirement. The maintenance of GPS Time relative to UTC(USNO) is to be kept to one micro-second and knowledge of the difference to 100 nanoseconds is therefore subject to operational expediency. Even though it has been demonstrated that the system is capable of maintaining knowledge of UTC(USNO) to less than 25 nanoseconds. Figure 1 shows the value of UTC broadcast by the GPS satellites compared to UTC(USNO).

Figure 2 shows the processes involved in maintaining and disseminating time within the GPS system. Time recovery as discussed here is Time Comparison (4) in figure 2. This is a result of the overall process, in which time must be maintained throughout. Each of the steps in the process are possible sources of systematic and noise-like errors.

The errors as discussed above in final link with the user is the source of additional errors limiting performance. There are two types of residual errors in the link to the users. The first, as seen in figure 3, is systematic errors observed in a pass of data. The points shown are the broadcast time, relative to a station clock, for one satellite as it passes from horizon to horizon. This systematic effect has been called the "bowing effect". The effect is the most extreme using a single frequency C/A receiver but is apparent in data from dual frequency receivers as well. By smoothing individual 13 minute ranging data points over a satellite pass and many passes, USNO has been able to recovery UTC(USNO) on a daily basis from the GPS message as shown in Figure 1 to less than 25ns for long periods. It is important to note that this level of time recovery is achieved by processing all data (< 96 tracks per day), and that with limited data (< 10 per day), the time recovery is likely to be closer to 100ns than 25ns. The bowing effect appears to be dealt with in the short term of a few days by the smoothing process, but the mechanism of the effect is not well understood. The principal cause for this effect has been attributed to propagation effects. However, some data, Figure 4, suggests a ramping effect not easily attributable to propagation. Possible long term effects have not been fully investigated.

To investigate a possible long term noise-like effect, an ensemble of clocks is needed since the behavior could be below the noise level of a moderately well behaved cesium. Figure 5 is data taken over the same timeframe from reliable clock ensembles in the western pacific. The magnitude of the effect varies between them, but the frequency shifts appear to correlate. There are other sites in this region that show similar correlations. These data would normally be a measure of each site versus UTC(USNO) and the frequency performance would be considered to be due to individual sites. However, these sites are considered to be very stable. The correlation of the frequency shifts between the sites would generally tend to indicate that the movement was in either the GPS time base or UTC(USNO). USNO compared by common view to UTC(BIPM) shows no similarities with the western Pacific is shown during this period. The common view comparison should eliminate GPS time error contributions thereby tending to isolate the GPS time base as a possible source of the performance.

DATA COLLECTION

Data has been gathered for this investigation on actual performance for isolation of the error effects through the cooperation of USNO, the GPS Master Control Station, Defence Mapping Agency (DMA), Detachment C of the Naval Space Operations Center, the Australian Orroral Valley Observatory, and the Physics and Engineering Laboratory, Lower Hutt, New Zealand. These data included received GPS observations and data on the clocks used in data collection. Single frequency receivers were used in the overseas sites and a dual frequency at USNO. Precise ephemerides for the GPS satellites were provided by DMA and GPS broadcast data were obtained through the Cooperative International GPS Tracking Network computer center at the National Geodetic Survey.

Data collection from the various sources began 1 October 1990 with the intention of having coincident data from all sources. Prior to data collection site surveys were conducted at all sites to verify calibration and a clear understanding of all the parameters effecting the analysis of the data.

GPS TIME MEASUREMENT

The GPS time recovery by pseudoranges (PR)

time values at a stationary site can be expressed as,

$$t_{LC} - t_{GPS} = PR_M - PR_C$$

This produces the difference in clock values between the local clock (LC) used for the measured pseudorange (PR_M) and GPS time (t_{GPS}) as maintained by the satellite. The calculated pseudorange term (PR_C) is,

$$PR_C = c^{-1}P(x_G) - \Delta t_a + \Delta t_{SV},$$

$P(x_G)$ = Geometric range between SV position and receiver position, Δt_a = the ionospheric and tropospheric propagation correction, and $\Delta t_{SV} = t_{GPS} - t_{SV}$ (Actual time on the space vehicle (SV)) + Δt_{Rel} (a correction for the relativistic time delay accounting for non-spherical nature of actual satellite orbit). The SV time is related directly to UTC-(USNO), t_{UTC} by,

$$t_{UTC} - t_{GPS} = \Delta t_{GPS}$$

So that,

$$\Delta t_{SV} = t_{UTC} - \Delta t_{GPS} - t_{SV} + \Delta t_{Rel}$$

The value of t_{LC} is the final desired value of the difference between a local clock and GPS time or UTC(USNO).

In normal system operation, the calculated term for the pseudorange is based on the data broadcast by the system. To examine the error components the DMA post processed data was used with the broadcast data. The post-processed derived pseudorange (PR_{DMA}) with the SV position information and the SV clock correction term is,

$$PR_{DMA} = c^{-1}P(x_D) + \Delta t_D$$
$$= c^{-1}P(x_D) + (t_{GPS} + t_{SV})$$

where, $P(x_D)$ = Position from post-processed tracking data, and Δt_D = The clock correction term derived from the post-processed estimate of clock offset.

The broadcast derived pseudorange (PR_B) provides predicted position information, predicted clock correction, and propagation corrections for ionospheric and tropospheric delays, so that,

$$PR_B = c^{-1}P(x_b) + \Delta t_B + \Delta t_a$$

where, $P(x_B)$ = SV broadcast position, Δt_B = the broadcast SV clock correction term, which is the predicted time error (t_E) between SV clock and t_{GPS}, and Δt_a = the propagation correction term.

The measured pseudorange (PR_M) contains the composite terms,

$$PR_M = c^{-1}P(x_B) + \Delta t_B + \Delta t_{LC} + \Delta t_{Aa} + E$$

where, Δt_{LC} = local clock offset, Δt_{Aa} = actual propagation delay, and E = other residual error including instrumentation effects.

DATA COMPARISON

These measurement data sets may be combined for comparison and examination of the components. The pseudoranges from the broadcast and the post-processed should, in theory, be equivalent, as,

$$PR_B - PR_D = \Delta P(x_B - x_D) + c(\Delta t_B - \Delta t_D).$$

These data being a function of the position vectors and clock offset of the SV clock from GPS Time. Comparing the radial position component of these two data sets, which should be the most sensitive component, for GPS SV3 and SV16 are shown in figures 6 and 7. The residual error in these figures shows a significant error in the broadcast positions. These errors appear for all SV's examined, and longer data sets are to be investigated.

The SV clock comparison term above for SV 11, 16, 20 and 21 are shown in figures 8 to 11, respectively. These data are residuals to a linear fit to the post-processed data, and a linear term was used because the broadcast clock term are likewise linear. Higher order models will be used

in further comparisons. The spread in these data suggests some difficulty in estimating the clock term. The magnitude of the errors vary between the SVs, but the minimums appear to be on the order of 10-15 nanoseconds over the prediction interval. These data further suggest that the post-processed solution is quite noisy and not indicative of clock-like behavior. This behavior being indicated data from the USNO receiver was used for comparison and verification of the post-processed solution. One comparison is shown in figure 12. In these data the USNO appears to have a different daily rate, but overall, following the general trend of the clock. Further investigation showed that if the relativistic correction (for orbit eccentricity) were removed from the USNO observations the data shows a better correlation as shown in figure 13. The residuals of the data from figure 13 are shown in figure 14.

The orbital and clock comparisons shown in figure 15 demonstrate the correlation between the orbital position and clock error ambiguity present in this passive system. These data also show the overlap period used in the post-processing to prevent discontinuities between the orbit fit periods. in this overlap period, the clock residual differences at the same observation time are the result of the orbit fits made in the two successive weeks using largely different data. The variation between the two fits could be a significant error contributor coupled with other sources.

These comparisons suggest the presence of structured signals in the error components for orbit and clock offset, exclusive of propagation effects. These error signals that would be unseen by the timing user, directly effect the short term performance and possible introduce a long term feature.

Data taken by participating sites to contribute to component error resolution compared to UTC as broadcast by GPS are shown in figures 16 through 19. The long term structure in these data do not appear to correlate with GPS Time or UTC(USNO) variations, although the analysis is not complete. These sites are in general using an ensemble of clocks and the data is to be examined. Independent calibrations or comparisons are being developed for comparative analysis.

PRELIMINARY CONCLUSIONS/RESULTS

Data collection and analysis for this effort is continuing. Considerable analysis and data correlation is yet to be performed. However, the preliminary conclusions made are that the Bowing effect may have major contribution from a number of different contributors. Orbit and clock residual error signatures imposed on the pseudoranging residuals appear to be the large contributors. Analysis of the propagation factors which would classically be expected to produce a bowing type effect has yet to be fully investigated.

It does seem reasonable to conclude that the spread of data and error growth in orbit, clock and possibly propagation factors is larger as observed in the southern pacific area. Data spread and noise from the data there appears to bear this out. A spectrum analysis of the error structure is to be examined in looking for a long term error structure.

REFERENCES

[1] Federal Radionavigation Plan, Departments of Transportation and Defence, 1990.

[2] ICD-GPS-200 NAVSTAR GPS Space Segment /Navigation User Interfaces, Revision A, 26 September 1984 (Corrected).

[3] G.O. Ajayi, A. Hedberg, and G. Hamberg, "Accurate determination of ionospheric effects on satellite-based positioning systems in terms of residual range error," Radio Science, vol. 15, no. 5, pp 1009-1016, September-October 1980.

[4] J.M. Moran and B.R. Rosen, "Estimation of the Propagation Delay through the Tropo-

sphere from Microwave Radiometer Data," Radio Science, vol. 16, no. 2, pp 235-244, March-April 1981.

[5] ICD-GPS-202, NAVSTAR GPS Control Segment/U.S. Naval Observatory Time Transfer Interfaces, 21 November 1984 (Proposed edition)

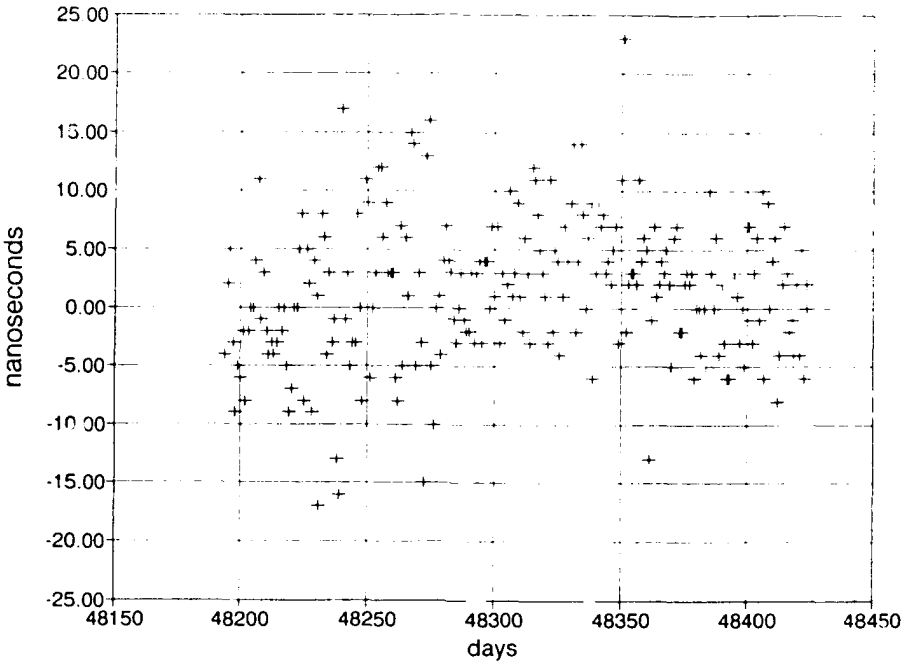


Figure 1, Universal Coordinated Time as Disseminated by GPS vs UTC(USNO)

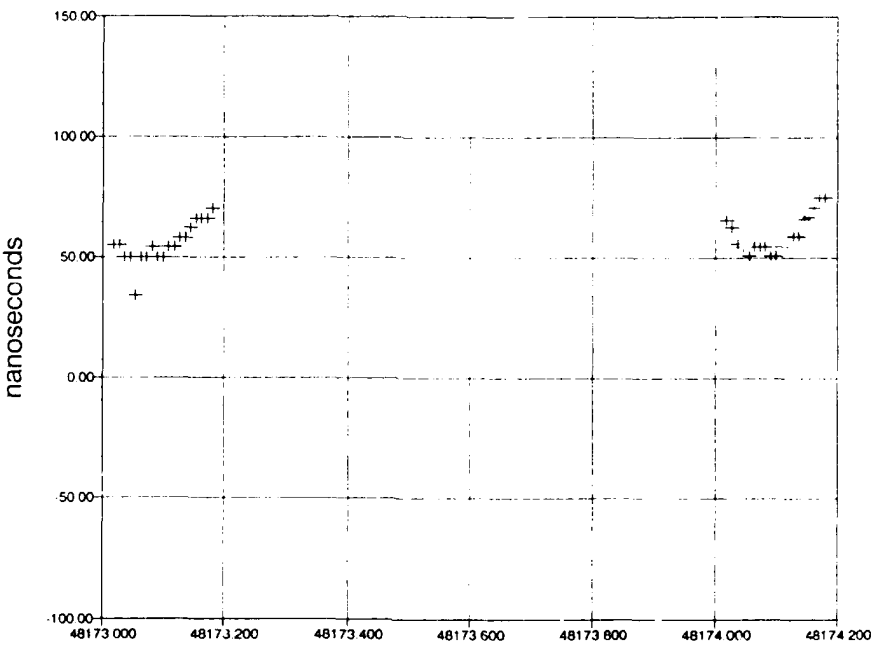


Figure 3, Single Frequency Receiver at GPS Master Control Station Observation of SV21

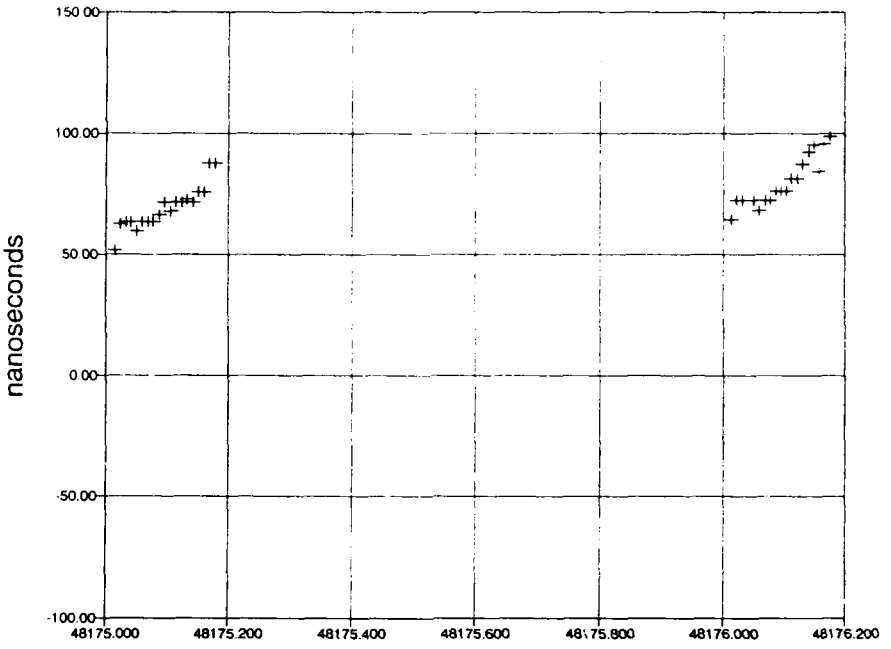


Figure 4, Single Frequency Receiver at GPS Master Control Station Observation of SV21

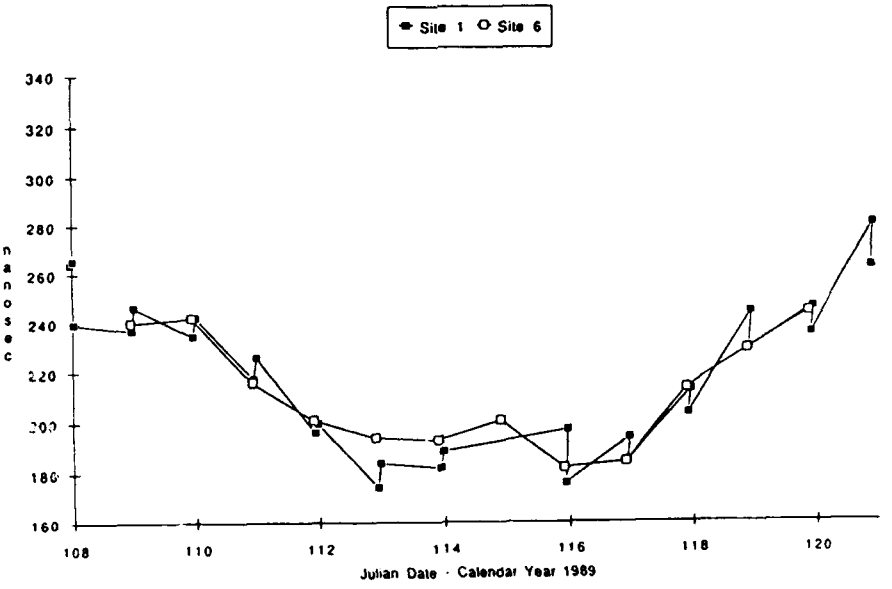


Figure 5, Observations of SV09 at Two Sites in the Southern Pacific

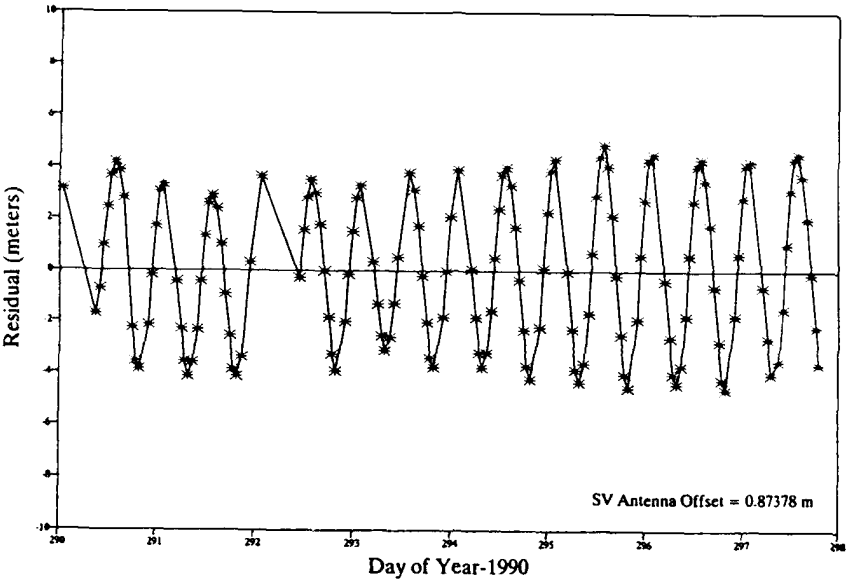


Figure 6, SV03 Radial Position Difference (Broadcast vs DMA)

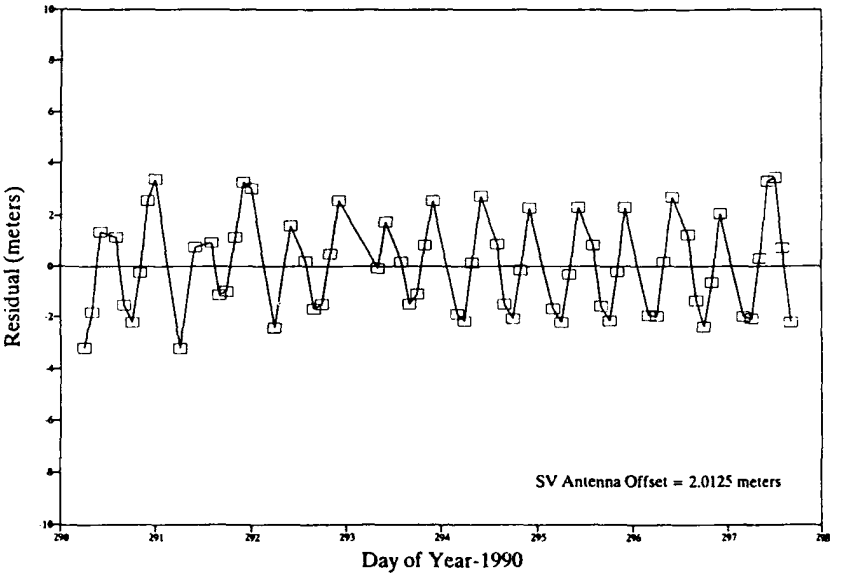


Figure 7, SV16 Radial Position Difference (Broadcast vs DMA)

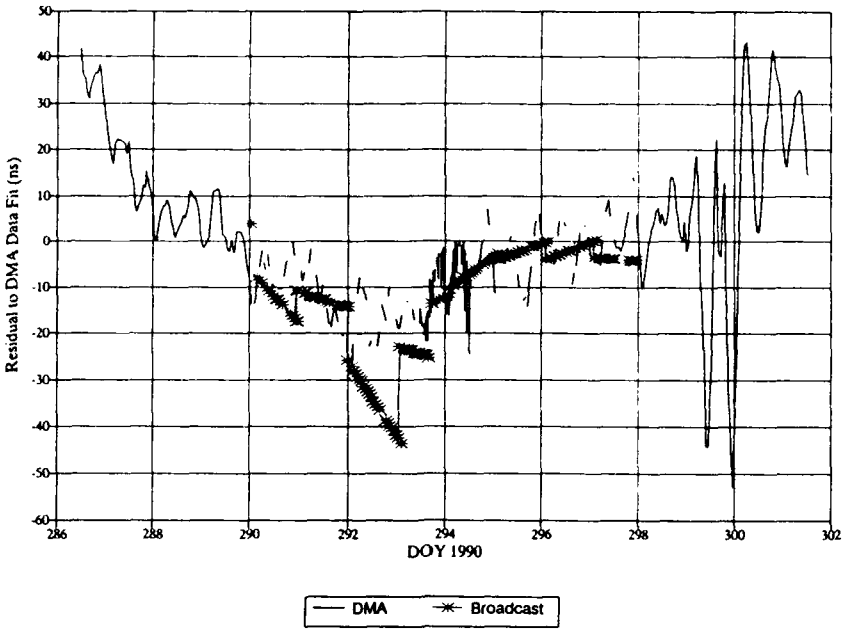


Figure 8, SV11 Clock Residuals (Broadcast vs DMA)

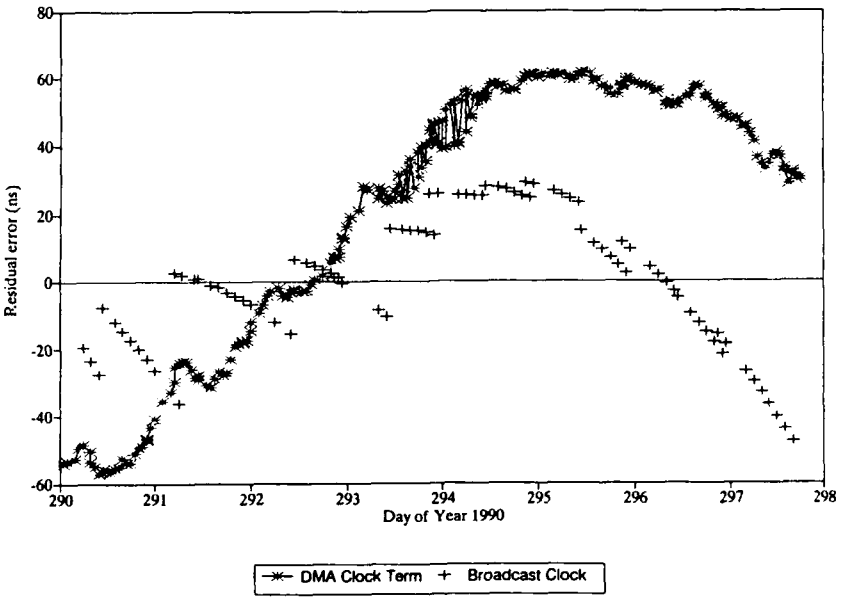


Figure 9, SV16 Clock Residuals (Broadcast vs DMA)

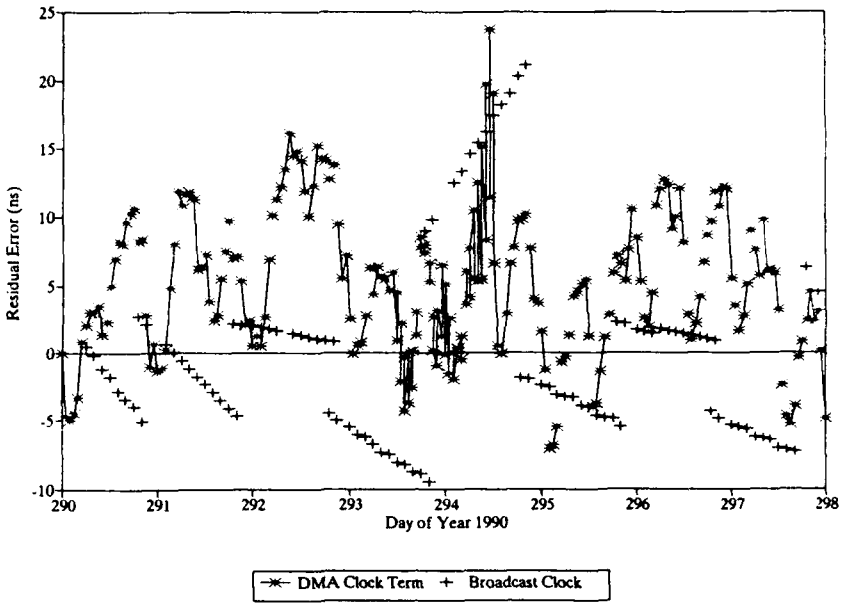


Figure 10, SV20 Clock Residuals (Broadcast vs DMA)

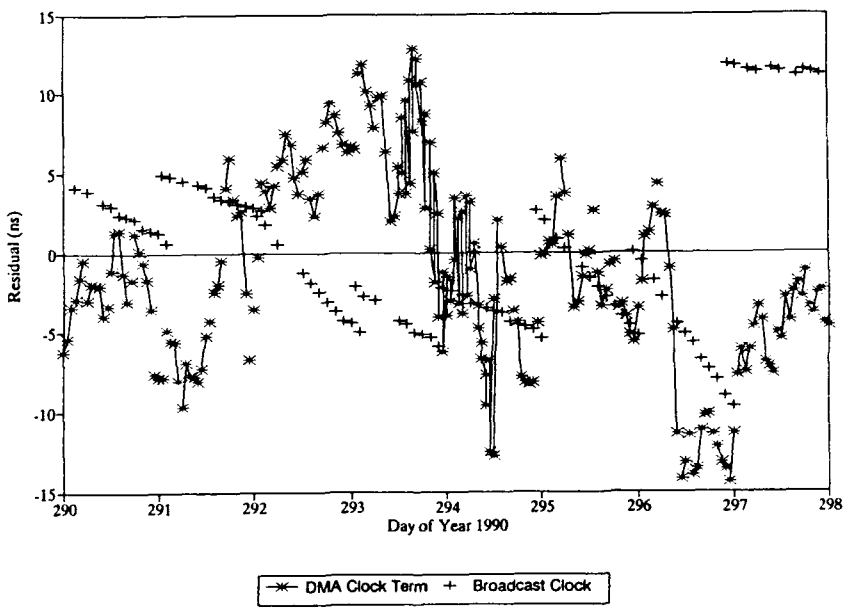


Figure 11, SV21 Clock Residuals (Broadcast vs DMA)

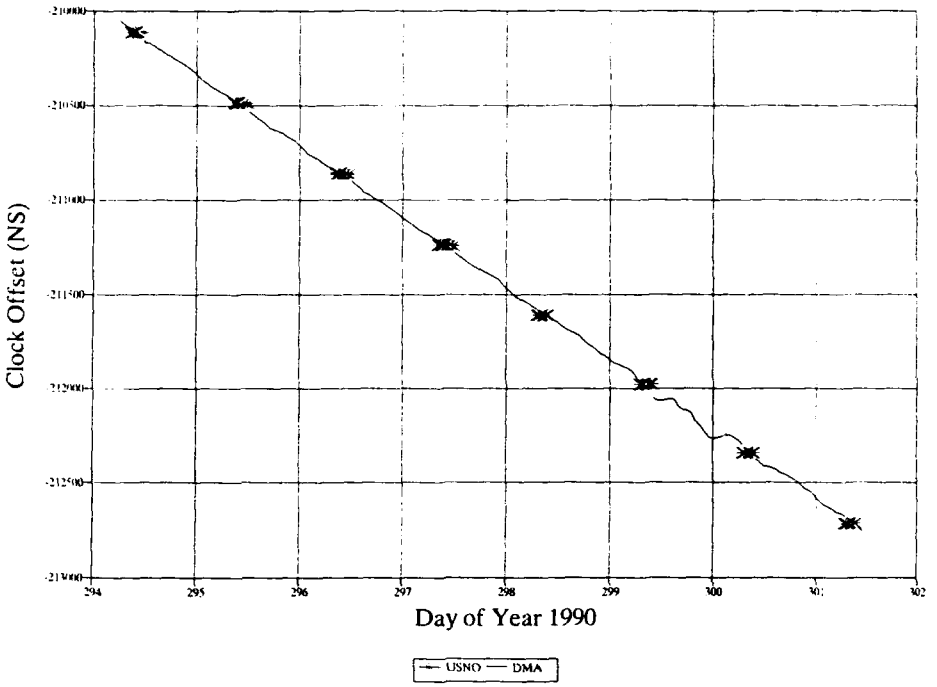


Figure 12, SV11 Clock Offset to GPS Time, DMA and USNO
Observations corrected for Relativity

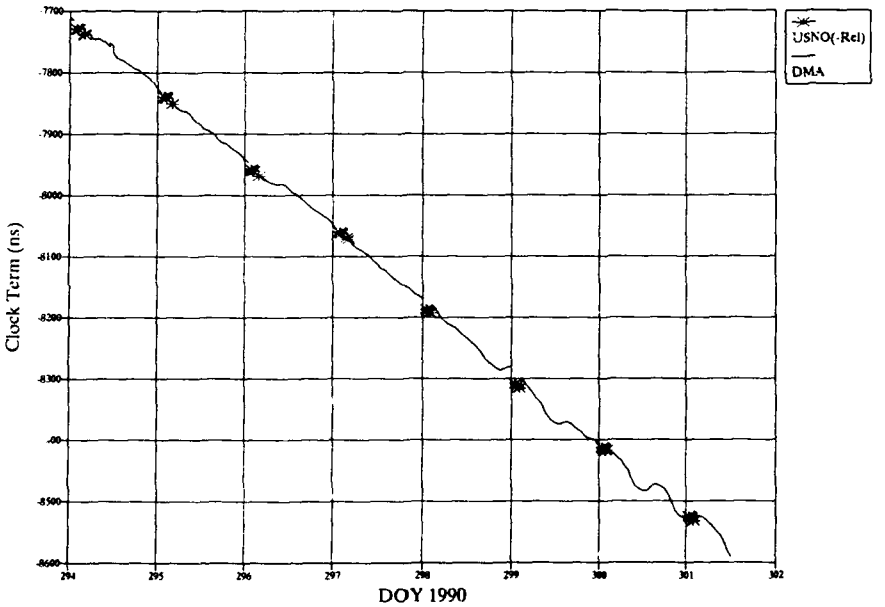


Figure 13, SV21 Clock Offset to GPS Time, DMA and USNO
Observations Uncorrected for Relativity

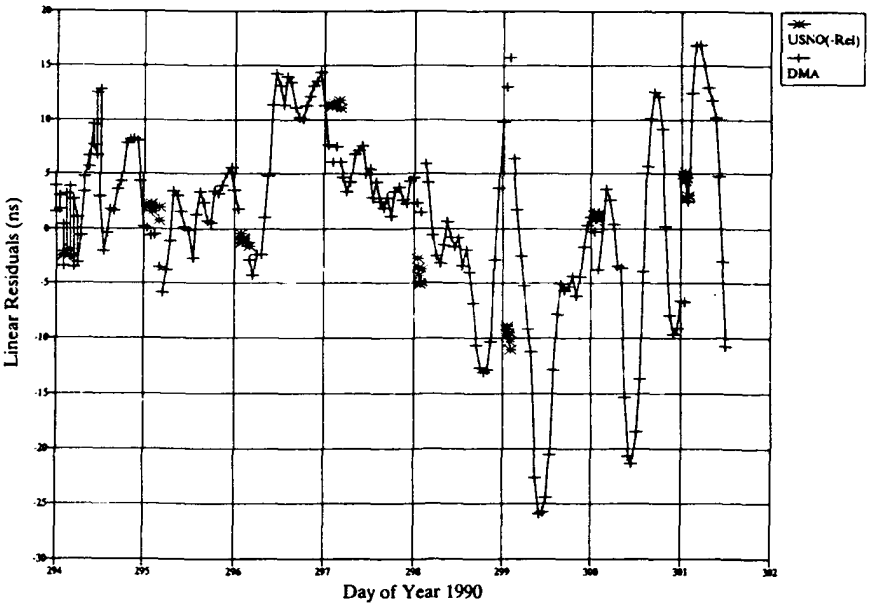


Figure 14, SV21 Clock Offset Residuals to GPS Time, DMA and USNO Observations Uncorrected for Relativity

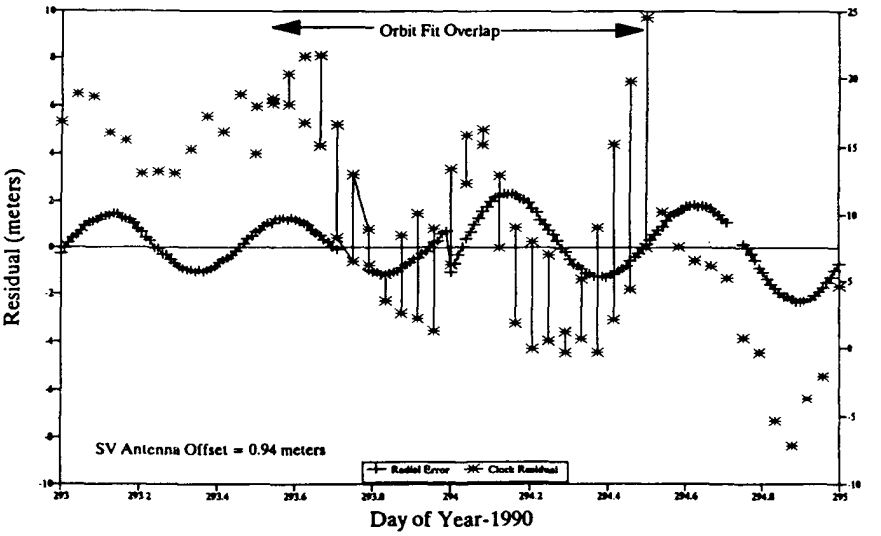


Figure 15, SV11 Radial Position Difference and Clock Offset to GPS Time Residuals in Nanoseconds

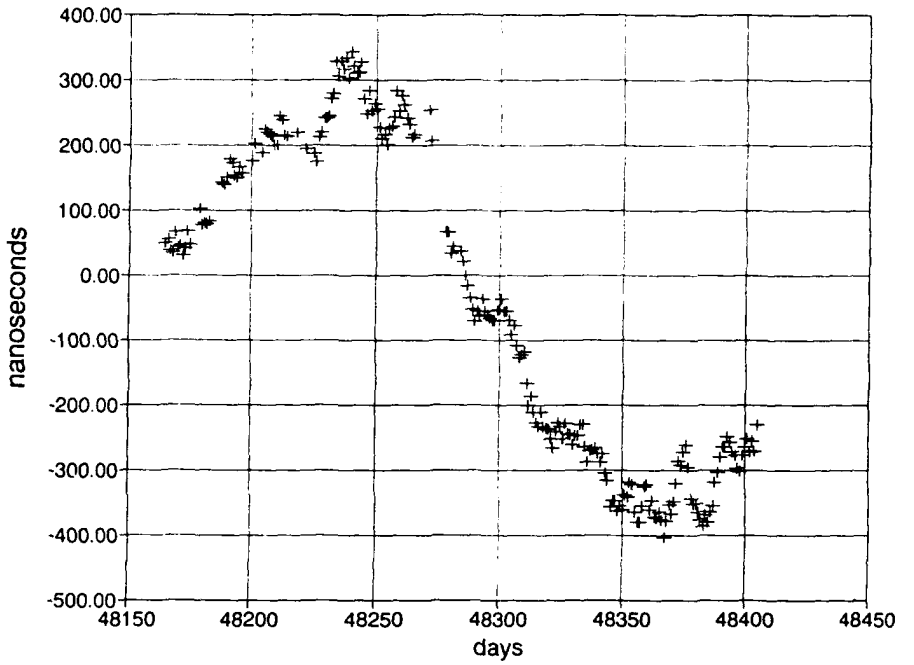


Figure 16, New Zealand Site vs UTC as Disseminated by GPS

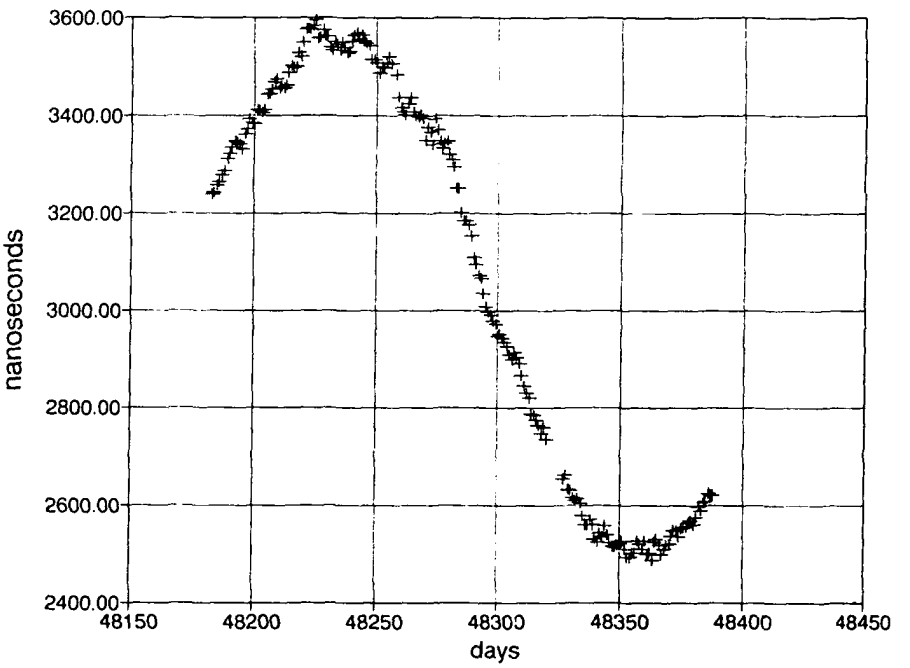


Figure 17, Orroral Valley vs UTC as Disseminated by GPS

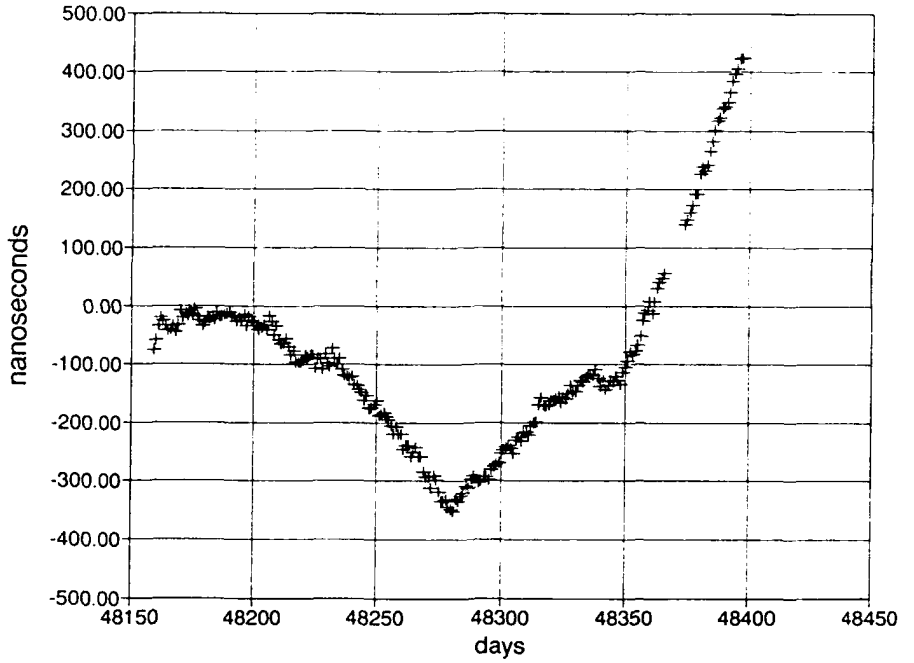


Figure 18, GPS Master Control Station Receiver vs UTC as Disseminated by GPS

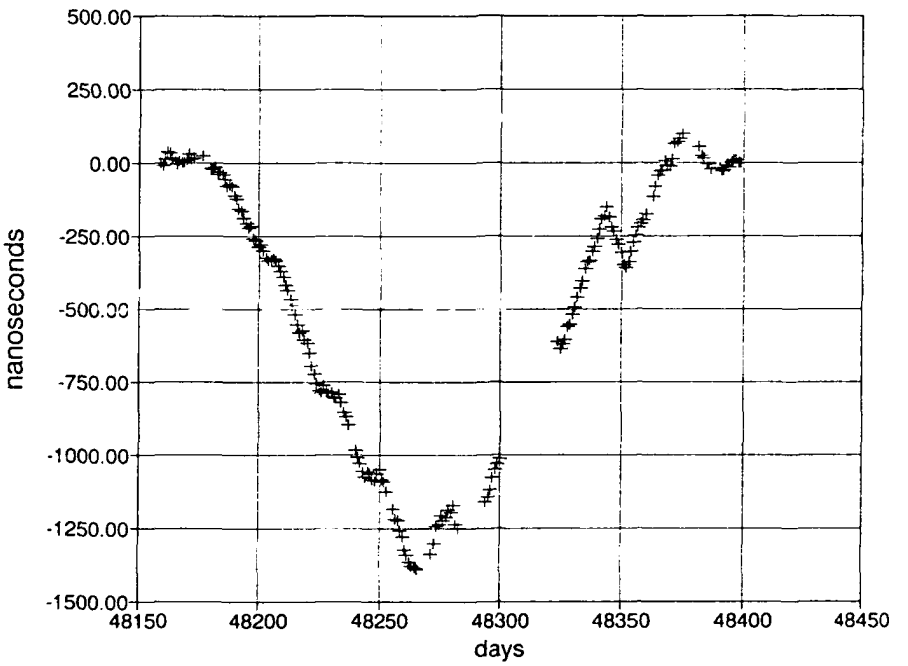


Figure 19, Hawaii Site vs UTC as Disseminated by GPS

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AUTOMATIC HIERARCHY CONTROL OVER STRATUM 1 REDUNDANT TIME & FREQUENCY SYSTEMS

E. PELED M. ZELITZKI A. NEMESH

TFL Time & Frequency Ltd. 14, Habanal St. Industrial Zone, Holon P.O.B. 1792 Holon 58117, Israel

Abstract

The paper describes a new system for Stratum 1 in telecommunication or for any high redundant frequency source system.

Traditionally, Primary Reference Systems are ensembles of three frequency sources in a 2:1 redundant configuration. One of the sources is used while the other two are hot standby substitutes. The three frequency sources can be any clocks ensemble, (such as Cesium, Rubidium or OCXO) or of the same kind. Whatever the structure is, the sources quality can change in time. (Due to aging, atomic tubes degradation etc.).
TFL, recently, developed an "aid" for such systems that can find out which of the three clocks is the best by various means, and change priorities in accordance with this scaling.

Introduction

Such system has been recently installed in BEZEK - the Israeli PTT at its International Digital Exchange as Stratum 1. System's Block diagram is shown in Fig. 1.

The system is based on an IBM PC, includes Time Interval Counter, and a sophisticated switch (controlled transfer unit) activated by the computer.
Three sources (RF) outputs are connected to the TF-3006B (Controlled Transfer Unit). This unit is a "smart" switch that can supply three outputs at any structure (channel one through channel three). The clock (source) connected to Ch1 will be the "Master" at the system's outputs. Ch2 will be second priority and Ch3- third priority.

Time Interval (TI) measurement between one source to the other sources, are constantly taken every 100 seconds by the Time Interval Counter. The measurements are stored in the computer hard disk. Using software clock AT1 time scale algorithm at the Nist methods, the hierarchy is established by the "waits" of the sources. Changeovers are done with no interference to the system's outputs (phase, amplitude or dead time). Thus a three frequency source system,that is built to conform to CCITT Rec. G.811, is improved by assuring the best (out of three) clock to be the one chosen at the system's outputs, while the other two are not standing by.

An additional input to the transfer unit, enables the three sources comparison towards another reference such as a GPS receiver (using special method of "ignoring its jitter") or an auxiliary Cesium Clock, as well as a "feed-back" from the exchange itself. Calibration toward UTC becomes easily available; as well as comparison to other remote similar systems (Via transmission lines- "Master Slave" operation), and ability to assure the stability of the exchange output.

The frequency source controller (FSC) software is part of TF-3002A - Test & Control Unit, which is a part of the TF-3100B System. The main purpose of this system is to provide the customers a reliable, precise frequency source, derived from the best source, out of three.

The phase aligning and the combining process has already been discussed in our previous paper from last year (44th annual symposium). In that system the clock that was connected to channel one input of the phase aligner and switch module, was 1st priority and Master to the system's outputs as long as it existed. It could be changed by a cross connection of the patching cables of TF-3006A (Transfer Unit), so that Ch1 input can be connected to clock (source) no. 2 or 3. The schematics block diagram of TF-3006B is shown in Fig 2, this is a computer controlled transfer unit.

As seen by Fig 2. the priority can be changed by a changeover of the coaxial switches. Any source can be connected to any channel so that in a case of "bad operation" of the first priority source it can be replaced by any other better source. The decision of which is done by the Test &

CH2965-2/91/0000-626 $1.00 © 1991 IEEE

Control Unit. (No phase or amplitude jumps occur, as a result of any change over!).

While the three 3÷1 coaxial switches are "in charge" of the cross connections and hierarchy establishing, the other two 4÷1 switches are continuously activated by the computer enabling the Time Interval Counter taking measurements, i.e. the time intervals (phases) between one frequency source to another (TIE).

These measurements enable the software to perform two tasks:

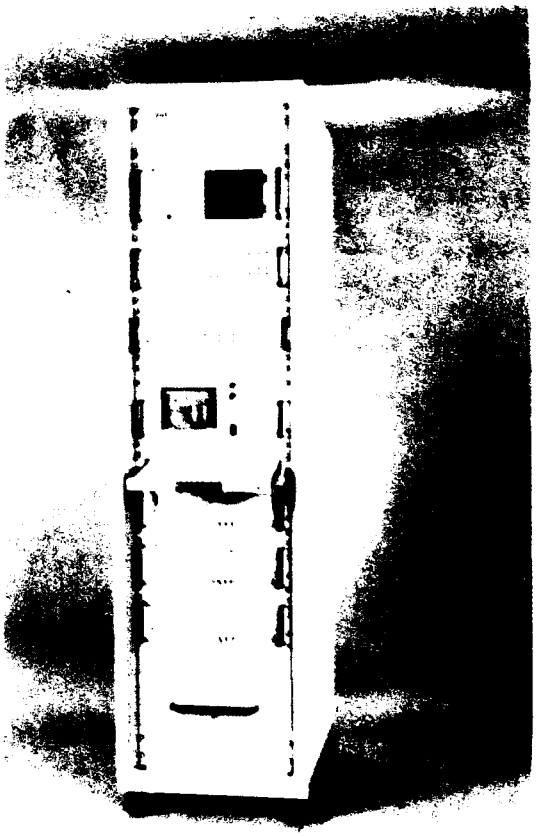
1. Creation of a software clock (based on NIST AT1 algorithm) and the accumulation of all sources history, in comparison to the software clock, and/or to UTC.
2. On line, two weeks back, determination of all sources by means of Frequency Offset, Weights, testing Short Term Stability and Faults.

The first task is the prediction of a Cesium tube failure as well as the calculation of the "very" Long Term Stability (over years).

The second task enables setting the hierarchy in accordance with clocks weights and the detection of a faulty source by mean of its offset and STS (short term stability). Such a clock will be disconnected to prevent changeover to an unaccurate or noisy source.

An option for which the software has already been written, is a modem PC card that enables any authorized executive to contact the control unit via a dedicated telephone line, and to get all graphs and status on his personal computer, just by dialing a telephone number.

The system is also capable of setting an alarm using that modem, on any predefined PC as long as it is connected to a telephone line and having a modem card, connected to its bus.



TF-3100B FREQUENCY & TIME GENERATING SYSTEM

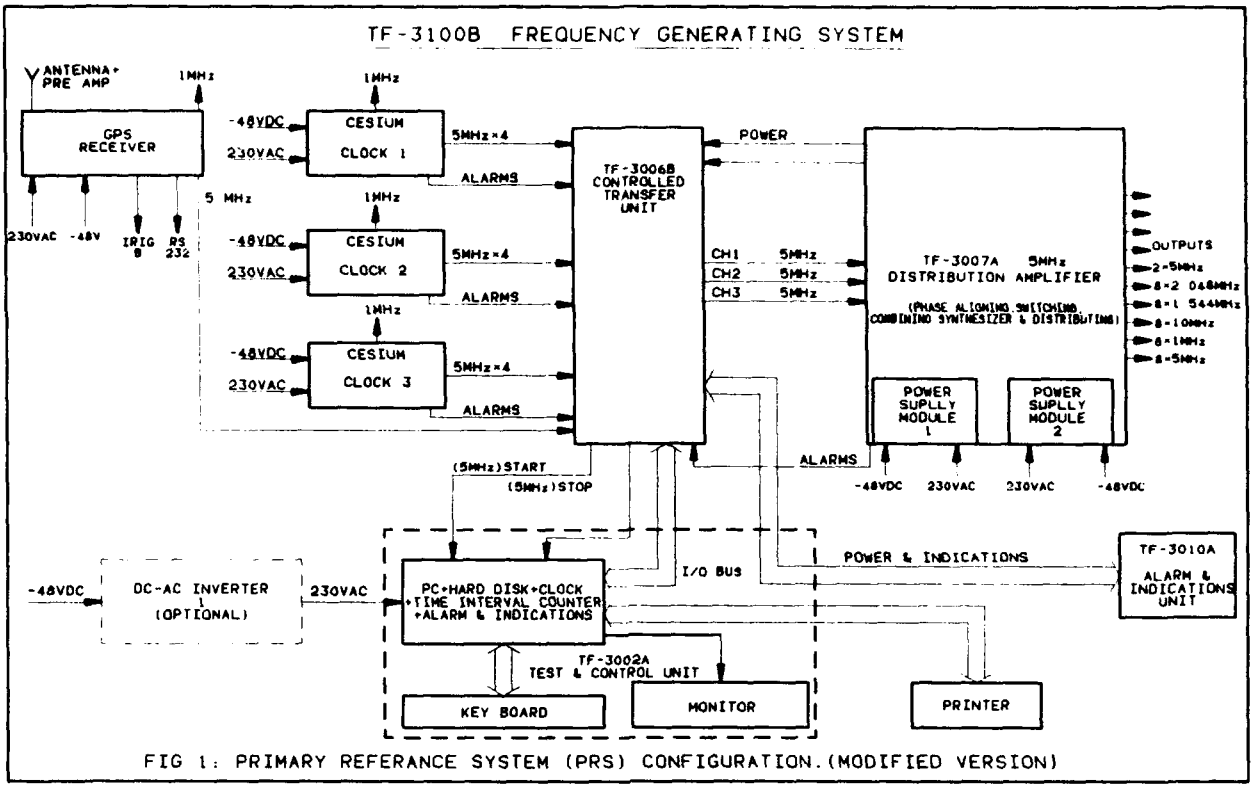


FIG 1: PRIMARY REFERANCE SYSTEM (PRS) CONFIGURATION.(MODIFIED VERSION)

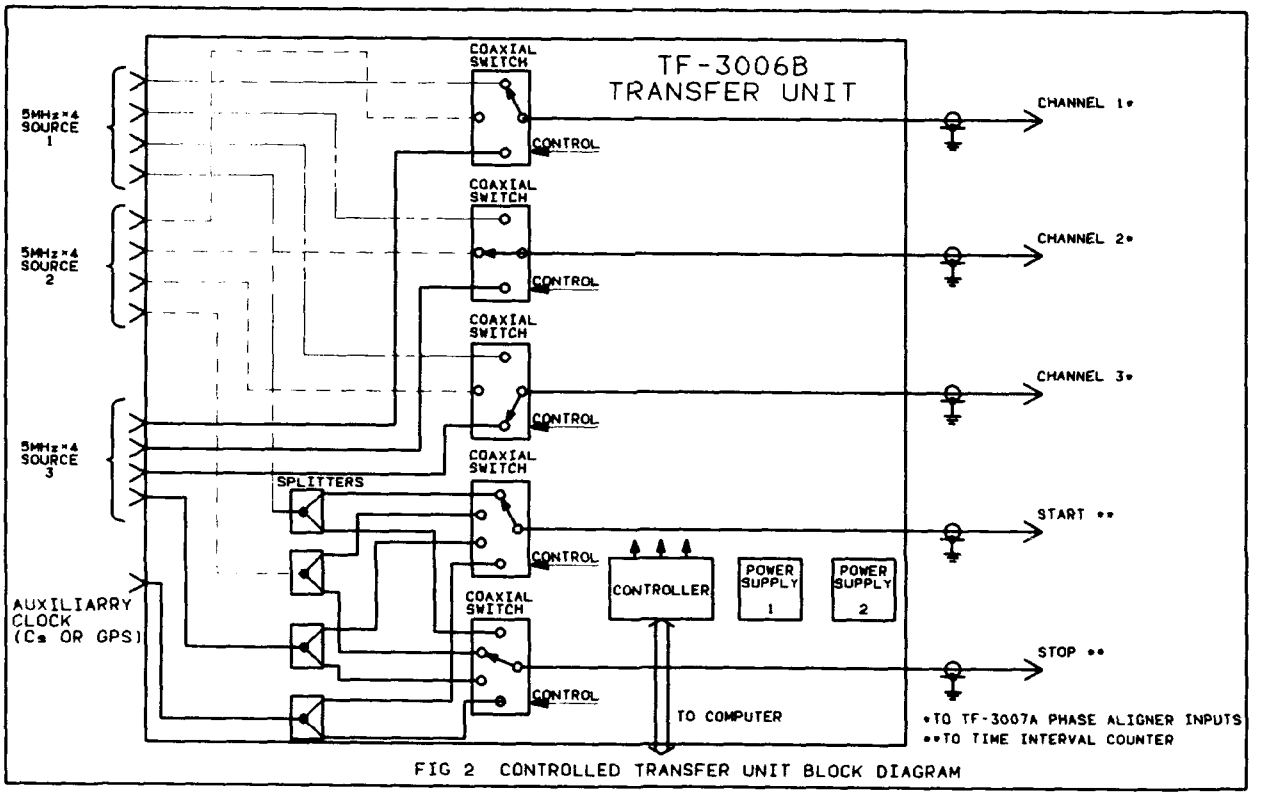


FIG 2 CONTROLLED TRANSFER UNIT BLOCK DIAGRAM

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Phase Noise Modeling in Frequency Dividers

William F. Egan

ESL A Subsidiary of TRW
495 Java Drive, Sunnyvale CA 94088-3510

Abstract

Noise within frequency dividers can be an important detrement to the spectral purity of frequency synthesizers. A set of noise sources that can be used to predict the phase noise of such dividers, as operating conditions vary, has been developed and values have been determined experimentally for one type of ECL divider. This set of sources is shown to give useful predictions of the performance of the tested type as well as of noise reported for other ECL dividers.

Introduction

Need for Design Information

Why it is required

In order to determine the spectral purity that can be obtained from a particular frequency synthesizer design, the designer must know two things. First the designer must know the noise profile of the various components of which the synthesizer is composed and, second, he or she must know how that noise influences the output. The latter is given by control theory but the former requires that the component noise be characterized. Such information has long been available for most components, including VCOs, op amps, and resistors but relatively little is available concerning frequency dividers.

Effect of divide ratio

Fortunately divider noise was not a significant contributor in many designs but its importance grows with higher divide ratios because, within the loop bandwidth, the output phase noise tracks the effective input noise of the divider, which equals the output noise multiplied by the divide ratio (Fig. 1). Thus a relatively large phase deviation would be required at the synthesizer output to cancel a relatively small deviation from a source at the divider's output.

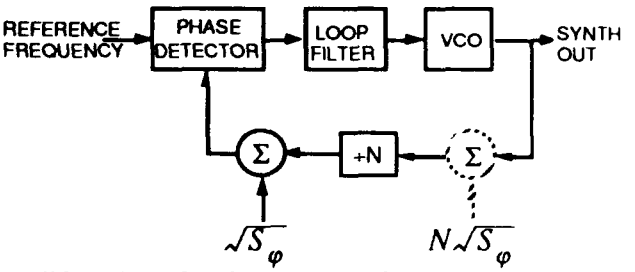


Fig. 1. Equivalent Noise at Divider Input

Historical Perspective

Available information

The first widely available noise data came from Scherer[1] in 1978. It consisted of curves of power spectral density versus modulation frequency for one TTL and two ECL devices. These were presented in seminars given by Hewlett Packard Co. and were reproduced in two books on frequency synthesis. While this information was most welcome, it left open the question of how the noise levels would change with operating conditions, such as input and output frequencies. What was needed was theory, of the kind that Leeson had provided for oscillator noise, which allowed reasonable estimates of performance to be made based on measurements made under different operating conditions. Information of this type was presented in a 1980 paper by Kroupa[2] and in a 1986 paper by Robins[3].

CH2965-2/91/0000-629 $1.00 © 1991 IEEE

Correlation of Information

Estimated phase power spectral density at the noise floor, as given in these papers, is shown in Fig. 2. While this theory was a valuable addition, one was led to wonder how universally applicable it was, how applicable were these few curves to the whole range of dividers of interest. In fact, when Scherer's data is plotted on the same graph, we see an estimate within 5 dB of the data in one case but also deviations of as much as 25 dB (the arrow indicates that the noise floor was not reached in the measurements and is somewhere below its tip). If we add data that has become available more recently, some of which is for other divider types (Fig. 8), we see even greater deviation so that Kroupa's curve is left to serve more as an upper limit to the measured data.

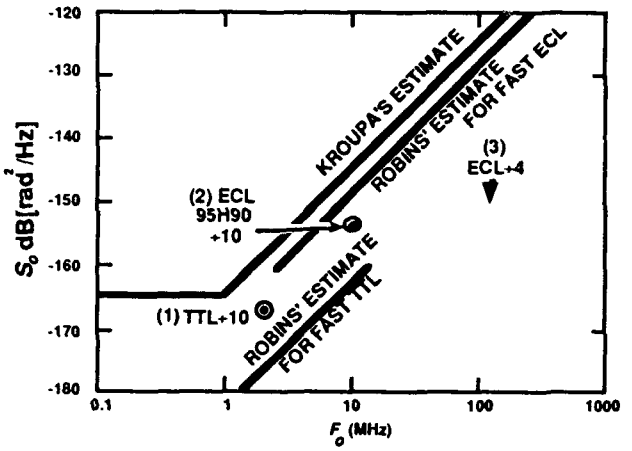


Fig. 2. Early Estimates of Noise Floor and Data

Research to be Discussed Here

Because of the inadequacy of the available information and because of the possibility of dependencies other than those that had been shown, some colleagues and I at GTE made measurements on various frequency dividers in the period from 1982 through 1984 and I attempted to determine the functional dependence of the measured noise on operating parameters. The main results were published earlier this year and the reader is referred to that referenced paper[4] for those details while this paper, while providing some new information, will attempt to place the topic in perspective so the reader can determine if further consideration is desirable.[5]

Modulation Frequency To Be Considered

While we need to know the noise level as a function of modulation frequency, this variation might be adequately modeled by a set of parameters in the 1/f region plus a set further out at the noise floor. In any case, for simplicity we will concentrate on the noise floor in this paper.

Formulation of the Problem

Sources to be formulated

To model divider noise in a manner that is most useful for design, a minimum number of noise sources are to be sought. The objective is not to aid in the design of dividers but in the design of synthesizers using the dividers so if many sources internal to the divider produce a net effect such that they are indistinguishable outside of the divider, then only one source is to be used to represent all the internal sources. Only when two noise sources depend differently on design parameters should they be specified separately so the designer can determine the resulting overall noise level as the various parameters (such as divide ratio, operating frequency, drive level, etc.) change. The plan then was to come up with a minimum set of sources, measure various dividers under various operating conditions, determine the values of the sources that would best correspond to the measured results, and finally to compare the performance indicated by these sources to measured performance, both for the measured dividers and for similar dividers from which the source values were not obtained.

Assumed Divider

The assumed divider chain is shown in Fig. 3. Each block is a divider that is driven asynchronously by the preceding block (i.e., the clock driving any block does not also go to another block, otherwise the two blocks would be taken as one synchronous divider).

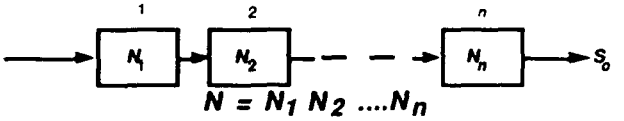


Fig. 3. Basic Divider

Formulation of Sources

Jitter

The first kind of noise that we might consider is time jitter within the divider circuitry (Fig. 4).

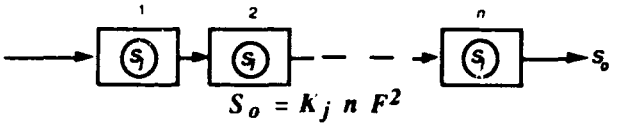
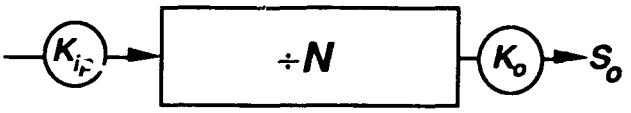


Fig. 4. Jitter (identical stages)

The variances of the jitter in each of the stages will add to produce a total jitter so that, if the stages are identical, the overall phase power spectral density will be proportional to the number of stages. Time jitter is converted to phase variation by division by the signal period, which is equivalent to multiplication by the signal frequency, at whatever point in the divider is being considered. (This is why phase deviation is divided by N in going through the divider.) Thus the effect of a fixed time jitter on phase deviation at the output is proportional to the output frequency and the power spectral density at the output is proportional to the square of the output frequency as well as to the number of stages. Therefore this type of noise can be represented at the output by a constant K_j, which is characteristic of the divider, multiplied by the number of identical stages and the square of the frequency. This is the type of noise that is represented by Robins' curves and by the sloped part of Kroupa's curve; S_O rises 6 dB per octave of F_O.

Output and Input Phase Noise

The second type of noise shown in Kroupa's curve is independent of F_O. This corresponds to a phase noise generator at the output of the divider (Fig. 5). A similar generator might exist also at the input of the divider (Fig. 5) and would be differentiated from the previously considered source by its dependence on divide ratio.



$S_o = K_{ip} / N^2$ $\qquad$ $S_o = K_o$

Fig. 5. Input and Output Phase Noise

Input Additive Noise

The fourth source is included to show dependence on the amplitude of the driving signal, since dividers in synthesizers are often driven by sine waves rather than by logic signals. This source represents an additive noise at the input rather than a pure phase noise and, therefore, its conversion to phase noise depends on the amplitude of the driving signal. The output produced by this source is similar to that produced by the input phase noise except that it is inversely proportional to the driving amplitude (Fig. 6).

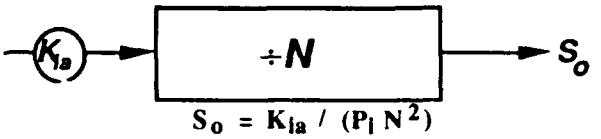


Fig. 6. Input Additive Noise

Independence of Sources

None of the four sources has the same dependence on operating parameters as any other. The effect of K_O on the output is independent of these parameters. The effect of K_{ip} is proportional to $1/N^2$. The effect of K_{ia} has an additional dependence on driving power and the effect of K_j is inversely proportional to F_O^2 as well as to the number of identical non-synchronous stages. Thus, not only is this set no larger than necessary but each member can be differentiated from each other by experiments in which these various parameters are varied.

Measurements

Noise was measured by a method similar to that used by Shearer (Ref. 1, Fig. 9). The two frequency dividers are driven by a common source whose noise will hopefully cancel when the phase of one divider is measured against that of the other. Their counts are offset to provide quadrature signals to drive the balanced mixer phase detector. The balanced mixer was chosen because its noise properties are well understood. However one disadvantage is that the quadrature relationship required between the two signals causes each to have a transition at a different time so that high-frequency noise modulation from the source is not perfectly canceled.

Additional Types of Sources

Phase Sampling

The data revealed dependencies not covered by the four postulated noise sources. These are apparently due to the effective sampling of the phase noise at the output frequency of the divider.[6,7] The phase is sampled because it can only be seen at an output transition. Sampling produces aliasing; noise at frequencies higher than half of the sampling frequency is converted by this process to frequencies below half of the sampling frequency and added to the noise already existing at those frequencies.

Example

For example, if the divider output frequency is 1000 Hz, phase noise that has a modulation frequency of 1,000,010 Hz will produce the same rate of change of the output phase as will noise at a 10 Hz rate. The same is true for all noise frequencies that offset by 10 Hz from multiples of 1000 Hz. Thus noise at many higher frequencies is aliased on top of the low frequency noise. If the output frequency in this example should drop from 1000 to 100 Hz, the number of frequencies within any given broad band of noise that would produce noise at a 10 Hz rate would be ten times greater. The number of bands that are aliased into the low-frequency band (below half of the output frequency) from a broad band of noise is thus proportional to the output frequency. Thus a new dependency arises, one that contains a factor F_O (in S_O) to account for this aliasing. To be general, the four constants obtained above must be retained to account for narrow-band noise, where the aliasing contributes little, but to each must be added another constant that differs only in the additional dependence of the output phase noise on the output frequency. Thus, for any given noise frequency, eight constants are required in general, although any of them may be found to be negligible for a particular divider.

General Results for ECL

For the ECL dividers that were measured, the most significant noise constants were those for additive output noise (corresponding to the horizontal part of Kroupa's curve) and sampled additive input noise, which varies inversely as the input power, the output frequency, and the square of the divide ratio. The next most significant contributor was sampled output noise, which is inversely proportional to output frequency. Finally input phase noise was given an estimated value. The other four constants, including the jitter constant that produces the 6 dB/octave slope as given by Kroupa's and Robins' curves, were estimated to be negligible.

Data Extraction

To extract the values of the constants from the measured data, a computer-generated listing of the data was used. Fig. 7 shows a fragment from such a list. The first line is

Sb	P	Fi	Fo	N	n	Ka	Kas	Ko
-149.5	900	1.07E+07	6.69E+05	16	1	774.6	829.8	-149.5
-143.1	10	1.00E+07	6.25E+05	16	1	-109	-54.1	-143.1
-146.7	1	8.00E+07	2.50E+06	32	1	-115.6	-54.6	-146.7

Fig. 7. Part of Data List

for a ÷16 configuration where the divider was driven by a square wave. This is assumed to act like a very large sinusoid, based on the slopes at the level where the divider is triggered, so a very large number was entered for the signal power tc eliminate effectively any dependence on signal power. Three of the eight constants are visible on the right. Each of the values listed there equals the value that would produce the measured output noise by itself (without contribution from other sources). A list of this type that has sufficient entries, made under sufficiently different conditions, enables the experimenter to go down each column under the noise constants and look for the smallest value. The true value can be no larger than this smallest constant because, in the line in which it appears, it is sufficient to contribute all the observed noise. After the maximum possible value for each constant is obtained in this manner, a somewhat different list is run. That list shows the theoretical contribution from the specified noise sources for each set of operating conditions. The total theoretical noise and the measured noise are shown together for comparison and major contributors are highlighted. This listing facilitates studying the effects of the various constants in order to improve the estimates.

Constants for F100136 ECL

Measurements were made on a synchronous divider, composed of F100136 4-stage integrated-circuit dividers. Forty-nine measurements were taken with division ratios from 4 to 32 and input powers from -8 dBm to +10 dBm, plus square wave. Input frequencies varied from 1.4 to 80 MHz. The predicted noise levels, based on these data, are given by the equation

$$S_O = K_{ip}/N^2 + K_O + K_{ias}/(F_O N^2 P_i) + K_{OS}/F_O ,$$

where $K_{ias} = 10^{-5.16}$ rad^2-mW
and $K_O = 10^{-15.31}$ rad^2/Hz

were the most significant contributers followed by

$K_{OS} = 10^{-10.13}$ rad^2

and then by $K_{ip} = 10^{-14.47}$ rad^2/Hz.

Except for four points at the lowest input levels, the predictions match the data to within +1.3 and -1.0 dB while varying over 25 dB. Eleven more measurements with division by 64 and input powers between -6 and +10 dBm showed a spread of ±1.7 dB but the measured mean was about 2.5 dB low at this divide ratio, which we were unable to explain.

It is hard to justify the existance of unsampled white noise, except at the divider's output, since noise originating elsewhere appears only through modulation of the switching instant. For this reason, and because it is not apparent why the divider should have input phase noise separate from additive noise and jitter, there is a suspicion that K_{ip} might represent band-limited noise from the test generator.

Use of These with Other ECL Types

Estimates of the phase noise levels corresponding to data points for other ECL dividers, based on these constants obtained by measuring the F100136s, are shown in Fig. 9. The measured points are assumed to have been obtained with square wave or high level inputs so that the additive input noise has no appreciable effect, leaving three constants to be used in the estimates. However, none of the frequencies is low enough for the sampled output noise to be important. For the lowest-frequency data (2, 12, 13, 14), only the unsampled output noise source is significant so the estimated level there is independent of output frequency (as is much of the measured data). These points have divider ratios of 10 or 20, which suppress the effect of input phase noise, but that source causes the estimate for the ÷4 (point 3,4) to rise slightly and it becomes the dominant contributor to the estimate for the ÷2 (point 17).

TTL Data

Data was also taken on TTL devices. The resulting constants for four types are given in the appendix even though they are very tentative for four reasons: (1) data are from earlier measurements made during the evolution of the measurement techniques; (2) it is more difficult to get good data on TTL devices because they are so quiet (perhaps that is why designers have gotten along as well as they have without divider noise data); (3) discrepancies between measured values and values predicted from the constants are relatively large, as indicated in the last column of the table; and (4) the estimated values based on types 7474, 74163, and 74LS163 predict levels that are higher by 15 to 21 dB than the only TTL data in Fig. D.

Hope for Additional Data

It is hoped that others will perform similar measurements in order both to verify our measurements and give some indication of the spread to be expected and to determine constants for other divider types. Ideally such information would be included in divider specification sheets, but until such a time, designers will have to depend on their own measurements or estimates based on what may become available in the literature.

Appendix : Tentative TTL Data

Although the confidence level in the following values is low, considering the scarcity of TTL data they are being provided as reference points for future research and possibly for cautious use in design where nothing better is available.

ESTIMATED VALUES OF CONSTANTS FOR TTL

	Kip	Kips	Kia	Kias	Ko	Kos	Kj	Kjs	INPUT	WORST ESTIMATES High by	Low by
74LS163	-131.5	-65.5			-145.5	-87.5	-277.6		Logic	4.3	5.0
74ALS74		-73.4			-150.5		-278.6	-206.2	note A	4.0	4.5
74S74	-131.5	-69.4			-150.5	-90.5	-283.6	-213.5	note A	5.1	22.3
7474		-68.4			-153.5	-96.5	-276.6		note A	1.4	3.2
74163		-85.5		-69.7	-157.3			-218.2	note B	5.2	8.3

LOWEST MEASURED VALUES FOR TTL (≈ upper bounds)

	Kip	Kips	Kia	Kias	Ko	Kos	Kj	Kjs	
74LS163	-131.5	-65.5			-145.5	-87.5	-277.6	-210.5	Logic
74ALS74	-132.5	-73.4	-122.5	-63.4	-150.5	-96.5	-278.6	-209.5	note A
74S74	-131.5	-69.4	-121.5	-59.4	-150.5	-90.5	-283.6	-213.5	note A
7474	-131.5	-68.4	-121.5	-58.4	-150.5	-93.5	-276.6	-210.5	note A
74163	-140.2	-85.5	-133.5	-69.7	-157.3	-103.3	-284.2	-218.2	note B

Note A: Estimated input levels. Note B: Various input levels.

Table 1. Tentative Values of Noise Constants for Some TTL Dividers

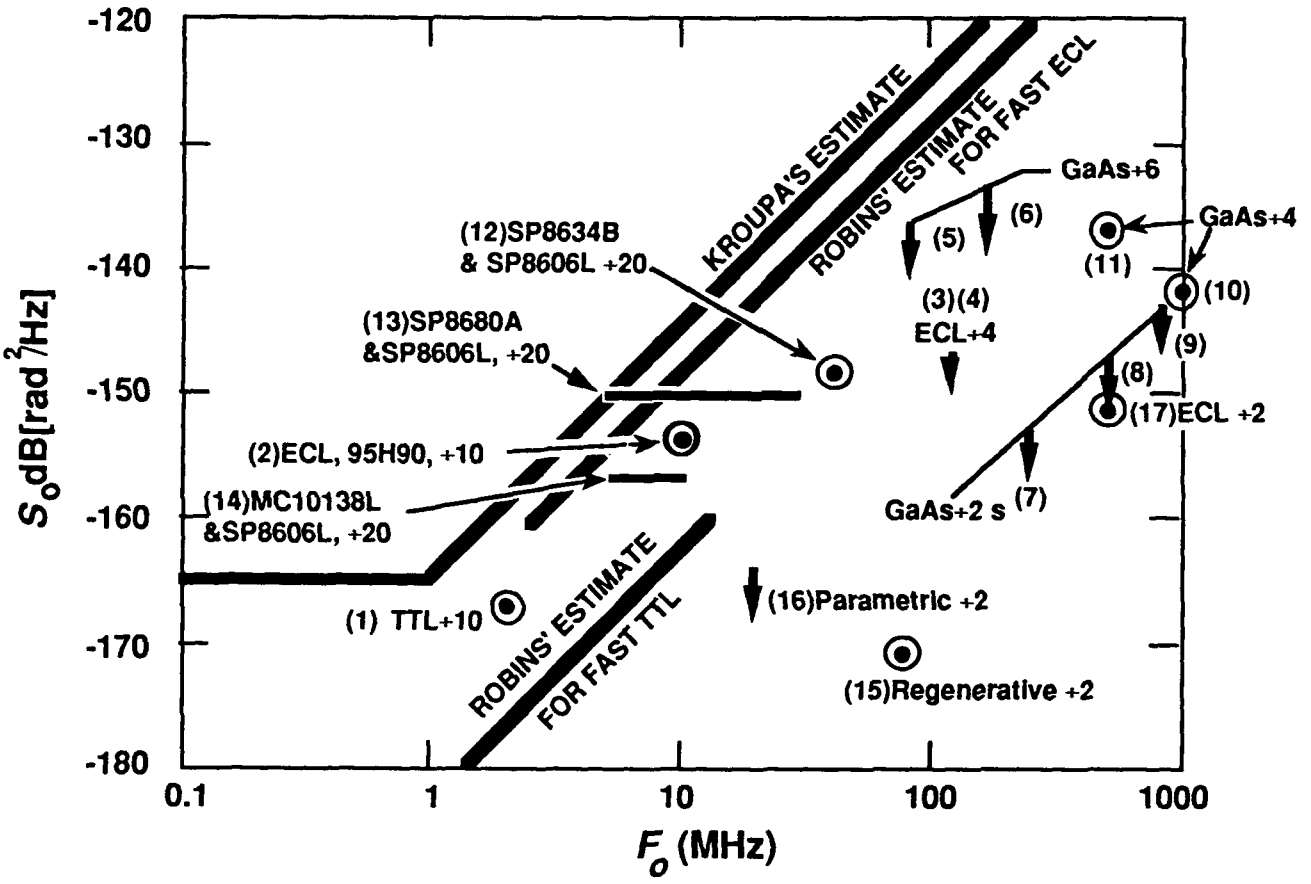


Fig. 8. Reported Data Superimposed on Earlier Estimates
Curves 1-14, see Table 1 in Ref. 4. Curves 15 and 16, ref._[8]. Curve 17, ref._[9]

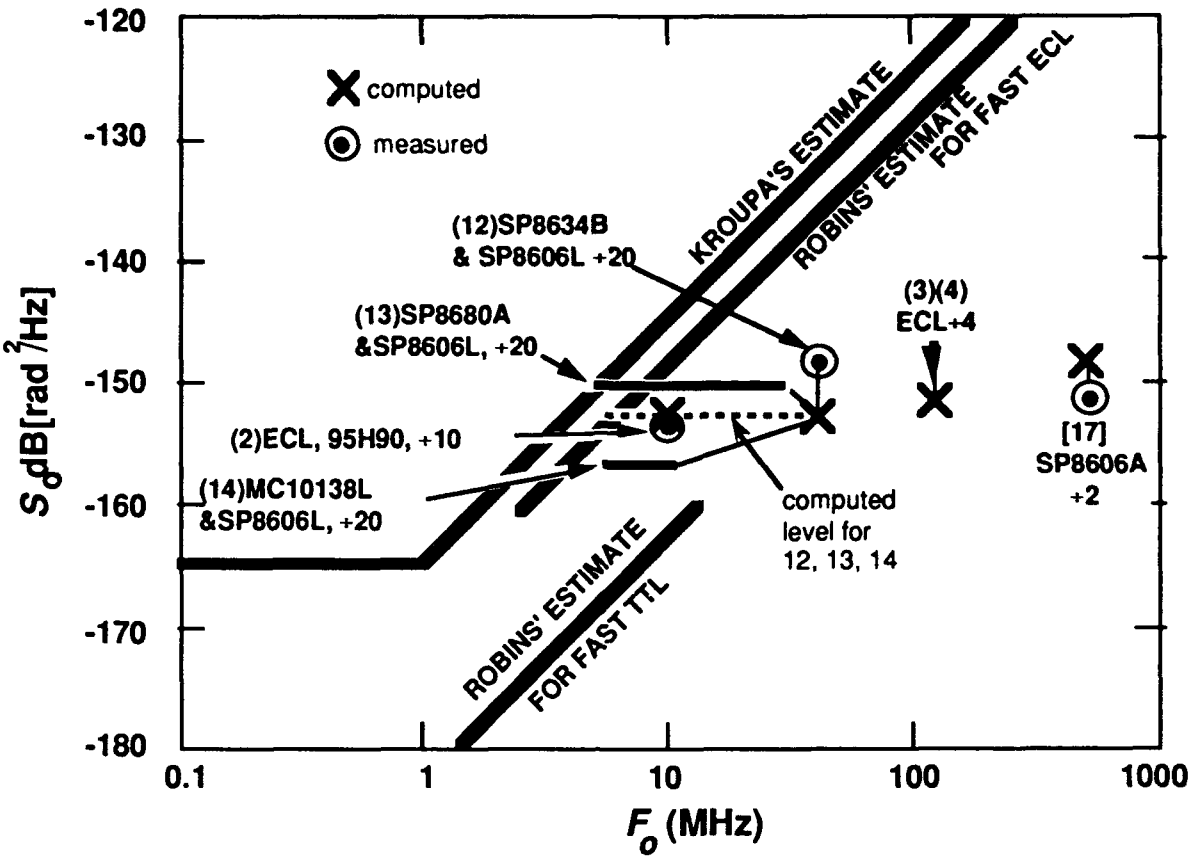


Fig. 9. Comparison of Predicted and Measured Values

---

[1]D. Scherer, "Design Principles and Test Methods for Low Phase Noise RF and Microwave Sources," in *RF & Microwave Measurement Symposium and Exhibition*, (Hewlett Packard, Inc., Palo Alto, CA: probably in 1978), p. 20

[2] V. Kroupa, "Noise in Standard Sources," in Digest of the Conference on Precision Electromagnetic Measurements, 1980, pp. 147-151

[3]W. Robins, "Synthesizer Phase Jitter Contributed by TTL & ECL Components " in Digest of the Colloquium on Low Noise Oscillators and Synthesizers, 1986, pp. 7/1-7/7.

[4]W. Egan, "Modeling Phase Noise in Frequency Dividers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, July 1990, p. 307-315. CORRECTION NOTE: in Table 1, right column, "a" should be "b".

[5]See also D.E. Phillips, "Random Noise in Digital Gates and Dividers," *Proceedings of the 41st Annual Frequency Control Symposium*. IEEE, 1987, pp. 507-511.

[6]W. Egan, Frequency Synthesis by Phase Lock. New York: Wiley, 1981 and Malabar, FL: Krieger, 1990, pp. 75-81.

[7]W. Egan, "The Effects of Small Contaminating Signals in Nonlinear Elements Used in Frequency Synthesis and Conversion," *Proc. IEEE*, Vol. 69, pp. 802-805 and 809-811, July 1981.

[8]M.M. Driscoll, "Phase Noise Performance of Analog Frequency Dividers," *43rd Annual Symposium on Frequency Control, 1989*, p. 342-348

[9]G. K. Montress, T. E. Parker and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components," *43rd Annual Symposium on Frequency Control, 1989*, p. 349-359

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH SPECTRAL PURITY FREQUENCY SOURCES USING LOW NOISE REGENERATIVE FREQUENCY DIVIDERS

J. Groslambert*, M. Olivier* and E. Rubiola**

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32,avenue de l'Observatoire - 25000 Besançon - France
** Dipartimento di Electtronica - Politechnico di Torino
c/so Duca degli Abruzzi 24 - 10129 Torino - Italy

Abstract

There is a need for reference frequency sources with high spectral purity in the frequency range of 5 to 10 MHz. The regular oscillators operating at these frequencies exhibit a spectral purity with a phase noise floor of the order of -155 dB (0 dB = 1 rad²/Hz) for Fourier frequencies higher than 1 kHz. On the other hand, oscillators at higher frequencies (VHF, UHF or Microwaves) have a fractional frequency fluctuation power spectrum $S_y(f)$ which is better at 1 kHz from the carrier and farer. Regenerative frequency dividers with input frequency in the range of 100 MHz to 1 GHz were studied. The phase noise of such dividers shows a 1/f spectrum corresponding to -155 dB. These performances are much better than those of GaAs digital frequency dividers. Using these regenerative frequency dividers at the VHF output of a synthesizer, a frequency source of high spectral purity was achieved, working in the range of 5 to 20 MHz. It can be used to test the spectral purity of oscillators in this frequency range.

1. Introduction

In frequency metrology low noise dividers are used especially in synthesized signal generators. Analog dividers show some advantages compared to digital ones, due to the higher input frequency and lower 1/f noise and noise floor performance. In the last years monolithic implementations of the regenerative divider have been proven to be feasible, by means of various technologies [1, 2], thus making this scheme more attractive.

Although the regenerative scheme appeared in the literature a long time ago and was explained in recent papers [3, 4] only a few experimental results dealing with phase noise have been published. An overview of the frequency dividers operation is given in section 2. We have realized and measured the phase noise of a regenerative frequency halvers chain. The experimental results are presented in section 3.

This regenerative divider chain can be used to obtain, in the R.F. domain, a synthesized signal with very high spectral purity. The noise floor of the synthesized signal is of the order of -160 dB. An application of these dividers is presented in the last part of this paper.

2. Frequency dividers operation

The block diagram of a regenerative frequency divider by N + 1 is shown in Fig. 1.

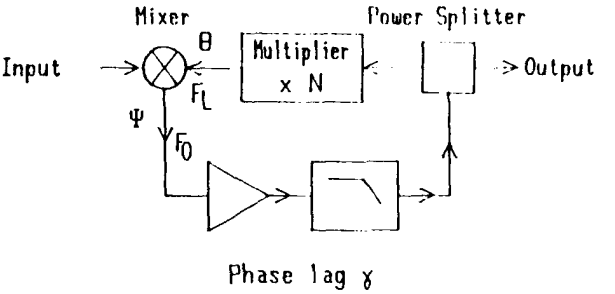


Fig. 1 :
Block diagram of a regenerative frequency divider

The mixer is a nonlinear 3-port device. The combined action of mixer and filter is such that the IF frequency f_0 is

$$f_o = f_R - f_L \qquad (1)$$

At the output of the multiplier, $f_L = Nf_0$, and a synchronous mode of operation is established, with $f_0 = f_R / N + 1$.

Regeneration normaly starts from the thermal noise, and the loop gain at low level must be greater than unity ; this condition is not always realized when a multiplier is inserted in the loop. Additionaly, the loop gain must be less than unity for zero output when the input signal is removed.

We were interested by a chain of frequency dividers by 2, which give us a signal in the frequency range 5 - 20 MHz from the output of a synthesizer at 160 MHz. The frequency dividers by 2 have no multiplier in the loop : a low-pass filter can be introduced at the output of the mixer.

The transfer function of this n^{th} order low pass filter is of the form $H(f) = 1 / 1 + j (f / f_c)^n$. the mixer introduces, at first order, a translation of the low-pass frequencies around the input frequency f_0.

CH2965-2/91/0000-636 $1.00 © 1991 IEEE

We obtained a total frequency response $H_T(f)$

$$H_T(f) = \frac{1}{1 + j(f/f_c)^n} \cdot \frac{1}{1 + j\left(\frac{f_o - f_c}{f}\right)^n} \qquad (2)$$

This is a band-pass response. The phase lag introduced by this filter, including the effect of the amplifier, can easily met the requirement of stability over a large bandwidth.

The noise at the output of the frequency divider is the phase noise of the input signal, divided by the division ratio, and increased by the contribution of the noise generated in the loop, the most significant source of noise being the amplifier. The close loop phase condition of the divider is

$$\Psi + \gamma = 0 \pm 2K\pi \qquad (3)$$

where γ is the phase lag introduced into the loop, θ and Ψ are the phase at the L.O. input and at the I.F. output of the mixer, the phase of the RF input being used as the phase reference. A phase gain $G = \partial\Psi/\partial\theta$ can be defined.

At the output of an ideal mixer, $\Psi = -\theta$, and the phase gain $G = -1$. The output phase noise spectrum $S_\theta(f)$ is related to the internal phase noise spectrum $S_\gamma(f)$ by the relation

$$S_\theta(f) = \left|\frac{1}{1-G}\right|^2 S_\gamma(f) = \frac{1}{4} S_\gamma(f) \qquad (4)$$

$S_\gamma(f)$ is the spectral density of the phase noise contribution of the amplifier, being the other noise sources within the loop generally negligible.

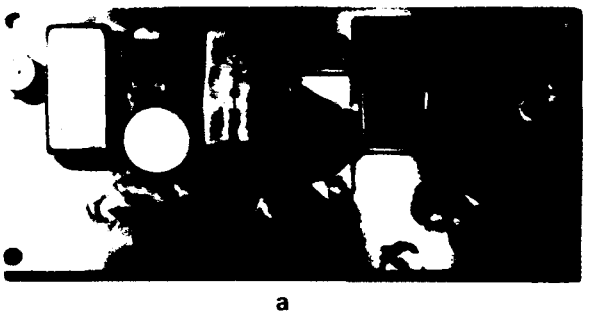
In fact, in a real mixer, many harmonics of the two input signals give a contribution to the output, each one with its own phase. This is the base of a phase-gain phenomenon. The loop gain can attenuate the phase perturbations arising inside the loop [5]. We can choose a working point with a strong harmonic distorsion in order to achieve $G < -1$, and this improves the noise performances of the divider.

We prefered to keep the harmonic levels small, this approach being recommended for wideband applications. The noise level of the divider is still very low, and appropriate for our application.

3. Frequency divider noise performances

We have built two identical chains of frequency dividers by 2 operated at 1 GHz input frequency (Fig. 2).

The first divider has a power splitter with a limited bandwidth of 600 MHz. It acts as a low pass filter. The low pass filter of the second divider is built with two toroidal cores and a capacitor.



a



b

Fig. 2 : A frequency divider chain
(a) Topview
(b) Bottom view

The phase noise of each divider is measured, using the classical differential scheme with two identical frequency halvers. The phase quadrature condition at the input of the mixer has been realized, using a different equilibrium state for each frequency halver.

The phase noise spectrum of the output of a frequency divider chain appears to be independent of the number of frequency dividers in the chain. Figure 3 represents the output noise of a divider by eight : 1 GHz - 125 MHz. We observe a noise floor of -165 dB, and a flicker noise of -165 dB for a Fourier frequency of 1 kHz. The harmonics of the output signal are rejected at least by 30 dB without using an external filter (Fig. 4). The operating frequency range of this divider chain is of the order of 300 MHz.

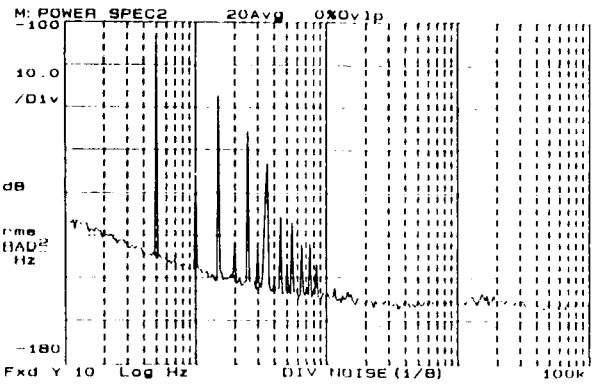


Fig. 3 : Phase noise spectrum of a divider by 8

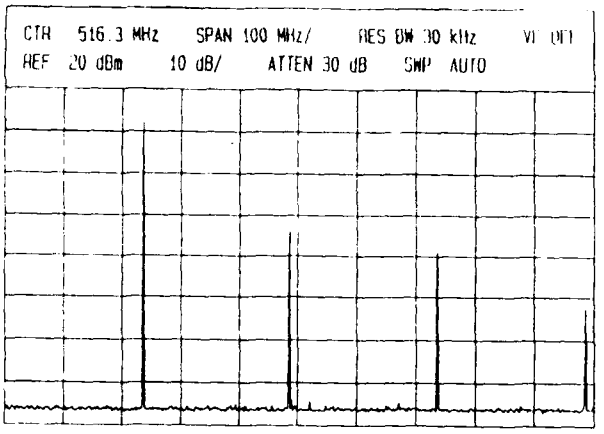


Fig. 4 : Output signal spectrum of a divider chain

This low noise regenerative divider, which can be built with a monolithic implementation is very attractive in comparison with an emitter-coupled logic divider. Apart from the better noise performances, the output signal is quasi sinusoidal, which is an advantage over the ECL divider.

4. A regenerative frequency divider application

Modern synthesizers generally use an heterodyne synthesis for output frequencies lower than 100 MHz and direct synthesis for frequencies upper than 100 MHz. The noise floor versus the output frequency of this type of synthesizers is represented on Fig. 5.

This synthesizer exhibits its best performances for output frequencies in the range 100 - 300 MHz. A synthesized signal with very high spectral purity could thus be obtain in the R.F. frequency domain, 10 MHz to 30 MHz for example, by using the regenerative divider whose effect is represented on Fig. 5 by arrows. A divider ratio of 16, that is 2^4, rejects the noise floor of a 160 MHz output signal of the synthesizer by 24 dB, and the limitation comes from the internal noise floor of the divider, which is -165 dB as measured previously.

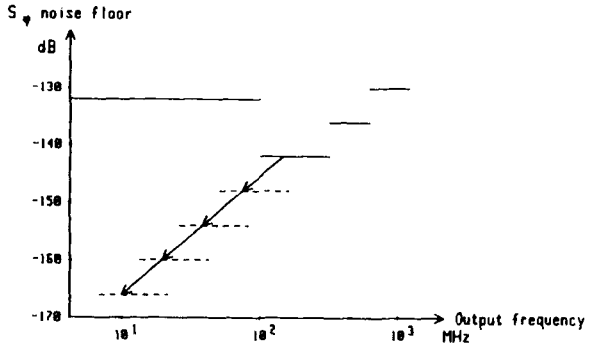


Fig. 5 : Noise floor of a frequency synthesizer

We built a regenerative divider chain with a divider constant of 16. The input frequency is centered at 160 MHz and the operation bandwidth is 75 MHz. The output signal at 10 MHz of this device is used as a spectral reference signal to measure the phase noise spectrum of a 10 MHz quartz crystal oscillator. The result is shown on Fig. 6, and this spectrum must be compared to the phase noise spectrum of the quartz oscillator, obtained by the classical method using two identical oscillators (Fig. 7). The noise floor at 160 dB of the quartz oscillator is correctly measured with the synthesizer and regenerative divider chain. Conversely, the same measurement has been made using ECL divider chains, and the best noise floor we can obtain is -152 dB. The phase noise of the ECL divider appears to be, in any case, preponderant.

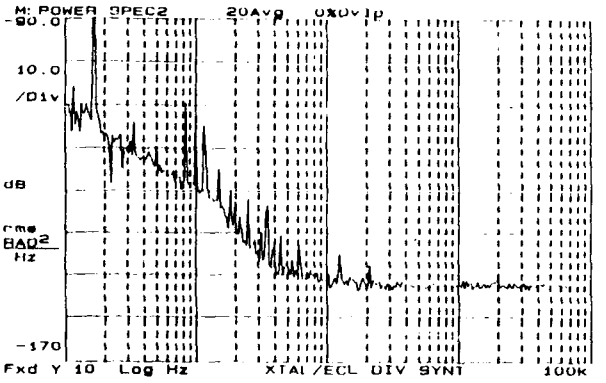


Fig. 6 :
Phase noise spectrum of a 10 MHz quartz oscillator
Reference : output signal of the divider chain

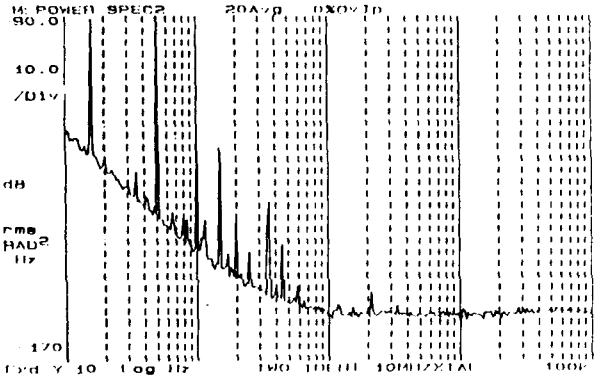


Fig. 7 :
Phase noise spectrum of a 10 MHz quartz oscillator
Measurement method using 2 identical oscillators

Conclusion

The measurement of the phase noise performance of regenerative frequency dividers indicates that this type of analog dividers is capable of providing much lower phase noise than conventional ECL dividers. The regenerative divider can be used to obtain a quasi sinusoidal signal of high spectral purity in the R.F. domain, when the VHF input frequency of the divider is delivered by a modern frequency synthesizer.

This synthesized signal can be used to test the spectral purity of an oscillator whose frequency lies anywhere in the range 5 to 20 MHz, the noise floor of the measurement being of the order of -165 dB.

References

[1] R.H. Derksen, H.M. Rein, K. Wörmer, "Monolithic integration of a 5,3 GHz regenerative frequency divider using a standard bipolar technology", Electronics Letters 21 (22), 1037-1039 (1985).

[2] T. Ohira, K. Arahi, T. Tanaka, H. Karo, "14 GHz band GaAs monolithic analog frequency divider", Electronics Letters 21 (22), 1057-1058 (1985).

[3] G. Immovilli, G. Mantovani, "Analysis of the Miller frequency divider by two in view of applications to wideband FM signals", Alta Frequenza 19 (11), 313-323 (1973).

[4] R.C. Harrison, "Theory of regenerative frequency dividers using double balanced mixers", IEEE Trans. on Microwave Theory and Techniques, MTT-S Symposium Digest 1, 459-462 (1989).

[5] E. Rubiola, M. Olivier, J. Groslambert, "Phase noise of regenerative frequency dividers", 5th European Frequency and Time Forum, Besançon, 1991.

EFFECTS OF CAVITY Q AND COUPLING ON OSCILLATOR NOISE MEASUREMENTS

Charles Cook
Johns Hopkins University
Baltimore, MD

ABSTRACT: This paper derives the response of single oscillator phase noise test setups. Both AM and PM noise responses are described as a function of the noise frequency and setup scattering parameters. The response of a three resonant cavity setups are compared to the response of the delay line setup. The effect of cavity Q and coupling on the setup response is shown for three resonant cavity configurations.

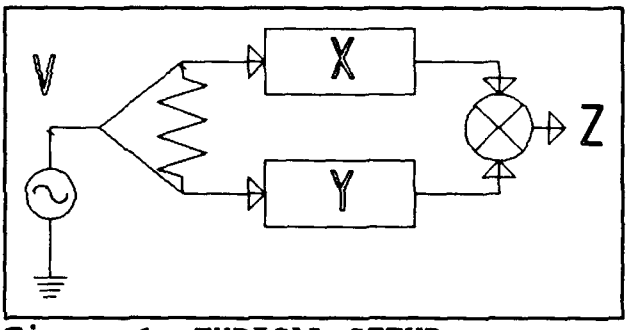


Figure 1: TYPICAL SETUP

INTRODUCTION

Oscillators have been described by a single carrier contaminated with amplitude noise (AM) and phase noise (PM) [1]. AM and PM noise will appear as sidebands when the oscillator output is viewed on a spectrum analyzer. Accurate measurement of these noise spectrums has typically required hardware which translates the noise spectrum from around the carrier to zero frequency. For AM noise this is accomplished using a simple square law detector. PM noise measurement requires more complicated setups using a frequency discriminator and mixer.

A typical frequency discriminator phase noise setup is shown in figure 1. The oscillator output is split in half using a power divider. The two halves pass through separate networks and are then multiplied together using a mixer. The mixer output Z contains a filtered version of the phase noise power spectrum. A well known example of this technique replaces network X of figure 1 with a short and network Y with a coaxial delay line. Another method replaces network Y with a resonant cavity [1][2] as shown in figure 2.

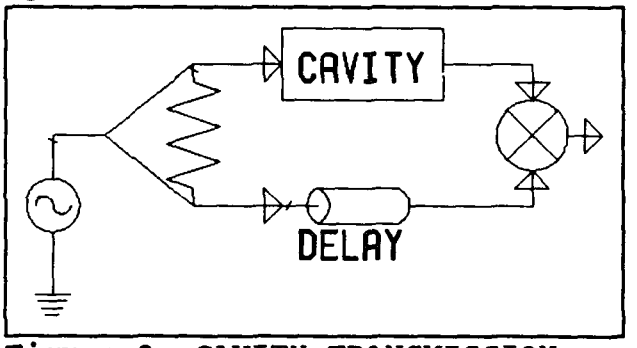


Figure 2: CAVITY TRANSMISSION

This paper modifies the network shown in figure 1 to the network shown in figure 3. The network of figure 3 models the oscillator as an ideal source with inputs for AM and narrow band PM modulation.

Networks X and Y are assumed to be linear functions of frequency. The noise output Z as the sum of a filtered AM and PM spectrums. This technique was used by Lesage and Audion [3] to describe phase noise

CH2965-2/91/0000-640 $1.00 © 1991 IEEE

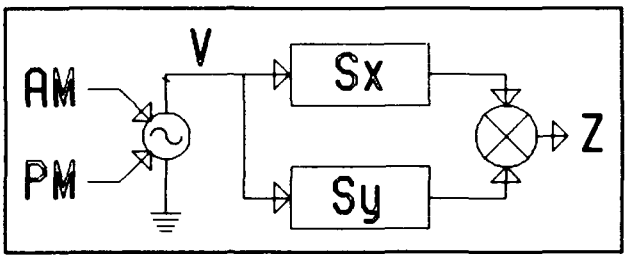


Figure 3: SETUP MODEL

measurement using filters. This approach assumes low levels of AM and PM modulation to develop a linear system response. The output spectrum at Z can be described using the following equations.

$$S_\epsilon(\omega) = \textit{OSCILLATOR AM NOISE PSD}$$
$$S_\phi(\omega) = \textit{OSCILLATOR PM NOISE PSD}$$
$$S_{z\epsilon}(\omega) = \|H_\epsilon(\omega)\|^2 S_\epsilon(\omega) \quad \textit{OUTPUT AM}$$
$$S_{z\phi}(\omega) = \|H_\phi(\omega)\|^2 S_\phi(\omega) \quad \textit{OUTPUT PM}$$
$$S_z(\omega) = S_{z\epsilon}(\omega) + S_{z\phi}(\omega)$$

(1)

OSCILLATOR MODEL

The oscillator is modeled as a cosine wave with AM and PM modulation.

$$\begin{aligned} v(t) &= (1+\epsilon)\cos(\omega_o t+\phi) \\ &= (1+\epsilon)\,[\cos(\omega_o t)\cos(\phi) \\ &\quad -\sin(\omega_o t)\sin(\phi)\,] \\ &\approx (1+\epsilon)\cos(\omega_o t) - \phi\sin(\omega_o t) \end{aligned}$$

If the modulation levels are small in amplitude and at low frequencies the oscillator can be approximated as follows.

$$\epsilon = \epsilon_o \cos(\eta t+\alpha) \quad \textit{AM MODULATION}$$
$$\phi = \phi_o \cos(\eta t+\beta) \quad \textit{PM MODULATION}$$

ω_o = *OSCILLATOR FREQUENCY*
η = *MODULATION FREQUENCY*
ϵ_o, α = *AM AMPLITUDE , PHASE*
ϕ_o, β = *PM AMPLITUDE , PHASE*

This model describes PM noise using the accepted narrow band PM modulation approximation. Converting v(t) into the frequency domain yields:

$$\begin{aligned} V(\omega) = &\frac{1}{2}\delta(\omega_o) + \frac{1}{2}\delta(-\omega_o) \\ &+ u\,\delta(\omega_o+\eta) + l\,\delta(\omega_o-\eta) \\ &+ u^*\delta(-\omega_o-\eta) + l^*\delta(-\omega_o+\eta) \end{aligned}$$

$$l = \frac{\epsilon_o}{4}e^{-j\alpha} - \frac{\phi_o}{4j}e^{-j\beta}$$
$$u = \frac{\epsilon_o}{4}e^{j\alpha} - \frac{\phi_o}{4j}e^{j\beta}$$

Where the terms **u** and **l** contain the upper and lower sideband products due to the AM and PM modulation.

Test Setup Model

The modulated oscillator output is connected to the input of the test setup. Since networks X and Y are linear, the output of those networks in the frequency domain is obtained by multiplying the oscillator spectrum by the network scattering parameters. If the X and Y are networks real then the following is true.

$$S(\omega) = S^*(-\omega)$$

The outputs of networks X and Y are multiplied in the time domain by the mixer. The equivalent operation in the frequency domain is convolution. Performing the convolution and keeping only the terms at the modulation frequency yields:
Terms at twice the modulation frequency appear but are at levels proportional to the square of the AM and PM levels. These terms are extremely small and have been neglected. DC and double frequency terms have also been dropped. Dividing the above result by the oscillator modulation functions yield the desired system response functions.

$$Z(\omega) = (1h_1 + u^* h_u)\,\delta(-\eta) + (1^* h_1 + u h_u)\,\delta(\eta)$$

$$h_1 = \frac{1}{2}\left(S_x^*(\omega_o)\,S_y^*(\omega_o - \eta) + S_x(\omega_o - \eta)\,S_y^*(\omega_o)\right)$$

$$h_u = \frac{1}{2}\left(S_x^*(\omega_o)\,S_y(\omega_o + \eta) + S_x(\omega_o + \eta)\,S_y^*(\omega_o)\right)$$

$$H_\phi(\eta) = \frac{(h_1^* - h_u)}{2j} \quad \textit{PM RESPONSE} \tag{2}$$

$$H_\epsilon(\eta) = \frac{(h_1^* + h_u)}{2} \quad \textit{AM RESPONSE} \tag{3}$$

The power spectral density of the noise at the mixer output Z is found by inserting equations 2 and 3 into equation 1. The ideal PM noise measurement setup would set equation 2 equal to one and equation (3) equal to zero. None of the configurations described in this paper achieve those goals. The ideal AM noise measurement setup is a square law detector where $S_x = 1$ and $S_y = 1$. Resulting in equation (2) equal to zero and equation 3 equal to one.

The performance of a delay line discriminator can be derived by inserting the following scattering parameters for a short and delay line respectively.

$$S_x = 1/\sqrt{2}$$

$$S_y = \frac{e^{j\omega t}}{\sqrt{2}}$$

The 3db power divider loss is included in these equations. The resulting AM and PM responses match previously published results [4].

PHASE NOISE MEASUREMENT USING

$$\| H_\epsilon \|^2 = \cos^2(\omega_o \tau)\cos^2(n\tau/2)$$

$$\| H_\phi \|^2 = \sin^2(\omega_o \tau)\sin^2(n\tau/2)$$

RESONANT CAVITIES

The performance of phase noise measurement setups using resonant cavities can be evaluated using cavity scattering parameters and the results of equations 1,2 and 3. The following formulae describe the transmitted and reflected waves of a two port cavity network.

$$S_{21}(\omega) = \frac{t}{1 + jQ_1(\omega/\omega_c - \omega_c/\omega)} \quad \textit{TRANS.}$$

$$S_{11}(\omega) = \frac{r + jQ_1(\omega/\omega_c - \omega_c/\omega)}{1 + jQ_1(\omega/\omega_c - \omega_c/\omega)} \quad \textit{REF.}$$

$r = S_{11}(\omega_c)$ *RETURN AT RESONANCE*
$t = S_{21}(\omega_c)$ *TRANSMISSION*
ω_c = *CAVITY RESONANT FREQUENCY*
Q_u = *CAVITY UNLOADED Q*

$$Q_1 = \frac{Q_u}{1 + k_i + k_o} \quad \textit{LOADED Q}$$

$$k_i = \frac{(1-r)^2}{1 - r^2 - t^2} \quad \textit{INPUT COUPLING}$$

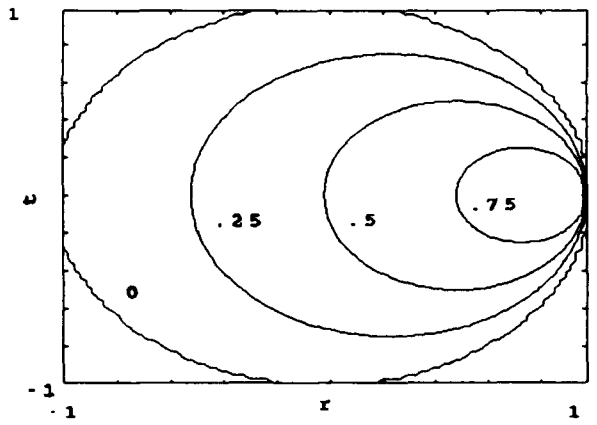
$$k_o = \frac{t^2}{1 - r^2 - t^2} \quad \textit{CROSS COUPLING}$$

The coupling parameters **r** and **t** control the cavity coupling and the loaded Q of the cavity. The following plot shows the ratio of loaded Q to unloaded Q as a function of **r**(horiziontal axis) and **t**(vertical axis). Energy conservation requirements limit **r** and **t** to inside the unit circle.

$$1 \geq r^2 + t^2 \quad \textit{ENEGRY CONSERVATION}$$

PERFORMANCE COMPARISONS

The equations developed above for test setup performance and network scattering were computer analyzed. The results reflect the performance of setups with ideal



RATIO Q_l/Q_U vs. t and r

components. No mismatches or component noise effects are included. Three cavity discriminator and one delay line test setup are compared.

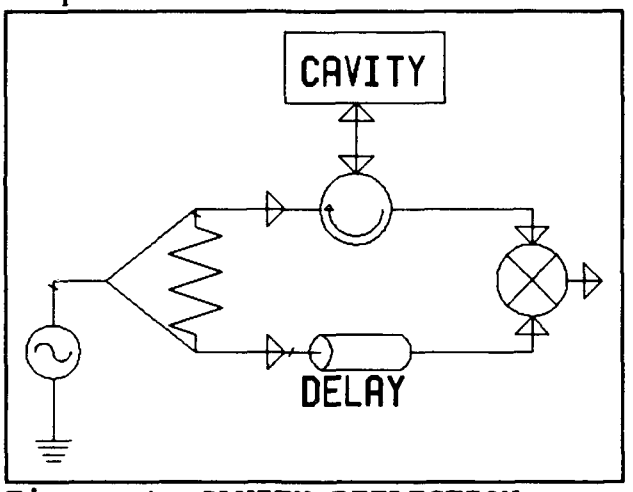


Figure 4: CAVITY REFLECTION

The first cavity discriminator setup and the delay line setup have been described by Walls [1] and others. The second and third cavity setups were proposed by Dick [5] as STALO oscillator configurations. Figures 4 and 5 shown the second and third cavity discriminator configurations. Equations 2 and 3 may also be applied to the phase stabilization loop of a STALO to determine

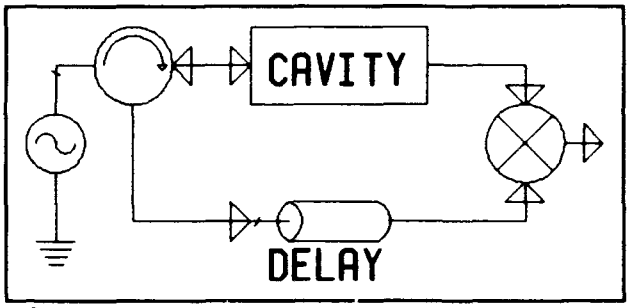


Figure 5: CAVITY TRANS. & REF.

the magnitude and phase response of the phase stabilization loop. The cavity coupling parameters used in the following figures were selected by computer analysis. Values were selected that yielded the maximum PM response at low noise frequencies. These operating points, given in table I, caused the cavity Q to fall to one quarter of the unloaded Q.

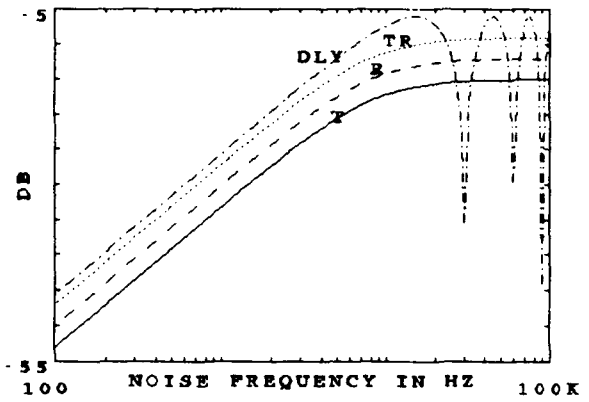
Table I Setup analysis data

	r	t
transmission	0.0	0.707
reflection	-0.5	0.0
trans. & ref.	0.183	0.683

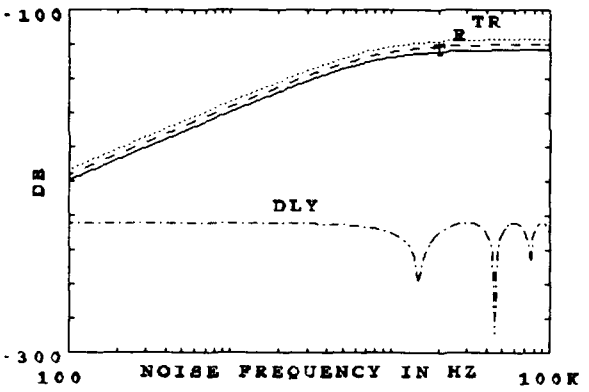
cavity Q_u=250,000
cavity f_c=1 GHZ

delay line =33.3 usec

The following two figures plot the setup AM and PM responses in DB on the vertical axis. The horizontal axis is the noise frequency. The cavity discriminator setups proposed by Dick [5] yield better PM performance than the standard transmission configuration. All the cavity configurations show reduced AM rejection at large frequency offsets. The sensitivity of the cavity discriminator setups is improved by increasing the Q of the resonator as described by Walls [1].



PHASE NOISE RESPONSE



AM NOISE RESPONSE

SUMMARY

This paper has developed a technique for estimating the performance of AM and PM carrier noise measurement setups. Setup performance can be estimated using computer models for the components or actual measured S parameter data. Computer modeling of the networks and using actual measured data has the advantage that the effects of mismatches are included in the results. This analysis assumed ideal components with perfect matching and no internal noise contributions. Hence the results of this paper represent theoretical setup performance; real components with mismatches and internal 1/f noise will degrade performance.

ACKNOWLEDGEMENT

This work was supported by the Time and Frequency group in the Space Department of the Johns Hopkins University Applied Physics Laboratory.

REFERENCES

[1] F. Walls, A. Clements, C. Felton, M. Lombardi and M. Vanck, "Extending the Range and Accuracy of Phase Noise Measurements," in Proceedings of the 42 Annual Symposium on Frequency Control, 1988.

[2] S. Stein, "Frequency and Time - Their Measurement and Characterization," in Precision Frequency Control, Vol. 2, New York, Academic Press, 1985, pp. 191-416.

[3] P. Lesage and C. Audion, "Characterization and Measurement of Time and Frequency Stability," Radio Science, Vol. 14, No. 4, pp. 521-539, 1979.

[4] A. Lance, W. Seal and F. Labar, "Phase Noise and AM Noise Measurements in the Frequency Domain," Infrared and Millimeter Waves, Vol. 11, pp. 239-289,1984.

[5] G. Dick, J. Saunders and T. Tucker, "Ultra-Low Noise Microwave Phase Stabilizer Using Sapphire Ring Resonator," in Proceedings of the 44th Annual Symposium of Frequency Control, 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ULTRALINEAR SMALL-ANGLE PHASE MODULATOR

John Lowe and F. L. Walls

National Institute of Standards and Technology
Boulder, Colorado 80303

Abstract

A dc electric field applied to a quartz plate resonator causes changes in the elastic constants which can lead to a change in the frequency of the resonator. This effect, known as the polarizing effect, has been shown to be extremely linear. We have used this effect to build a phase modulator with 2nd-harmonic distortion that is at least 117 dB below the fundamental modulation and low added phase noise. A description of the modulator as well as methods of measurement are discussed.

INTRODUCTION

To find the center of a resonance curve accurately, either phase or frequency modulation techniques are used to guide the servo [1-6]. As the resolution requirements increase so do the requirements on the spectral purity of the modulation reference and the linearity of the modulator. 2nd-harmonic distortion in sine wave modulators can lead to offset errors. The design goals for NIST-7, a new optically pumped cesium beam frequency standard, is that the servo resolve line center to an accuracy of 1 ppm. This requires that the modulation reference and the modulator maintain 2nd-harmonic distortion approximately 114 dB below the fundamental modulation [1,2]. The 2nd-harmonic distortion requirements on the demodulator are approximately -65 dB and not a significant problem [1,2]. Traditional methods of phase modulation have not demonstrated such high linearity; thus a new technique is required. We describe a new type of small-angle analog phase modulator, based on the polarizing effect in quartz resonators, which is exceptionally linear [7-12]. Measurements on three such devices show that the 2nd-harmonic distortion is at least 117 dB below the fundamental modulation. The phase modulator has very low added noise so as to minimize noise contributions at the 2nd-harmonic of the modulation frequency [3]. We also discuss the measurement techniques to verify this performance.

MODULATION ERRORS

If phase modulation of the form

$$\Delta\phi = B \cos \Omega t \qquad (1)$$

is used and the error detection is done at the fundamental, 2nd-harmonic distortion with phase cos $2\Omega t$ leads to offset errors [1,2]. Real phase modulators have small nonlinearities which generate small components of modulation at multiples of the modulation frequency.

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

Assume that the realized phase modulated signal is of the form

$$\omega = \omega_1 + B \cos \Omega t - M_{2s} \sin 2\Omega t + M_{2c} \cos 2\Omega t, \qquad (2)$$

where $\Omega/(2\pi)$ is the modulation frequency and $\omega_1/(2\pi)$ is the average frequency of the probe oscillator. Coefficients M_{2s} and M_{2c} contain the effects of second harmonic distortion in the modulation process under the assumption that the residual modulation at harmonics of Ω in the reference signal are small compared to that imposed by the modulator. Neglecting higher harmonic effects and nonlinearities in the detection process, the offset in the error signal is [2]

$$\text{Frequency offset} = -1/2\ M_{2c}. \qquad (3)$$

The NIST-7 servo requires resolving line center to an accuracy of 1 part in a million. Therefore the 2nd-harmonic distortion of the phase represented by M_{2c} must be less than $2 \times 10^{-6} B$ or at least 114 dB below the fundamental modulation [1,2].

POLARIZING EFFECT MODULATOR-THEORY

When a vibrating piezoelectric resonator is subjected to a dc electric field, ΔE, it responds by a change, $\Delta\nu$ in its resonance frequency ν [7-13]. The polarizing effect in quartz resonators is so nearly linear that it can be adequately described using only the 1st- and 2nd-order terms of the Taylor expansion about $E = 0$. Therefore,

$$\Delta\nu/\nu = P_L\, \Delta E + P_Q\, \Delta E^2, \qquad (4)$$

where

$$P_L = 1/\nu(d\nu/dE) \text{ at } E=0 \text{ and} \qquad (4A)$$

$$P_Q = 1/(2\nu)\ (d^2\nu/dE^2) \text{ at } E=0. \qquad (4B)$$

For a resonator to exhibit a polarizing effect there must be a component of E along the x-axis of the quartz resonator. Additional discussion including the effects of ion migration altering the sensitivity for dc applied electric fields are found in [11-13]. Typical values for P_L and P_Q are on the order of 10^{-11} to 10^{-12} and 10^{-20} to 10^{-21}, respectively, depending on the crystal cut and overtone used [7-9].

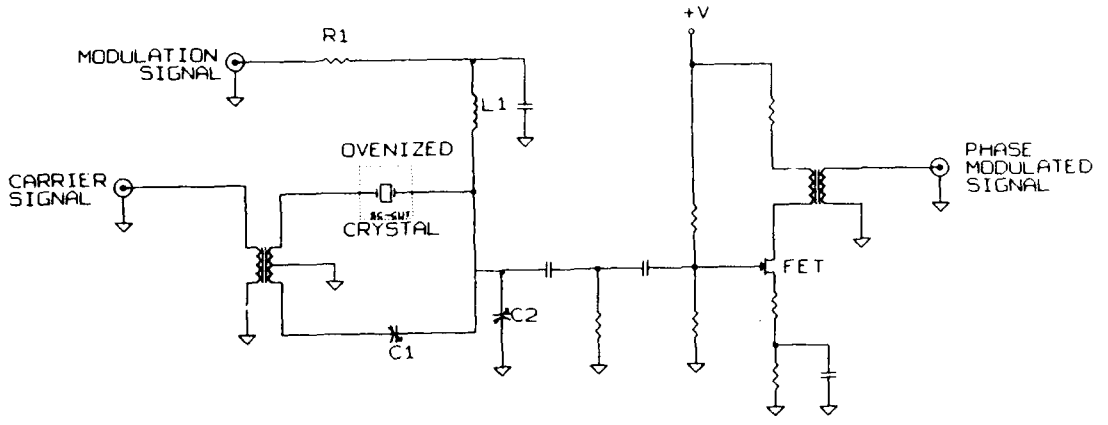


Figure 1. Block diagram of the new phase modulator.

Near line center the steady state phase shift across a high Q resonator is

$$\Delta\phi = 2Q\,(\nu - \nu_0)/\nu_0, \tag{5}$$

where ν is the applied frequency, ν_0 is the resonance frequency of the resonator, and Q is the loaded quality factor of the resonator. Thus a frequency change $\Delta\nu$ in the frequency of the resonator generated by applying an electric field E across the resonator results in a phase shift

$$\Delta\phi = 2Q(P_L\,\Delta E + P_Q\,\Delta E^2). \tag{6}$$

POLARIZING EFFECT MODULATOR-PRACTICAL REALIZATION

Figure 1 shows a generalized schematic of the polarizing phase modulator. A balanced transformer and adjustable capacitor C_1 are used to minimize off resonance transmission. A 3rd overtone SC-cut 5 MHz resonator was chosen because the polarizing coefficient was much larger [8] and both the static and dynamic temperature coefficients are much better than for comparable AT-cut units [13]. The resonator was temperature controlled in a standard crystal oven at 80°C and tuned to resonance by adjusting C_2. Tuning to resonance was determined by nulling the amplitude modulation observed on a detector diode placed at the output when a slow modulation voltage is applied across the resonator through R_1 and L_1. No other alignment or adjustment is necessary. The isolation amplifier on the output of the modulator provides overall unity gain and removes the effects of changing loads on the tuning of the resonator.

For NIST-7, a phase modulation of approximately 0.55 mrad is required at 5 MHz to achieve a phase modulation of 1 radian at the final output frequency of 9.2 GHz. The desired modulation frequency $\Omega/(2\pi)$ = 49 Hz is much larger than the half-bandwidth of the resonator. At such high modulation frequencies the phase shift across the resonator becomes

$$\Delta\phi = [2Q(P_L\Delta E + P_Q\Delta E^2)][1 + ((Q\Omega/(\pi\nu_0))^2]^{-1}, \tag{7}$$

where $\nu_0/(2Q)$ is the loaded half-bandwidth of the resonator.

Figure 2 shows the relative amplitude of the phase modulation as a function of modulation frequency. These results follow Eq. (7) within ± 0.5 dB for modulation frequencies from 0.1 up to at least 400 Hz.

Approximately 7 V (6400 V/m) RMS is required to achieve the desired modulation level using a, 3rd overtone SC-cut 5 MHz quartz crystal and a modulation frequency of 49 Hz. This corresponds to a dc linear polarizing sensitivity of P_L = 2.6 $\pm$0.3 x 10^{-12} m/V. The difference between our data (which is about a factor of 5 smaller) and that found by Hruska [7,8] is probably due to the difference in the electrode patterns, cut angle, and the shape. The measurements of P_L for a particular resonator can easily be made to an accuracy of a few percent with our technique, which is free from effects due to ion migration [11-13]. From the relative values of P_L and P_Q from [8] and the shape of Fig. 3 we would expect that the second harmonic distortion due to the resonator to be approximately 126 dB below the fundamental. Even lower harmonic distortion would be expected if the modulation frequency was lower since a lower voltage would be required to achieve the same phase modulation angle.

The added phase noise is somewhat lower (typically $\mathcal{L}(10)$ = -134 dBc/Hz and $\mathcal{L}(100)$ = - 154 dBc/Hz) than the present 5 MHz oscillator. (Techniques for measuring the added phase noise of components are discussed in [14,15].) Further improvements in the phase noise appear possible. The added white frequency contribution of a passive standard due to the present level of added phase noise at the second harmonic of the modulation frequency (2Ω = 100 Hz) is approximately 2 x 10^{-13} $\tau^{-1/2}$ [3].

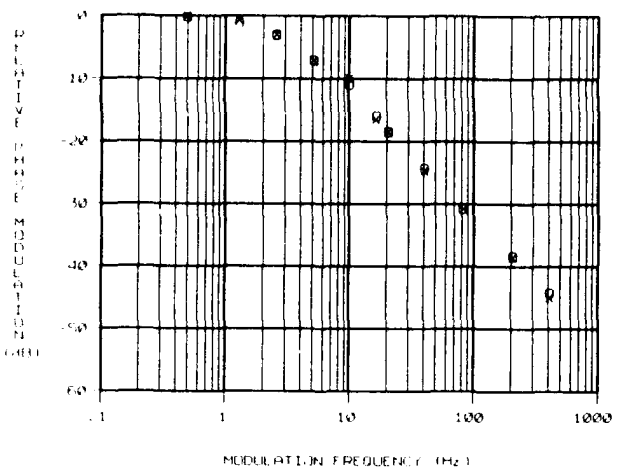


Figure 2. Relative phase modulation of the output signal as a function of the frequency of modulation. The open circles are the predictions using Eq. (7) and the half-bandwidth $\nu_0/(2Q)$ measured from transmission experiments. The solid points are the phase modulation levels obtained from phase bridge measurements [14,15].

MEASUREMENTS OF HARMONIC DISTORTION

The dynamic range of most spectral analyzers--from 60 to 100 dB--is not enough to observe 2nd-harmonic levels expected to be 120 dB below the fundamental. Some filtering is therefore required to extend the dynamic range to this level. The general approach is to arrange the various tests so that the fundamental modulation signal is nulled against another signal by at least 40 dB before interacting with a phase detector, spectrum analyzer, or other nonlinear device. We compared 3 different devices to determine the distortion characteristics of each modulator. Figure 3 shows the test setup used to compare the difference in phase modulation of two modulators. The mixer is sensitive to only the difference in the phase modulation level between the two modulators. The input reference level to one modulator is adjusted to achieve the required 0.55 mrad phase modulation. The input modulation signal to the other modulator is adjusted in amplitude and phase to null the demodulated fundamental signal at the output of the mixer. Typical reduction of the fundamental was 40 to 60 dB.

The out-of-phase distortion products between two modulators at frequency Ω/π were measured by driving the second modulator with the output signal from the first modulator, applying modulation signals that are 180° out of phase and using the unmodulated signal from the power splitter to drive the reference side of the mixer. See Fig. 4. Great care must be taken not to introduce spurious signals due to residual amplitude modulation since typical double balance mixers only provide 30 to 40 dB of AM suppression. The worst case 2nd-harmonic distortion between the three modulators was -117 dB relative to the fundamental modulation. A typical result is shown in Fig. 5. The fundamental modulation corresponds to -6 dBV, yielding anet sum for the second harmonic distortions of -126 dB for this pair.

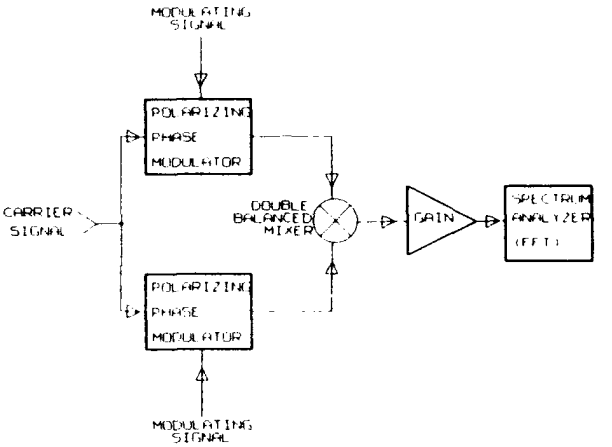


Figure 3. Block diagram of a parallel phase bridge measurements to determine the difference in phase modulation between two phase modulators [14,15].

In this configuration the output of the mixer is of the form

$$\phi_{mod} = V_0\cos\omega t + (B_1\text{-}B_2)\cos\Omega t + (M^1_{2C} + M^2_{2C})\cos 2\Omega t + \quad (8)$$

$$(M^1_{2S} + M^2_{2S})\sin 2\Omega t$$

under the assumption that all of the 2nd harmonic distortion is due to the modulation process.

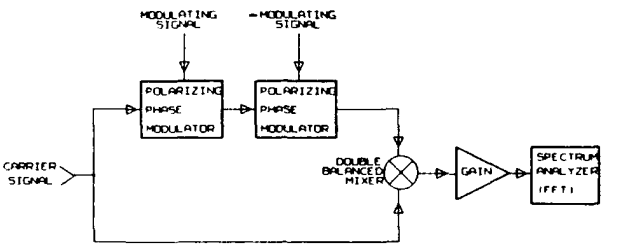


Figure 4. Block diagram of a series phase bridge measurements to determine the sum of the 2nd harmonic in phase modulation between two phase modulators [14,15].

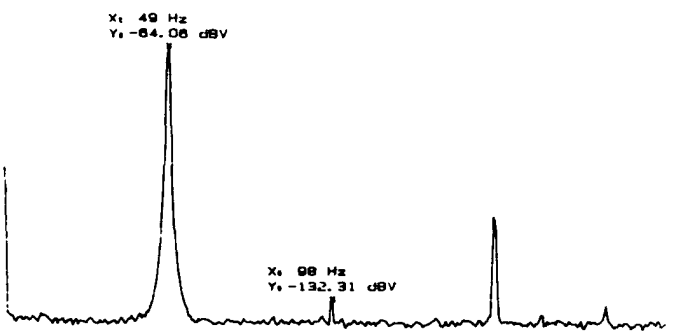


Figure 5. Typical result of a series phase bridge measurement.

CONCLUSIONS

We have introduced a method for measuring the polarizing effect in high-Q SC-cut 5 MHz quartz resonators for small ac electric fields and found that the sensitivity and the linearity is in general agreement with the previous work at dc [7-11]. We have demonstrated that this polarizing effect can be used to make an ultralinear small angle phase modulator with low added phase noise. The second harmonic distortion was measured to be less than -117 dB relative to the fundamental in three different units even though the modulation frequency was more than 20 times larger than the half-bandwidth of the resonators. Using low modulation frequencies, the second harmonic distortion could probably approach -140 dB relative to the fundamental modulation. The high linearity, low noise, and simplicity of the circuit combine to make this new small angle analog phase modulator an interesting tool in precision frequency metrology.

ACKNOWLEDGEMENTS

We are grateful to K. Hruska for the initial measurements which stimulated this work and to J.-J. Gagnepain for fruitful discussions.

REFERENCES

1. F. L. Walls and S. R. Stein, "Servo Techniques in Oscillators and Measurement Systems," NBS Tech Note 692, 1976.
2. F. L. Walls, "Errors in Determining the Center of a Resonance Line Using Sinusoidal Frequency (Phase) Modulation," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 592-598 (1987).
3. C. Audoin, V. Candelier, and N. Dimarco, "A Limit to the Frequency Stability of Passive Frequency Standards," CEPM Digest, 93, IEEE Pub. 90CH2822-5.
4. A. DeMarchi, G. D. Rovera, and A. Premoli, "Effects of Servo Loop Modulation in Atomic Beam Frequency Standards Employing a Ramsey Cavity," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 582-592 (1987).
5. C. Audoin, V. Candelier, and J. Vanier, "Effect of the Atom Transit Time on the Frequency Stability of Cesium Beam Frequency Standards," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 573-582 (1987).
6. J. L. Hall, "Stabilized Lasers and Precision Measurements," Science, vol. 202, pp. 147-156 (1978).
7. K. Hruska, "Polarizing Effect with Piezoelectric Plates and Second-order Effects," IEEE Trans. Sonics Ultrason., Su-18, 1-7 (1981).
8. C. K. Hruska, "On the Linear Polarizing Effect with α-quartz AT Plates," IEEE Trans. Sonics Ultrason., Su-28, 108-110 (1981).
9. C. K. Hruska and M. Kucera, "The Dependence of the Polarizing Effect on the Frequency of Quartz Resonators," Journal of the Canadian Ceramic Society, vol. 55, pp. 38-41, 1986.
10. J. Brendel, "Material Nonlinear Piezoelectric Coefficients for Quartz," J. Appl. Phys., 54, 5339-5346 (1983).
11. R. Brendel and J.-J. Gagnepain, "Electroelastic Effects and Impurity Relaxations in Quartz Resonators," Proc. 36th Ann. Frequency Control Symposium, June 1982, pp. 97-107.
12. J.-J. Gagnepain, "Sensitivity of Quartz Oscillators to the Environment: Characterization Methods and Pitfalls," IEEE Trans. UFFC-37, 347-354 (1990).
13. F. L. Walls and J.-J. Gagnepain, "Environmental Sensitivities of Quartz-Crystal-Controlled Oscillators," to be published UFFC, 1991.
14. Ed. by D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, "Characterization of Clocks and Oscillators," NIST Tech Note 1337, March 1990.
15. F. L. Walls, C. M. Felton, A. J. D. Clements, and T. D. Martin, "Accuracy Model for Phase Noise Measurements, Proc. of 21st Ann. Precise Time and Time Interval Planning Meeting, Redondo Beach, CA, Nov. 1990, pp. 295-310.

STOCHASTIC RESONANCE: A SIGNAL + NOISE IN A TWO STATE SYSTEM

Frank Moss
Department of Physics
University of Missouri at St. Louis
St. Louis, MO 63121

The term *Stochastic Resonance* (SR) describes an interesting stochastic phenomenon which occurs in bistable systems when driven by a weak periodic signal embedded in noise of comparable intensity. The "resonance" occurs when the period of the signal becomes comparable to an internal time scale established by the noise. For such a condition, the signal-to-noise ratio of the response of the bistable system is maximized.

Introduction

The idea of SR was first introduced by a group of Italian theorists[1,2] who proposed it as a possible explanation for the observed periodicity in the recurrences of the Earth's Ice Ages[3]. In this model, the Earth's global climate is represented by a model which consists of an energy potential having two local minima separated by a barrier. The locations of the two minima determine the two states, one metastable (representing the ice age) and the other globally stable (representing the Earth's normal climate). Random fluctuations from a "noisy" solar constant are supposed to drive this bistable system, so that switching events from one well to the other, for example, from normal to ice age climate and the reverse, occur at completely random times. However, the observed periodicity in the ice age recurrences at approximately 100,000 year intervals means that this seemingly random sequence must, in fact, be correlated to some degree with some periodic dynamical force to which the global climate system is exposed. In climate models, this periodic force is assumed to originate in a weak but periodic variation of the eccentricity of the Earth's orbit, the period of which is $\cong 10^5$ years.

The dynamics of SR are then established in a bistable potential by a stochastic force called the "noise", of magnitude comparable to that of the height of the barrier separating the two local minima, plus a weak periodic force often called the "signal", which acting together cause switching between the two states of the system. The switching events are noisy, but not completely random, being correlated to some degree with the signal. We can look a bit more closely at this phenomenon with the aid of Fig. 1, where I have depicted a typical bistable potential being "rocked" periodically by the signal.

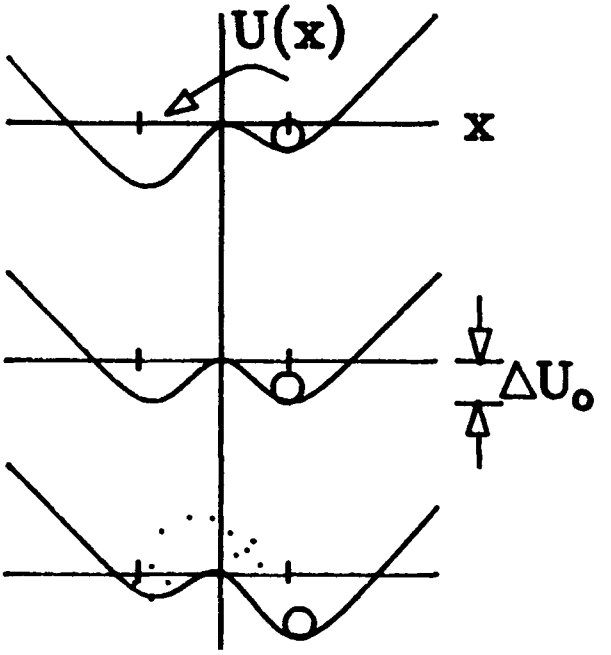


Fig. 1. A bistable potential with an added periodic component provided by a weak signal depicted at successive quarter cycle intervals of the signal. Switching events can only take place in the presence of noise as indicated by the arrows.

CH2965-2/91/0000-649 $1.00 © 1991 IEEE

The location of a "ball" near the bottom of one well represents the system state point. If the amplitude of the signal is smaller than the height of the barrier, the ball will not change states under the influence of the signal. If however, we now add a small amount of "noise", which we shall assume to be Gaussian distributed and white, we see that there is now always a non zero probability that the ball will change state. However, because the potential is being rocked by the signal, i.e. the barrier is alternately being raised and lowered at every half cycle of the signal, the probability of a switching event taking place is modulated by the signal. It is this modulation which causes the switching events to become correlated with the signal.

Moreover, if we interpret the sequence of switching events as the output of information which passed through the system, then we see that no information can be transmitted to the output in the absence of noise. It is also clear that in the limit of very large noise intensity the information content in the output will be degraded. There must be an optimum value of the noise intensity for which the information content is maximized. In the case of a single frequency signal, the information content is the signal frequency ω and its intensity A. A common measurement in SR experiments is the power spectrum of the output, from which the signal-to-noise ratio (SNR) can be obtained.

An early demonstration of SR in an electronic system was realized in a Schmitt trigger[4], but after this interest seems to have waned for a number of years. Recently, however, a demonstration experiment with a bistable ring laser was accomplished[5]. Publication of this observation stimulated renewed interest, including a number of theoretical works[6-10] and two analog simulations[11-14].

A dynamical system representing an infinitely damped particle moving in the standard quartic potential with an added single frequency signal provides a simple demonstration of SR. The potential, schematically depicted in Fig. 1, is given by

$$U(x) = -\frac{x^2}{2} + \frac{x^4}{4} - xA\sin\omega t, \qquad (1)$$

where the magnitude of the of the unperturbed barrier is $\Delta U_0 = 1/4$. The dynamics of a particle moving in this potential with infinite damping and under the influence of the noise $\xi(t)$ is determined by the Langevin equation

$$\dot{x} = -\frac{dU}{dx} + \xi(t)$$
$$= x - x^3 + A\sin\omega t + \xi(t), \qquad (2)$$

where the noise is assumed to be Gaussian distributed with zero mean and of infinite bandwidth, i.e. "white". The noise intensity D is defined by its autocorrelation function

$$\langle \xi(t)\xi(s)\rangle = 2D\delta(t - s). \qquad (3)$$

This simple example will be used throughout this paper, however more complex cases involving particles with inertia[9,15] as well as band limited, or "colored", noise[9] have been treated. The Langevin equation governing the inertial case is given by

$$\ddot{x} + \gamma\dot{x} + \frac{dU(x)}{dx} = A\sin\omega t + \xi(t). \qquad (4)$$

If the noise is colored, its intensity is defined by the exponential correlation function,

$$\langle \xi(t)\xi(s)\rangle = \frac{D}{\tau}\exp[-|t-s|/\tau], \qquad (5)$$

with correlation time τ.

Stochastic resonance has also been demonstrated for potentials other than the so called "standard quartic" given by Eq. (1), a recent example being the "soft" potential[16], often used in neuron models

$$U(x) = \frac{x^2}{2} - \eta\ln(\cosh x) - xA\sin\omega t - x\xi(t). \qquad (6)$$

This potential is called "soft", because it increases only linearly with x for large enough x; in contrast, for example, to the infinitely "hard" Schmitt trigger potential for which $U(x) \to \infty$, as $x \to \pm x_0$; where the two values of x_0 are the two states of the trigger. SR has also been demonstrated in the Schmitt trigger[4,6]. In addition to the bistable ring

laser[5], SR has recently been demonstrated also in an electron paramagnetic resonance system[17].

Theoretical Issues

In the absence of an external signal (A = 0), but with only the noise present, stochastic differential equations of the general type of Eq. (2) are solved for the stationary probability density $P_0(x)$, of the variable x, with the Fokker-Planck equation[18-20]:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[x - x^3]P(x,t) + D\frac{\partial^2}{\partial x^2}[P(x,t)] \quad (7)$$

where the bistable forcing is given by $F(x) = -dU(x)/dx = x - x^3$, for the case that $U(x) = -x^2/2 + x^4/4$ is not a function of time, and $D \neq D(x)$ [for a recent example where $D = D(x)$, see Ref. 21]. The stationary solution is obtained from

$$\frac{\partial P(x,t)}{\partial t} = 0, \quad (8)$$

which results in the exact solution,

$$P_0(x) = \sqrt{\frac{1}{\pi D}}\exp\left[-\frac{U(x)}{D}\right]. \quad (9)$$

Under these conditions, the mean rate of switching between wells is given by the Kramers rate[22]:

$$r_0 = \left[\frac{1}{\pi\sqrt{2}}\right]\exp\left[-\frac{\Delta U_0}{D}\right], \quad (10)$$

where ΔU is the barrier height in the absence of a signal, as shown in Fig. 1. The times at which switching events occur are completely random.

Unfortunately, the Fokker-Planck equation can be solved exactly for only a very restricted dynamics: one dimensional, stationary, and infinitely damped. This means that problems involving inertia, such as Eq. (4), or colored noise as in Eq. (5), or problems with time dependent potentials as shown in Eq. (1), must be treated in some approximation. These approximations are challenging, and for that reason, SR has been of interest to a number of theorists. An historically important paper dealing with time dependent stochastic processes was written in 1975 by Hänggi and Thomas[23]. A recent collection of papers deals with such approximation techniques and their applications to various physical problems as well as actual experiments and simulations[24]. The most straight forward approach to the problem of the time dependent potential is the "adiabatic" approximation, $\omega << r_0$. This means that the signal changes so slowly that we can consider the Fokker-Planck equation as relaxing to its stationary solution at every instant of time. Moreover, if the noise is assumed to be white, and the dynamics infinitely damped, then the problem is one dimensional. A further simplification can be achieved by considering a reduced dynamics: the "two state" dynamics. In this approximation, the only information required at the output of the SR system is which well the state point currently occupies. Motions of the state point within each well are specifically ignored and, in fact, replaced with a constant, for example, $x(t) = \pm x_0$, where the $\pm$ sign identifies which well. The exact form of the potential is no longer of consequence, and the only parameters are the unperturbed barrier height ΔU_0 and the steady state points $\pm x_0$. [It should be noted that the two state approximation is exact for the Schmitt trigger, which is the closest and simplest realization of an ideal two state system.]

We shall not here recite further the details of the various theoretical approximations, but instead refer the reader to the original literature. To date all theories have been restricted to Gaussian, white noise. The first modern theory, based on the adiabatic approximation and infinitely damped two state dynamics was due to McNamara and Wiesenfeld[6]. They provided very useful formulae which are most often used today for the interpretation of experimental results. Important non adiabatic results, valid also for the full dynamics, were the predictions by Hänggi and Jung[7,10] that the ideal line shape of the signal in the power spectrum was a delta function, and that for symmetric potentials (as, for example, the standard quartic, or

the neuron) there exists a sequence of lines located at odd harmonics of the signal frequency with exponentially decreasing strength. These results were later independently verified by simulations[12,13]. Additionally, Hänggi and Jung[15] and Marchesoni, et al[9,11] have obtained useful approximations and exact results for the case of inertial motion (i.e. non infinitely damped). Though a preliminary approximate result has been obtained for colored noise[11], no systematic study of SR dynamical systems driven by this important noise has been undertaken.

Recently, some calculations of the escape, or residence, times have been made. Here, we are concerned with the probability density of the times which the state point spends in one well, an object which can be easily measured on real physical systems[25]. An interesting generalization of this theory has recently been extended to include multi stable, or "washboard" potentials[26]. These results will have a direct bearing on future SR-type experiments, for example, in noisy, periodically modulated Josephson junctions. A reformulation of the problem in terms of Floquet theory has recently been accomplished and signal phase noise, which broadens the signal spectral line, has been treated in this framework for the first time[27]. The latter treatment is important, since most noisy physical signals to be encountered in the real world are expected to show phase noise as well as amplitude noise. A recent study of the phase relation between the signal and the response of an overdamped bistable system has been completed[28].

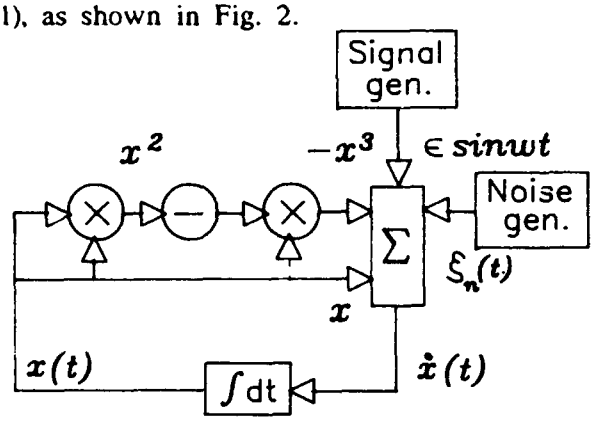
T.To date, all investigations of SR except one have been for a noisy signal composed of a sinusoidal function with added Gaussian white or colored noise. The single exception is an interesting study by R. Kapral and his associates[29] who looked at a damped oscillator moving in a bistable potential under the influence of what they call "periodic dichotomous noise". Periodic dichotomous noise is defined to be a function which can take on only two values, say $[a_0,a_1]$, with a fixed time, τ, between every switching event. The switching events are the stochastic variable, since they are defined to occur according to: $(a_1 \rightarrow a_0)$ with probability p, and $(a_0 \rightarrow a_1)$ with probability 1 - p. Entirely new resonance activation phenomena, showing fractal behavior, are revealed by the invariant probability densities P(x) of this system. The authors discuss various physico-chemical systems for which this new SR phenomenology might be realized.

Analog Simulations of Stochastic Resonance

Analog simulators are electronic circuits which are designed to mimic the dynamics of a certain Langevin equation. They are the modern reincarnations of the analog computers which were prevalent before the age of modern digital computing. Today modern chip technology has produced multipliers, for example the Analog Devices AD 534, multifunction chips which raise a voltage to an arbitrary power, trigonometric function converters, and logarithmic and antilog amplifiers with bandwidths in the hundreds of kHz combined with accuracies of a few tenths of a percent all on single chips. Rather than patch board programming an analog computer for each application, it is now convenient to simply build a single circuit for each application. While the accuracy is limited - usually to less than 2 percent in the non stochastic, steady state, and around 10 percent for noise driven dynamics where ensemble averaging of the appropriate quantity must be done - analog simulations offer advantages in speed, especially for multi dimensional stochastic problems over digital simulations. Moreover, large regions of the parameter space can quickly be scanned for interesting dynamical behavior, for example by turning the dial of a potentiometer. Finally, it is sometimes important that the analog simulator is after all a real physical system, replete with small asymmetries, noise, non ideal functional conversion, etc. This makes it impossible to observe singular effects, many of which can perfectly well be digitally simulated, for example any of a multitude of structurally unstable objects which have recently been predicted by nonlinear dynamicists.

In the typical application, the circuit outputs a set of voltages which are the approximate solutions of some set of Langevin equations, e.g. that of Eq. (2), (4) or (6). These are connected to analog-to-digital converters

(ADC) which digitize the appropriate voltages for a computer. At present we are using Data Translation DT2828 digitizer boards, mounted in 386-type PC's with ASYST software. An example is given by the schematic diagram of a circuit which mimics Eq. (1), as shown in Fig. 2.



$$\dot{x} = x - x^3 + A\sin\omega t + \xi_n(t)$$

Fig. 2. A schematic of the simulator of Eq. (1). The crosses represent analog multipliers. Integration is accomplished by collecting charge on a capacitor - the feedback element in an operational amplifier. The circuit is a closed loop system, with the dynamical approximation of Eq. (1) appearing as the voltages x(t) and the velocity $\dot{x}(t)$. The noise is supplied by a Quan Tech Model 400 noise generator, which produces Gaussian wide band (≅200 kHz) noise.

The ADC is connected to the output voltage x(t). In a typical scan, 4096 digitized points are accumulated at a rate of 20 μs per point. From this time series the power spectrum is computed and stored as a running sum. Ensemble averaging is accomplished by repeating this process as many times as necessary until the desired statistical accuracy is achieved. For the power spectra shown below, typically 200 such scans were used, however the probability densities shown later sometimes required up to 10,000 scans. It is important to note that the simulator as shown in Fig. 2 reproduces the full dynamics specified by Eq. (1), including the motion within the potential wells. This is shown by the time series reproduced in Fig. 3 (a). If, however, it is desired to examine the two state dynamics, a "two state filter" can be connected to the simulator output. This is simply a voltage comparator, which, in the present example replaces x(t) with + or - 1.0 V whenever x(t) > or < 0 respectively. The two state filtered output is shown in Fig. 3 (b). The resulting power spectrum is shown in Fig. 3 (c), where delta-like peaks show up at 500 and 1500 Hz, the latter being the first peak at the first odd harmonic of the signal frequency.

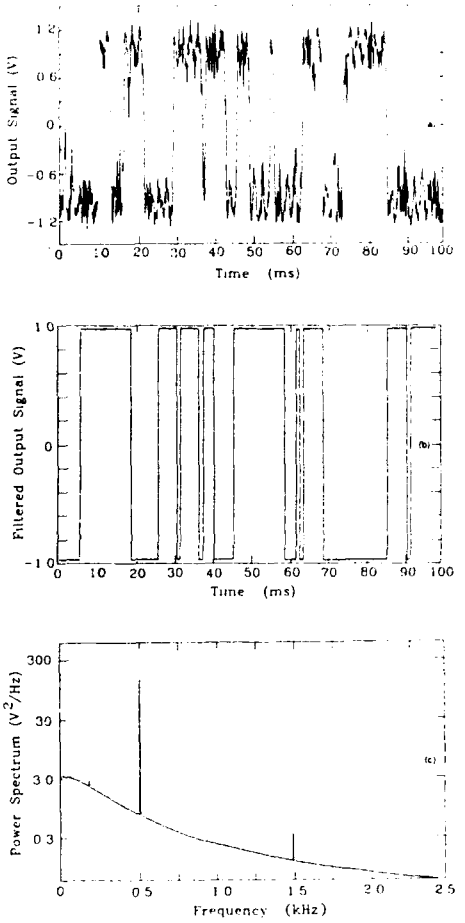


Fig. 3 (a) The measured output x(t) of the simulator shown in Fig. 2 for A = 0.35 V, $\langle\xi^2\rangle$ = 1.0 V^2, $\omega/2\pi$ = 500 Hz. (b) The same output (however, taken at a different time) after passing through the two state filter. (c) The power spectrum of the time series shown in (b) after averaging for 500 scans.

The signal for this simulation was provided by an ultra stable frequency synthesizer (1 part in 10^{-9}, short term stability). Our power spectra were assembled into 2500 frequency

bins of 1 Hz width each. The entire signal power, theoretically a delta function as predicted by Hänggi and Jung[7], was, therefore, concentrated into one bin and rendered a finite amplitude by integration of the delta function over a width of 1 Hz. The shape of the noise background is accurately a Lorentzian as predicted by McNamara and Wiesenfeld[6]. The signal-to-noise ratio (SNR) is obtained by measuring the amplitudes at the fundamental peak (S + N) and at the base of this peak (N). The SNR in decibels is then defined by

$$SNR = 10\log[(S + N)/N]. \qquad (11)$$

The results for this example are shown by the asterisk in Fig. 4, where we have plotted the SNR versus the noise intensity $\langle \xi^2 \rangle$. Note the prominent peak, which shows a 22 DB enhancement at an effective noise intensity of $\cong 0.1\ V^2$. The solid curve is the prediction of the adiabatic, two state theory of Ref. [6]. The plus signs show the results of measurements of the noise amplitude N, at the location of the signal frequency.

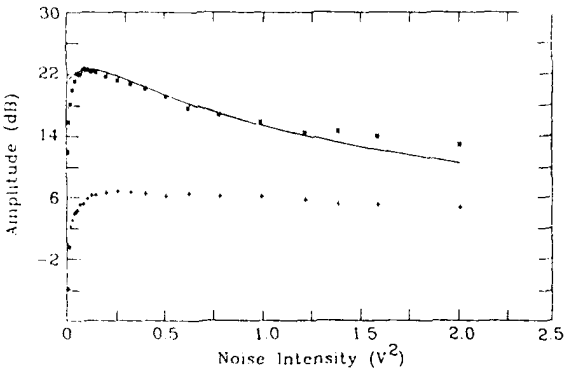


Fig. 4. The SNR (*) versus the noise intensity for the same conditions listed in the caption of Fig. 3. The amplitude of the noise alone at the signal frequency (+). The solid curve is the theoretical prediction.

Escape Time Measurements

Though the majority of experimentalists have always preferred to measure the power spectrum of their SR systems, recently a new object - the probability density of residence times - has been measured[13,25]. Figure 5 shows the output of a bistable system after two state filtering. For convenience, we identify the two states A and B as shown. The sequences of times labeled along the top and bottom of the figure indicate the only two possible residence time sequences available to a two state system. For convenience we define them as the ABBA sequence (upper set) and the ABAB sequence (lower set). Initially, we will concentrate on the ABBA sequence.

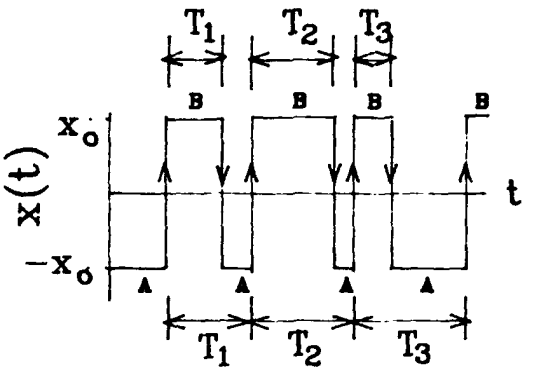


Fig. 5. The two residence time sequences. The top shown set, ABBA, is a measure of the time spent in one state only, in this example the B state. The lower set, ABAB, measures the residence time in A plus the next sequential residence time in B.

The residence times from such a sequence, either numerically or experimentally generated, can be assembled into a probability density P(T). We show, in Fig. 6, an example of such data measured on the analog simulator of Fig. 2 which mimics Eq. (2). A notable feature of these data is the repeated sequence of peaks of decaying amplitude located at odd integer multiples of $T_0/2$, where T_0 is the period of the signal. Data for the ABAB sequence (not shown here) are similar but phase shifted, the peaks being located at all integer multiples of T_0. The underlying reason for the difference in peak locations for the two different residence time sequences can be understood as follows: Consider the ABBA sequence. The switching events are random, but the most probable switching events are B → A when $\omega t = \pi/2$ (the signal is at its maximum positive amplitude), followed by the reset event A →

B at $\omega t = 3\pi/2$ (when the signal next arrives at its minimum value). The elapsed (or residence) time between these switching events is $T_0/2$. If, however, by chance the reset does not occur at $\omega t = 3\pi/2$, the next most probable reset event must wait until $\omega t = 7\pi/2$ (when the signal is next a minimum) so that the elapsed time (from the first switching event) is $6\pi/2$ or $3T_0/2$.

Figure 6 (a), (b) and (c) show the effect of increasing the noise intensity D.

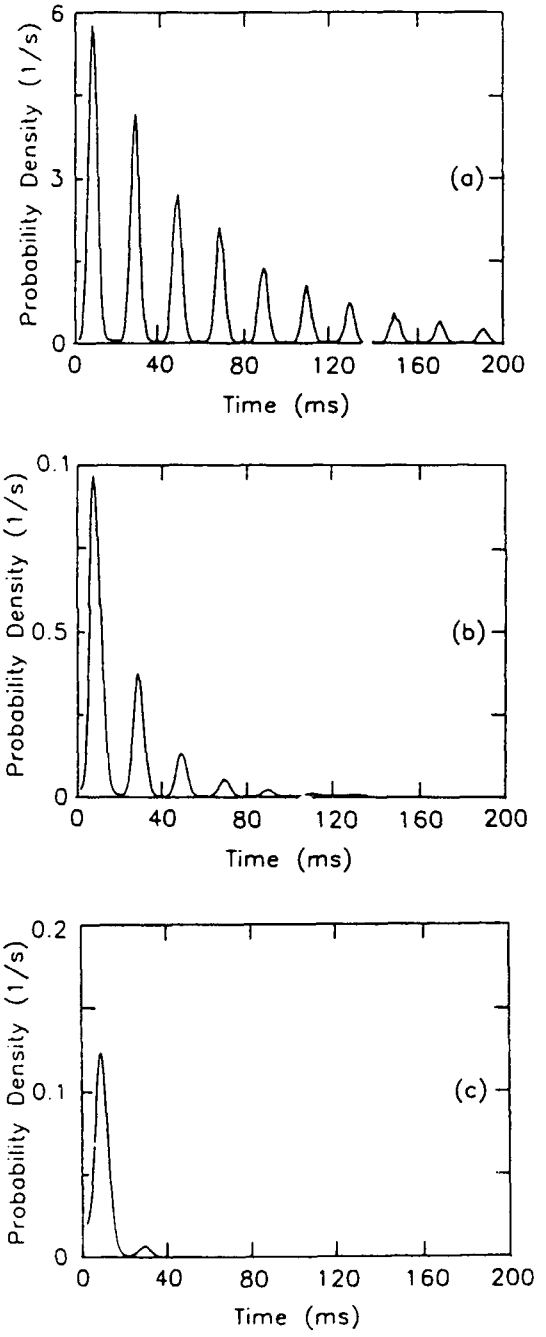


Fig. 6. Residence time probability densities measured on an analog simulator of overdamped motion in the standard quartic potential [Eq. (2)] for $\omega/2\pi = 50$ Hz and $A = 0.20$ V. The noise intensities are: (a) $D = 0.02$, (b) $D = 0.03$ and (c) $D = 0.05$.

We note the strong dependence of the decay constant on noise intensity. The decay is consistent with an exponential law. Such a law would be a strong indicaton that the higher order escape events are governed by rate processes similar to those encountered in chemical physics. In Fig. 7 we show the same data as in Fig. 6 (a) replotted on a semilogarthmic scale. Note that the straight line behavior of the peak amplitudes indicates a decay law,

$$P_{max} \propto \exp(-\lambda T). \qquad (12)$$

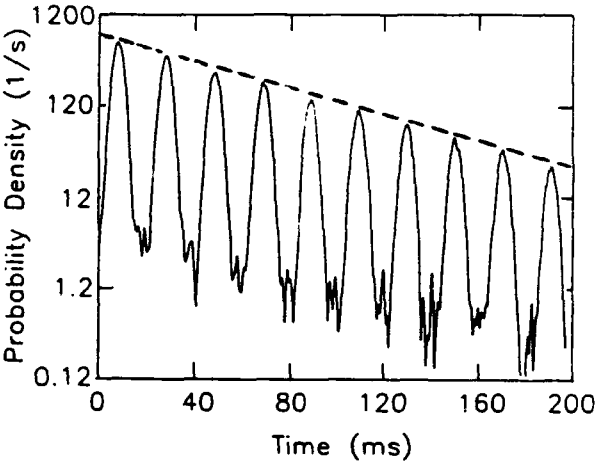


Fig. 7 Same data as Fig. 6 (a) replotted on a semilog scale. The peak-to-peak decrement of this measured data is 0.719 ± 0.003, which can be compared to the theoretical result calculated from Eq. (13) of 0.628. (The complete version of Eq. (13) is given in Ref. 25 where it is also shown that the discrepancy can be accounted for by an experimental under estimate of the value of D.)

An adiabatic theory for the residence time probability density has been put forward in Ref. 25. Here we quote only the result of that calculation to lowest order in the ratio A/D:

$$P_{ave} = \left[\frac{r_0}{\sqrt{2\pi A/D}}\right] \exp(A/D); T. \qquad (13)$$

where P_{ave} is P(T) averaged over all T. This quantity demonstrates the exponential decay law observed with the simulator. Moreover, the dependence on the potential barrier height ΔU_0 is built into this result through the unperturbed Kramers rate r_0 [see Eq. (10)]. Thus P(T) indicates the signal frequency ω, and is exponentially sensitive to the ratio A/D and ΔU_0. It therefore contains all the same information as the power spectrum, but is more sensitive to A/D but less precisely indicates ω.

There is also a "resonance" behavior of P(T). At low enough noise intensity, the first few peaks in P(T) become lower in amplitude than the higher order peaks. This is because the fundamental switching event becomes less probable than the next higher order one. That is, switching at every half cycle of the signal becomes less probable than switching at every three half cycles. (After all, for zero noise no switching is the most probable event.) The various peaks in the sequence, therefore, grow and decay with D. This is shown in Fig. 8, where the maximum amplitude of the second third and fourth peaks show the characteristic resonance peak at three different optimum values of the noise intensity.

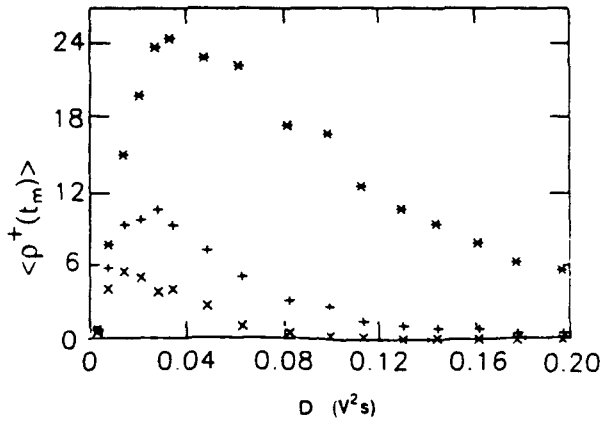


Fig. 8. The maximum amplitudes of the peaks in P(T) plotted versus D. The () signs are the second, the (+) are the third, and the (x) are the fourth peak maxima respectively.

Summary and Discussion

In this brief paper, I have discussed the phenomenology of stochastic resonance, outlined the theoretical questions, and demonstrated typical measurements on an analog simulator - measurements similar to those which have been performed on other physical systems. I have included all the relevant references as well, so that any interested person dealing with a noisy signal and a bistable detector could easily start his own SR research program. Such a person might wish to begin with a recently published brief but informative note in the popular press[30]. In this summary, I want to emphasize two items:

First, the residence time probability density displays all the same information as the much more widely used and well known power spectrum. Indeed, power spectral analysis is treated in every text book on noisy time series analysis, but when did you last see the probability density discussed? Yet, in some applications P(T) may be more useful, or at least more convenient, than $P(\omega)$.

Second, SR has never been demonstrated as naturally occurring in any system (discounting the ice age speculations). It would be interesting to search for such a phenomenon, for example, where Nature had somehow contrived to make use of the optimum noise intensity for information processing. It seems to me that the place to look is in evolutionary biology, where the necessary ingredients - bistable-like behaviors with D, A and ΔU_0 all of comparable magnitude - are most likely to be found. Certainly, all living organisms are noisy and perform information processing tasks of varying degrees of complexity in the presence of various "signals". It seems to me that after millions of years of evolution, successful organisms would have developed a mechanism, or mechanisms, for making good use of this inescapable and ubiquitous noise[31].

This work was supported by Office of Naval Research grant N00014-90-J-1327, and by NATO grant $\frac{1272}{90}$.

References

[1] R. Benzi, S. Sutera, and A. Vulpiani, "The Mechanism of Stochastic Resonance", J. Phys. A, Vol. 14, pp. L453-456, 1981.

[2] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, "Stochastic Resonance in Climatic Change", Tellus, Vol. 34, pp11-16, 1982.

[3] M. Ghil and S. Childress, Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics, New York: Springer-Verlag, 1987, pp. 335-338.

[4] S. Fauve and F. Heslot, "Stochastic Resonance", Phys. Lett. A, Vol. 97, pp. 5-9, 1983.

[5] G. Vermuri and R. Roy, "Stochastic Resonance in a Bistable Ring Laser", Phys. Rev. A, Vol. 39, pp. 4668-4674, 1989.

[6] B. McNamara and K. Wiesenfeld, "Theory of Stochastic Resonance", Phys. Rev. A, Vol. 39, pp. 4854-4869, 1989

[7] P. Jung and P. Hänggi, "Stochastic Nonlinear Dynamics Modulated by External Forces", Europhys. Lett., Vol. 8, pp. 505-510, 1989.

[8] R. Fox, "Stochastic Resonance in a Double Well", Phys. Rev. A, Vol. 39, 4148-4153, 1989.

[9] C. Presilla, F. Marchesoni and L. Gammaitoni, "Periodically Time-modulated Bistable Systems: Nonstationary Statistical Properties", Phys. Rev. A, Vol. 40, pp. 2105-2113, 1989

[10] P. Jung, "Thermal Activation in Bistable Systems under External Periodic Forces" Z. Phys. B, Vol. 76, pp. 521-535, 1989

[11] L. Gammaitoni, E. Menichella-Saetta, S. Santucci, F. Marchesoni, and C. Presilla, "Periodically Time-modulated Bistable Systems: Stochastic Resonance", Phys. Rev. A, Vol. 40, pp. 2114-2119, 1989.

[12] G. Debnath, T. Zhou and F. Moss, "Remarks on Stochastic Resonance", Phys. Rev. A, Vol. 39, pp. 4323-4326, 1989.

[13] T. Zhou and F. Moss, "Analog Simulations of Stochastic Resonance", Phys. Rev. A, Vol. 41, pp. 4255-4264, 1990

[14] F. Moss, "Stochastic Resonance", Proc. NATO Advanced Research Workshop, **Rate Processes in Dissipative Systems: 50 Years after Kramers**, edited by P. Hänggi and J. Troe, Ber. Bunsen-Gesellschaft Phys. Chem., in press.

[15] P. Jung and P. Hänggi, "Resonantly Driven Brownian Motion: Basic Concepts and Exact Results", Phys. Rev. A, Vol. 41, pp. 2977-2988, 1990.

[16] A. Bulsara, E. Jacobs, T. Zhou, F. Moss and L. Kiss, "Stochastic Resonance in a Single Neuron Model: Theory and Analog Simulation", J. Theor. Biol., in press.

[17] L. Gammaitoni, M. Martinelli, L. Pardi and S. Santucci, "Observation of Stochastic Resonance in Bistable Electron Paramagnetic Resonance Systems", preprint.

[18] H. Risken, The Fokker-Planck Equation: Methods of Solution and Applications, Berlin: Springer Verlag, 1989.

[19] N. G. van Kampen, Stochastic Processes in Physics and Chemistry, Amsterdam: North-Holland, 1981.

[20] R. Lefever and W. Horsthemke, Noise Induced Transitions, Berlin: Springer-Verlag, 1984.

[21] K. Sinha and F. Moss, "Analog Simulation of a Simple System with State Dependent Diffusion", J. Stat. Phys. Vol. 54, pp. 1411-1423, 1989.

[22] H. A. Kramers, Physica, Vol. 7, pp. 284, 1940

[23] P. Hänggi and H. Thomas, "Linear Response and Fluctuation Theorems for Nonstationary Stochastic Processes", Z. Phys. B, Vol. 22, pp. 295-300 (1975).

[24] F. Moss and P. V. E. McClintock, editors, Noise in Nonlinear Dynamical Systems, Vols 1 - 3, Cambridge: Cambridge University Press, 1989.

[25] T. Zhou, F. Moss and P. Jung, "Escape-time Distributions of a Periodically Modulated Bistable System With Noise", Phys. Rev. A, Vol. 42, pp. 3161-3169, 1990.

[26] P. Jung and P. Hänggi, "Escape in Periodic Potentials Perturbed by Periodic Forces", Proc. NATO Advanced Research Workshop, **Rate Processes in Dissipative Systems: 50 Years after Kramers**, edited by P. Hänggi and J. Troe, Ber. Bunsen-Gesellschaft Phys. Chem., in press.

[27] P. Jung and P. Hänggi, "Amplification of Small Signals via Stochastic Resonance", Phys. Rev. A, in press.

[28] M. I. Dykman, R. Mannella, P. V. E. McClintock and N. Stocks, "Phase Shifts and Nonlinear Effects in Stochastic Resonance",

preprint.
[29] A. J. Irwin, S. J. Fraser and R. Kapral, "Stochastically Induced Coherence in Bistable Systems", Phys. Rev. Lett. Vol. 64, pp. 2343-2346, 1990.
[30] Ivars Peterson, "The signal Value of Noise", Science News (Science Service, Inc., Washington, D.C.) Feb. 23, 1991, p. 127.
[31] A. Longtin, A. Bulsara and F. Moss, "Time Interval Sequences in Bistable Systems and the Noise Induced Transmission of Information by Sensory Neurons", preprint.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ABOUT SUM OF SIGNALS AND POSSIBLE USE IN TIME & FREQUENCY SYSTEMS

A. Stern

TFL, Time & Frequency Ltd
P.O.B 1792, Holon 58117, Israel

Abstract

1. The paper analyzes the frequency of the sum of sinusoidal signals. It is shown that if the signals are roughly phase aligned then the frequency of the sum is a weighted average of the frequencies of the signals and the fractional frequency noise is being reduced in the summation.
2. The "Allan deviation" of a signal whose phase is modulated by various wave shapes is analytically and numerically calculated. These include sine-wave, sawtooth-wave and a general periodic function modulation. The calcultions are used to analyze the stability of sum of signals in various cases.
3. We discuss a possible use of signals addition in time & frequency generating systems, where several frequency sources are used for redundancy. One can make optimal use of the frequency sources by simply adding the signals with predetermined amplitudes (weights) in order to achieve a better short-term stability and a redundancy at the same time.

1. Introduction

Frequency addition is very common in many applications, especially in frequency synthesis. It is usually performed with the aid of a non-linear analog element or with a digital element such as XOR. Signals (voltages) addition, on the other hand, is not very usefull since even a small frequency deviation between two added signals give rise to strong phase and amplitude modulation. However, if by some means, the phases of the signals can roughly be kept aligned, then the sum is very nearly a sinusoidal function. In this case a single frequency can be attributed to the sum and some very interesting features are revealed. Section 2 of this paper is devoted to this.

The need to analyze the frequency stability of a sum of signals is the motivation for the Allan variance calculations. These are given in section 4 and include various cases of periodic wave shape phase-modulations. The results of course, are of general interest not only in relation to the main subject of the paper.

The last section of the paper discusses a possible application to signals addition. Whenever several frequency sources are available one can devise schemes to best use all sources in order to improve the short-term frequency stability and at the same time maintain the redundancy and high reliability.

2. Sum of signals (voltage addition)

This section is devoted to an analysis of the sum of two signals in various limiting cases. It is assumed that the signals are in voltage and have the general sinusoidal form:

$$V(t) = V_0 \sin[2\pi\nu_0 t + \Delta\Phi(t)] \tag{1}$$

with the conventional notations: ν_0 is the nominal frequency, $\Delta\Phi(t)$ is the phase deviation from nominal and V_0 is the amplitude. V_0 is assumed to be constant or nearly constant. Usually it is assumed that $\Delta\Phi(t) \ll 2\pi\nu t$ [1], but at present we take $\Delta\Phi(t)$ not to be restricted.

Let $V_1(t)$ and $V_2(t)$ be two sinusoidals signals:

$$V_1(t) = V_{01} \sin[2\pi\nu_0 t + \Delta\Phi_1(t)] \tag{2}$$

$$V_1(t) = V_{02} \sin[2\pi\nu_0 t + \Delta\Phi_2(t)] \tag{3}$$

which are being added to give:

$$V_3(t) \equiv V_1(t) + V_2(t) \tag{4}$$

A phasor diagram is shown in fig. 1

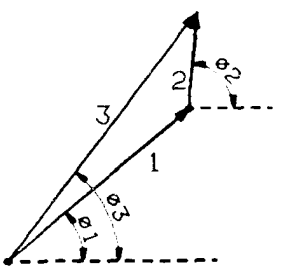


Fig. 1: Signals addition, a phasor diagram.

Both signals has the same nominal frequency ν_0. However, the following calculation applies also for different frequencies since $\Delta\Phi(t)$ has not been limited. Let:

$$V_{01} = 1 \qquad V_{02} = r \tag{5}$$

with no loss of generality.
Then, straightforward trigonomical calculation (fig 1.) gives:

$$V_3(t) = V_{03}(t) \cdot \sin[2\pi\nu_0 t + \Delta\Phi_3(t)] \tag{6}$$

CH2965-2/91/0000-659 $1.00 © 1991 IEEE

where:

$$V_{03}(t) \quad \sqrt{1 + r^2 + 2r\cdot\cos[\Delta\Phi_2(t) - \Delta\Phi(t)]} \tag{7}$$

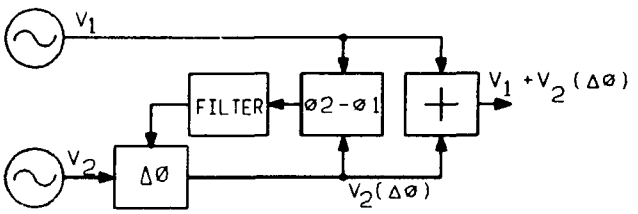
$$\Delta\Phi_3(t) \quad \Delta\Phi_1(t) + \tan^{-1}\left\{\frac{r\cdot\sin[\Delta\Phi_2(t) - \Delta\Phi_1(t)}{1 + r\cdot\cos[\Delta\Phi_2(t) - \Delta\Phi_1(t)]}\right\} \tag{8}$$

As is, the sum $V_3(t)$ which is given by these expressions is difficult to interpret and is of limited use. Hence, we consider the following two limits.

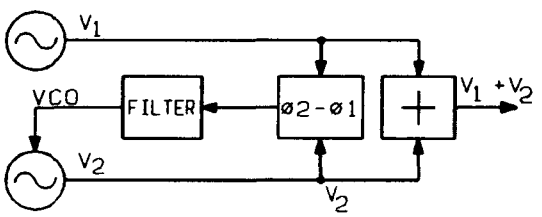
2.1 Phase alignment

In this case it is assumed that the phases of the two signals are very roughly being phased aligned before the addition:

$$|\Delta\Phi_2(t) - \Delta\Phi_1(t)| \sim< \pi/10 \tag{9}$$



a: Phase control



b: Frequency control

Fig. 2: Schematic diagram of phase-lock-loop followed by voltages addition.

Fig. 2a and 2b illustrate schemes of a phase-lock-loop followed by voltage addition. Phase alignment can be achived by discrete steps or in a continuous manner. In the first case the period between the steps is related to the reciprocal of the frequency diffrence between the two signals. For example, two 5MHz signals differing by 1×10^{-11} (in fractional frequency) require a phase correction every 1000 sec.
On the other hand, the loop feed back can be done in a continuous mode where the phase steps are being smoothed out or by continuous frequency control as in Fig. 2b.

Either way a loop time constant τ_L has to be chosen so that the relative phase excursions between the two signals is not greater than $\pi/10$. Then for averaging times longer than τ_L the phases are correlated, while for times shorter than τ_L the phases are not correlated.

Now, the phase alignment condition implies that the amplitude of the sum, $V_{03}(t)$, as given by eq.(7) may vary by no more than ~1%. This is small enough not to disturb the phase measurement significally. Thus $V_{03}(t)$ is considered as constant.

The instantaneous frequency deviation from the nominal is conventionally defined by:

$$\Delta\nu \quad \frac{1}{2\pi}\frac{d\Delta\Phi}{dt} \tag{10}$$

Then, the frequency deviation of the sum $\Delta\nu_3$, is derived from (8) to give:

$$\Delta\nu_3(t) = [1 - A(t)]\cdot\Delta\nu_1(t) + A(t)\cdot\Delta\nu_2(t) \tag{11}$$

where:

$$A(t) = \frac{r\ \{\cos[\Delta\Phi_2(t)-\Delta\Phi_1(t)] + r\}}{1 + r^2 + 2r\cdot\cos\ [\Delta\Phi_2(t)-\Delta\Phi_1(t)]} \tag{12}$$

Again, the phase alignment condition (9), limits the variation of A(t) by ~1%. Hence A(t) may be considered as constant.

Eq.(11) is a key result to this section. It shows that the frequency deviation of the sum of two signals with the same nominal frequency is a weighted average of the frequency deviations of the signals.
The weights are approximately proportional the signals amplitudes:

$$A \approx \frac{r}{1+r} = \frac{V_{02}}{V_{01}+V_{02}} \tag{13}$$

$$1-A \approx \frac{1}{1+r} = \frac{1}{V_{01}+V_{02}} \tag{14}$$

Since $\Delta\Phi(t)$ has not been restricted $\Delta\nu$ can take any value. Hence the signals may have different frequencies ν_1 and ν_2 and a weighted sum may be written also for the frequencies:

$$\nu_3 = (1-A)\cdot\nu_1 + A\cdot\nu_2 \tag{15}$$

Usually the frequency deviation is separeted into an offset term, a drift term and a random, non-deterministic term. The previous results are valid, of course, for each kind of term separately. However, the random-term case is also characterized by a noise reduction as shown in the following section.

2.2 Random frequency fluctuations

Consider the case of random frequency fluctuations. Let $\Delta\nu_f$ in eq.(11) denotes only the random terms. Then one can see that due to interferance a part of the noise is being cancelled in the sum. In the following, the Allan Variance is used to evaluate this noise reduction.

The Allan Variance $\sigma_y^2(\tau)$ is defined by [1,2]:

$$\sigma_y^2(\tau) = < \frac{1}{2}(\bar{y}_{k+1}-\bar{y}_k) > \tag{16}$$

where the fractional frequency y(t) is defined by:

$$y(t) \quad \frac{\Delta\nu(t)}{\nu_0} \tag{17}$$

and $\bar{y}_k$ is an average of y(t) over the k-th time interval of length τ. The angle brackets, $\langle\ \rangle$, denote an infinite time average.

The weighted sum, (11) can now be written in terms of $\bar{y}_k$:

$$y_{k,3} = (1-A)\cdot y_{k,1} + A\cdot y_{k,2} \tag{18}$$

where $y_{k,i}$ is derived from $\Delta\nu_i$, i=1,2,3.

Then:

$$(\bar{y}_{k+1,3}-\bar{y}_{k,3})^2 = (1-A)^2\cdot(\bar{y}_{k+1,1}-\bar{y}_{k,1})^2 + A^2\cdot(\bar{y}_{k+1,2}-\bar{y}_{k,2})^2 + 2(1-A)\cdot A\cdot(\bar{y}_{k+1,1}-\bar{y}_{k,1})(\bar{y}_{k+1,2}-\bar{y}_{k,2}) \tag{19}$$

Taking the time average of (19), it includes terms that vanish if no correlation exists between the two signals:

$$\langle \bar{y}_{k,1}\cdot\bar{y}_{j,2}\rangle = 0 \tag{20}$$

Therefore, a weighted sum is also implied for the Allan variances:

$$\sigma_{y3}^2 = (1-A)^2\sigma_{y1}^2 + A^2\sigma_{y2}^2 \tag{21}$$

An example: let two signals have the same amplitudes and the same noise: r=1, $A\approx 1/2$ and $\sigma_1=\sigma_2=\sigma$.

Then:

$$\sigma_{y3} = \sigma/\sqrt{2} \tag{22}$$

For N signals of the same amplitude and the same noise the noise of the sum is then found to be reduced by $1/\sqrt{N}$.

2.3 The small signal limit

A frequent problem arises when a small amplitude signal interferes with a larger main one. For this case we put:

$$r \ll 1 \tag{23}$$

and analyse the sum $V_3(t)$ (eq.(6)). In this limit, the amplitude, phase deviation and frequency deviation are approximated from expressions (7) and (8) to give:

$$V_{03} \approx V_{01} = 1 \tag{24}$$

$$\Delta\Phi_3(t)=\Delta\Phi_1(t) + r\cdot\sin[\Delta\Phi_2(t)-\Delta\Phi_1(t)] \tag{25}$$

$$\Delta\nu_3 = \Delta\nu_1 + (\Delta\nu_2-\Delta\nu_1)r\cdot\cos[2\pi(\Delta\nu_2-\Delta\nu_1)(t)] \tag{26}$$

i.e. the phase and the frequency are being modulated by the frequency difference, $\Delta\nu_2-\Delta\nu_1$, between the two signals. Hence, the peak frequency deviation (of the sum) is given by:

$$\Delta\nu_0 = r\cdot(\Delta\nu_2-\Delta\nu_1) \tag{27}$$

An example: Let two equal amplitudes signals of 10MHz have a relative frequency offset of 1×10^{-11}. Suppose further that the signals are interisolated by 60dB.

Then, for each signal $r = 10^{-3}$, $\Delta\nu_2 - \Delta\nu_1 = 10^{-4}$Hz and the peak fractional frequency deviation is 1×10^{-14}.

3. Frequency addition

The case of frequency addition is only discussed here in short for comparison. Let two signals be mixed in a non-linear element and filtered to remove the low frequency component. Then:

$$\nu_3 = \nu_1 + \nu_2 \tag{28}$$

$$\Delta\nu_3 = \Delta\nu_1 + \Delta\nu_2 \tag{29}$$

where $\Delta\nu_i$ is the deviation from the nominal ν_i.
Again, if $\Delta\nu_1(t)$ and $\Delta\nu_2(t)$ are not correlated then a partial cancellation of noise occurs. If for example, $\nu_1=\nu_2$ and $\sigma_{y1}=\sigma_{y2}=\sigma$ then one can show that $\sigma_{y3}=\sigma\sqrt{2}$.
However, in contrast to the case of voltages addition, the weights assigned to each frequency are equal and fixed and ν_3 is of course different than ν_1 or ν_2.

4. Allan deviation for periodic phase modulations

In the previous sections we have dealt with cases where a periodic phase modulation occured:
a. In order to achieve phase alignment periodic corrections to the phase of one of the signals are applied. This in turn causes a periodic modulation to the phase of the sum. This modulation can take a form of sawtooth, triange or sine wave.
b. A disturbance of a small signal to a large one is found to result in a sine-wave phase-modulation.

For these cases, but moreover for the general interest, we present calculations of the Allan deviation, $\sigma_y(\tau)$ of a signal whose phase is modulated by various periodic wave shapes.

4.1 Sine-wave modulation

The effect of a sine-wave modulation on $\sigma_y(t)$ has been given in [3]. However, the following calculation reveals more details and points to some ambiguities that might happen in the measurement of $\sigma_y(t)$.

A sine-wave phase-modulation is defined by substituting for $\Delta\Phi(t)$ in eq.(1):

$$\Delta\Phi(t) = \Delta\Phi_0 \sin(2\pi f_m t+\Delta\varphi_m) \tag{30}$$

$\Delta\Phi_0$ is the peak phase deviation (also called the modulation index). f_m is the modulation frequency, $\Delta\varphi_m$ is a modulation phase offset which is later related to the time intervals starting-point. Deriving eq.(30) we find that the frequency deviation is also sine-wave modulated:

$$\Delta\nu(t) = \Delta\nu_0 \cos(2\pi f_m t+\Delta\varphi_m)$$
$$\Delta\nu_0 = \Delta\Phi_0\cdot f_m \tag{31}$$

The fractional frequency error y(t) was defined in (17). The time error x(t) is conventionally defined by:

$$x(t) = \frac{\Delta\Phi(t)}{2\pi\nu_0} \tag{32}$$

Then the Allan variance can be written in terms of x(t) [2]:

$$\sigma_y^2(\tau) = \left\langle \frac{1}{2\tau^2} [x(t+2\tau) - 2x(t+\tau) + x(t)]^2 \right\rangle \tag{33}$$

Substituting eq.(30) and (32) into (33) one finds, as shown in appendix A:

$$\sigma_y(\tau) = \frac{\Delta\Phi_0}{\pi\nu_0\tau} \sin^2(\pi f_m\tau) = \frac{y_0}{\pi f_m\tau} \sin^2(\pi f_m\tau) \tag{34}$$

This result had already been given by [3] and is plotted in fig. 3.

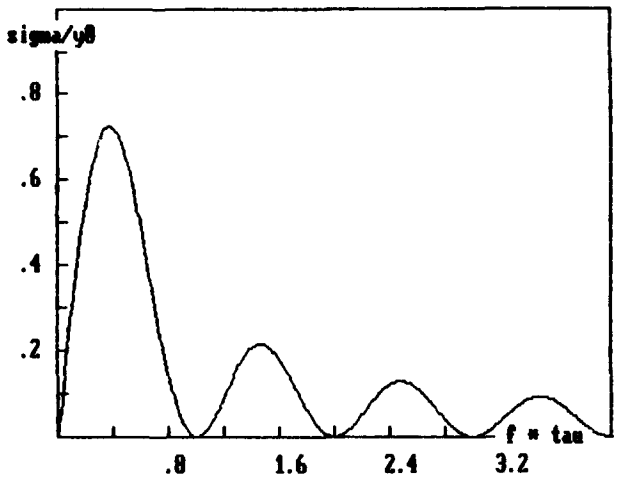


Fig. 3: Allan deviation for sine-wave phase modulation.

However in practice the Allan variance is often evaluated by the formula [2]:

$$\sigma_y^2(\tau) = \frac{1}{2(N-2)\tau^2} \sum_{i=1}^{N-2} (x_{k+2} - 2x_{k+1} + x_k)^2 \tag{35}$$

where:

$$x_k = x(k\tau) \tag{36}$$

Here N discrete reading are used to estimate the variance, and the time axis is divided into intervals of length τ.
In this case, the calculations, which are given in appendix A, result in:

$$\sigma_y^2(\tau) = \left(\frac{\Delta\Phi_0}{\pi\nu_0\tau}\right)^2 \sin^4(\pi f_m\tau) \cdot \left\{1 + \frac{1}{N}\cos[2\pi f_m(N+3)\tau + 2\Delta\varphi_m] \cdot \frac{\sin(2\pi N f_m\tau)}{\sin(2\pi f_m\tau)}\right\} \tag{37}$$

Fig. 4 presents computer simulations of $\sigma_y(\tau)$ based on eq.(37). Apart from a modulation, that appears around $\tau = \frac{2n-1}{2f_m}$ (with n, an integer) the graphs follow approximately the "true" dependence described by eq.(34). The additional modulation, originate from the additional term in the curly brackets in eq.(37) and show a dependence on $\Delta\varphi_m$ and N. $\Delta\varphi_m$ represents the offset in the modulation phase relative to the time interval starting point.

Thus, when estimating $\sigma_y(\tau)$ through eq.(35) one should be carefull with the ambiguous results obtained at $\tau = \frac{2n-1}{2f_m}$

This problem can be largely avoided, however, by using the method of overlapping samples for the calculation of the Allan variance [2]. It is important however that, the smallest spacing of the data, τ_0, should be much smaller then the modulation period, $T_m = 1/f_m$.
Then, it can be shown that the phase dependent term in the curly bracket in eq.(27) gives no contribution (this can still be a problem if strong high harmonics exist).

Eqs.(34) and (26) can now be used to find $\sigma_y(\tau)$ for the small-signal disturbance case:

$$\sigma_y(\tau) = \frac{r}{\pi\nu_0\tau} \sin^2[\pi(\Delta\nu_2 - \Delta\nu_1)\tau] \tag{38}$$

The largest stability degradation, according to (38) occurs at $\tau_{max} = 1.17/(\Delta\nu_2 - \Delta\nu_1)\pi$ where σ_{ymax} is equal to: $0.72 \cdot r \cdot (\Delta\nu_2 - \Delta\nu_1)/\nu_0$.

An example: Let two equal amplitude signals of 10MHz have a relative offset of $1X10^{-11}$ and interisolation of 60dB. Then each signal is sine-wave phase modulated and $\sigma_{ymax} = 0.72x10^{-14}$ and $\tau_{max} = 3200$ sec.

4.2 General periodic phase-modulation

Assume that the phase $\Delta\Phi(t)$ is modulated in a periodic manner. Expand $\Delta\Phi(t)$ by a Fourier series:

$$\Delta\Phi(t) = \sum_{n=1}^{\infty} [A_n \cos(2\pi n f_m t) + B_n \sin(2\pi n f_m t)] \tag{39}$$

where $f_m = 1/T_m$ and T_m is the modulation period. Then the Allan variance can be calculated to give (see appendix B):

$$\sigma_y^2(\tau) = \frac{1}{(\pi\nu_0\tau)^2} \sum_{n=1}^{\infty} (A_n^2 + B_n^2) \sin^4(\pi n f_m\tau) \tag{40}$$

This is, in effect, the discrete form of the well known transformation formula from phase-spectral-density $S_\phi(f)$ to Allan Variance [1,2]:

$$\sigma_y^2(\tau) = \frac{2}{(\pi\nu_0\tau)^2} \int_0^{\infty} S_\phi(f) \sin^4(\pi f\tau) df \tag{41}$$

Actually, both eq.(34) and eq.(40) could have been derived from eq.(41) with the aid of Dirac δ-functions. Eq. (40), however, exhibits explicitely the contributions of the in-phase terms A_n and the quadrature terms B_n.

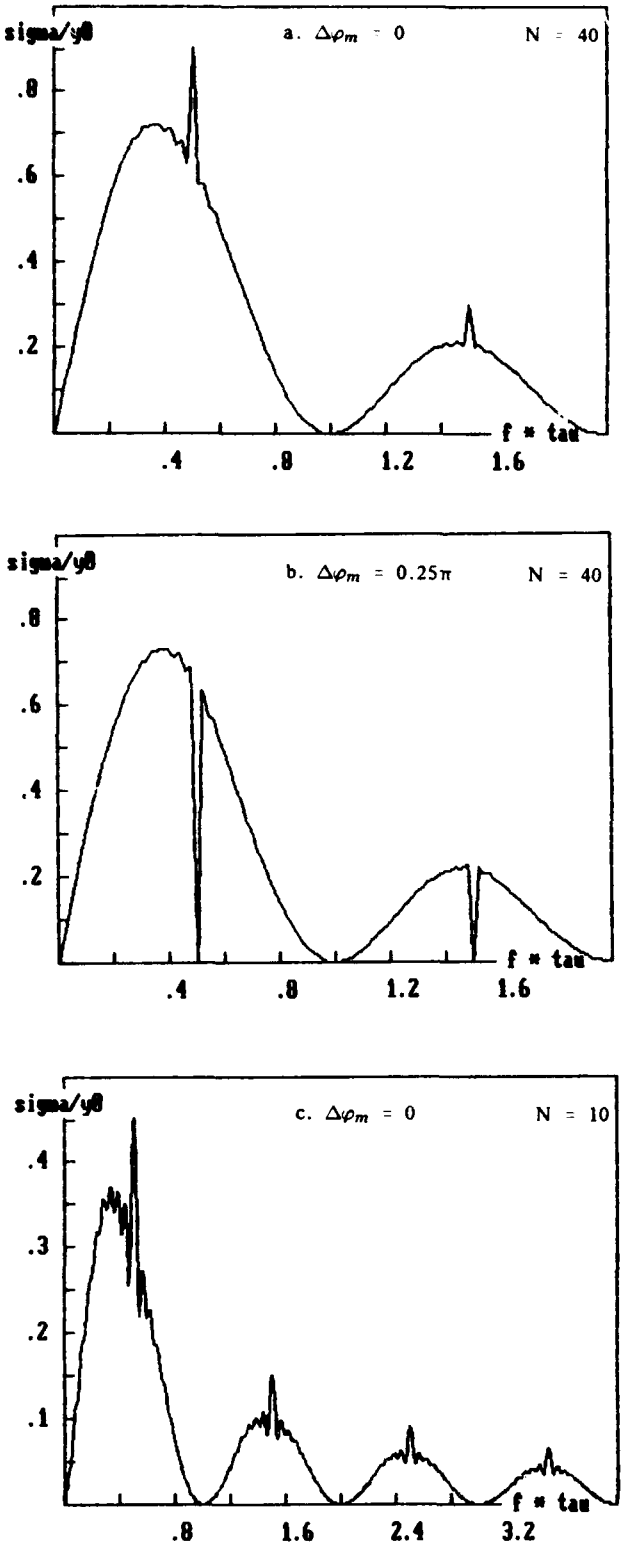


Fig. 4: Allan deviation for sine-wave phase modulation based on the estimation formula (35).

4.3 Triangle and sawtooth phase modulations

The behaviour of the phase and frequency for the cases of periodic triangle and sawtooth phase-modulations is illustrated in fig.5. In fig 6, computer simulations of $\sigma_y(\tau)$ for these modulations are presented. The simulations are based on the estimation formula (35), rather then the Fourier sum, in order to emphasise the irregularities that may arise using the estimation formula.

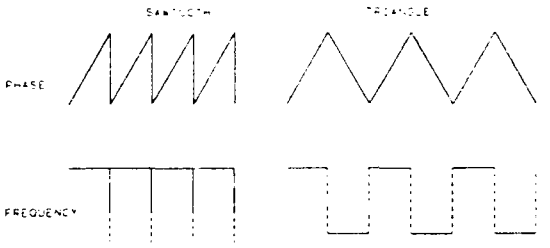


Fig. 5: Periodic triangle and sawtooth phase modulations: phase and frequency behaviour.

Sawtooth phase modulation is a limiting case of triangle modulation where infinitely narrow and infinitely high pulses occurs in the frequency domain. This gives rise to the divergent behaviour of the Allan deviation at $\tau=0$ as ovserved in fig 6c. On the other hand, the triangle-wave picture for $\sigma_y(\tau)$ shown in fig. 6a, is similar to the sine-wave case, shown in fig. 3.

4.4 Considerations of noise reduction

In section 2.2 a noise reduction was shown to occur in the sum of two signals. This however ignors the noise that arises due to the phase alignment corrections. If phase alignment is performed in discrete steps then the phase of the sum is sawtooth-wave modulated. This degrades considerably the short-term-stability around $\tau=0$ as seen in Fig. 6c. Thus, in order to benefit from a noise reduction (which occurs only in the short term) the phase steps should be smoothed-out.

This may lead to the conventional frequency control PLL, depicted in fig. 2b and discussed in section 2.1. However, for primary frequency sources such as Cesium, frequency control is not possible. In the later case we propose a special scheme of phase control which is described in section 5.

5. Applications

Key locations that provide reference frequency or time usually rely on several frequency sources for redundancy. One of the sources is normally being used to realize a frequency output while the others are kept for back-up. In case of three or more sources a "majority vote" can be performed to select the best source. On the other hand, a higher status national laboratories use ensemble of clocks to generate a software time scale [4,5,6]. A software clock provides time and frequency with more reliability, stability and accuracy then any single clock in the ensemble. A hardware clock can be realized by an additional oscillator or by a phase microstepper, both steered by the software.

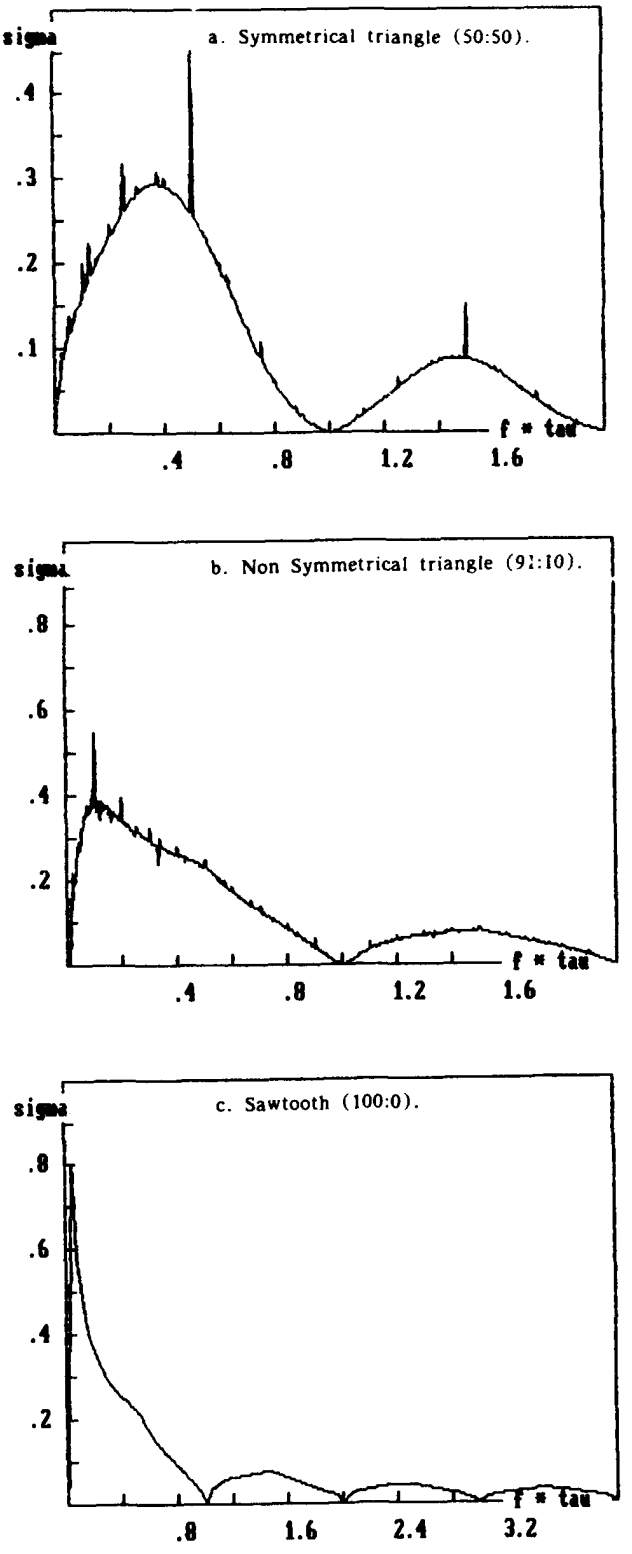


Fig. 6 Allan deviation for triange-wave phase modulation

Signals addition may be implemented to the above applications so as to best utilizes all sources available, in order to attain a better short term stability. in cases where *ovenized or rubidium oscillators are used*, the scheme of frequency steering (fig. 2a) is suggested. On the other hand, for higher hierarchy systems, where Cesium standards are used, a more sophisticated phase steering method is proposed.

In the following a preliminary short description of the later method is given. The general scheme is outlined in Fig. 7.

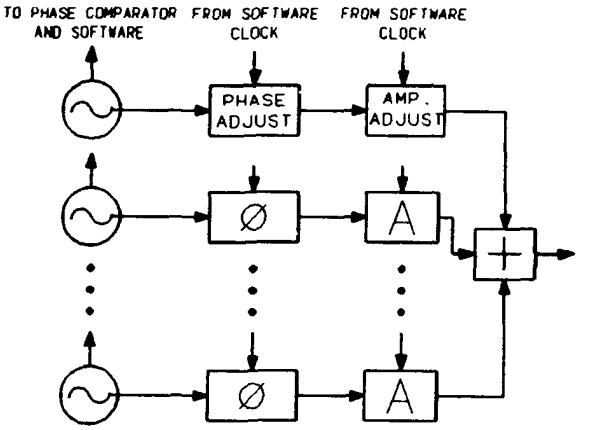


Fig. 7: A proposed scheme for realization of a single low-noise output from several highly accurate oscillators.

The output of several highly accurate sources is splitted into two channels. One channel is used to generate a software clock. In the other channel the signals are phase aligned, amlitude adjusted and summed to produce a signle final output.

Now, the software algorithm is designed to produce two sets of parameters: "long-term parameters" based on the long-term stabilities and "short-term parameters" based on the short-term stabilities. Then, the phase alignment of each signal is made against the long-term software phase so that the relative phase is kept within $\pi/10$.
Phase corrections are performed simultaneously and smoothly to all phases so as to cancel each other and to minimize the phase steps of the sum (relative to the software phase). On the other hand amplitudes adjustments to each signal are made according to the short-term weights assigned by the software.

The method just described has the following advantages:
a. A realization of a hardware clock which follows, in the long-term, the long-term software clock.
b. A reduction of the fractional frequency noise in the short-term. This can be rather substantial if a large number of sources are being used.
c. A high reliability. If a source fails the amplitudes and phases are automatically adjusted to assure a constant amplitude and continuous phase at the output.

References

1. James A. Barnes et al, IEEE Trans on Instrum and Measur, Vol IM-20, No 2, 1971, pp. 105.
2. Samuel R. Stein, Chapter 12, pp. 191 in, "Precision Frequency Control", Vol 2, edited by Eduard A. Gerber & Arthur Ballato, Accademic Press Inc., 1985.
3. David W. Allan, IEEE Transaction on Ultrasonics Ferroelectrics and frequency Control, Vol UFFC-34 No 6, 1987, pp. 647.
4. Fran B. Varnum et al, Proc. of the 9th Annual PTTI, 1987, pp.13.
5. M. Weiss and T. Weissert, Proc. of the 21st Annual PTTI, 1989, pp.343.
6. A. Shenhar et al, 42nd Annual Frequency Control Symposium, 1988, pp. 485.

Appendix A:
Calculation of $\sigma_y(\tau)$ for Sine-Wave Phase Modultion

a. Calculation based on the definition

We use the following definition for $\sigma_y(\tau)$:

$$\sigma_y^2(\tau) = \left\langle \frac{1}{2\tau^2} [x(t+2\tau) - 2x(t+\tau) + x(t)]^2 \right\rangle \tag{A1}$$

Sine-wave phase modulation is defined by:

$$x(t) = x_0 \cos(2\pi f_m t + \Delta\varphi_m) \tag{A2}$$

where a cosine is used for convenience and,

$$x_0 = \Delta\Phi_0 / 2\pi\nu_0 \tag{A3}$$

Then:

$$\frac{1}{x_0}[x(t+2\tau) - 2x(t+\tau) + x(t)] =$$

$$\text{Re exp } \{i\,[2\pi f_m(t+\tau) + \Delta\varphi_m]\}\cdot[\exp(i\,2\pi f_m\tau) - 2 + \exp(-i\,2\pi f_m\tau)] = \cos[2\pi f_m(t+\tau)+\Delta\varphi_m]\cdot[-4\sin^2(\pi f_m\tau)] \tag{A4}$$

Then:

$$\sigma_y^2(\tau) = \frac{8x_0^2}{\tau^2} \left\langle \cos^2 [2\pi f_m(t+\tau) + \Delta\varphi_m] \right\rangle \cdot [-4\sin^4(\pi f_m\tau)] \tag{A5}$$

The infinite time average gives:

$$\left\langle \cos^2[2\pi f_m(t+\tau)+\Delta\varphi_m] \right\rangle = \frac{1}{2} \tag{A6}$$

with no dependence on τ or $\Delta\varphi_m$
Substituting in A5 gives the final result:

$$\sigma_y(\tau) = \frac{2x_0}{\tau}\sin^2(\pi f_m\tau) = \frac{\Delta\Phi_0}{\pi\nu_0\tau}\sin^2(\pi f_m\tau) \tag{A7}$$

b. Calculation based on the estimation formula:

Using the estimation formula (35):

$$\sigma_y^2(\tau) = \frac{1}{2N\tau^2}\sum_{k=1}^{N}(x_{k+2} - 2x_{k+1} + x_k)^2 \tag{A8}$$

Where:

$$x_k \cong x_0\cos(2\pi f_m k\tau + \Delta\varphi_m) \tag{A9}$$

Then calculation like (A4) gives:

$$\frac{1}{x_0}(x_{k+2} - 2x_{k+1} + x_k) = \cos[2\pi f_m(k+1)\tau+\Delta\varphi_m] \cdot [-4\sin^2(\pi f_m\tau)] \tag{A10}$$

Now:

$$\frac{1}{N}\sum_{k=1}^{N}\cos\,[2\pi f_m(k+1)\tau+\Delta\varphi_m]^2$$

$$= \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{2} + \frac{1}{2}\cos\,[4\pi f_m(k+1)\tau + 2\Delta\varphi_m]\right\}$$

$$= \frac{1}{2} + \frac{1}{2N}\,\text{Re exp}\,[i\,8\pi f_m\tau+2\Delta\varphi_m]\sum_{k=1}^{N}\exp\,[i\cdot 4\pi f_m(k-1)\tau]$$

$$= \frac{1}{2} + \frac{1}{2N}\cos\,[2\pi f_m(N+3)\tau+2\Delta\varphi_m]\cdot\frac{\sin(2\pi f_m N\tau)}{\sin(2\pi f_m\tau)} \tag{A11}$$

Then substituting (A10) and (A11) into (A8) we obtain the final result:

$$\sigma_y^2(\tau) = \frac{4x_0^2}{\tau}\sin^4(\pi f_m\tau) \cdot \left\{1 + \frac{1}{N}\cos\,[2\pi f_m(N+3)\tau+2\Delta\varphi_m]\cdot\frac{\sin\,(2\pi N f_m\tau)}{\sin\,(2\pi f_m\tau)}\right\} \tag{A12}$$

Appendix B:
Calculation of $\sigma_y(\tau)$ for general Periodic Phase Modulation

Let $x(t)$ be a periodic function with a period $T_m = 1/f_m$. Expanding in a Fourier series:

$$x(t) = \sum_{n=1}^{\infty}[a_n\cos(2\pi n f_m t) + b_n\sin(2\pi n f_m t)] \tag{B1}$$

Denoting:

$$X_n(t) = \cos(2\pi n f_m t) \tag{B2}$$

$$Y_n(t) = \sin(2\pi n f_m t) \tag{B3}$$

and substituting (B1) into (A1) and rearranging sums*:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2}\left\langle \left\{\sum_{n=1}^{\infty}[a_n\,[X_n(t+2\tau) - 2X_n(t+\tau) + X_n(t)] + [b_n\,[Y_n(t+2\tau) - 2Y_n(t+\tau) + Y_n(t)]]\right\}^2 \right\rangle \tag{B4}$$

As in (A4), we obtain:

$$X_n(t+2\tau) - 2X_n(t+\tau) + X_n(t) = X_n(t+\tau)\,[-4\sin^2(\pi n f_m \tau)]$$
$$Y_n(t+2\tau) - 2Y_n(t+\tau) + Y_n(t) = Y_n(t+\tau)\,[-4\sin^2(\pi n f_m \tau)] \tag{B5}$$

Then expanding the square of the sum and exchanging infinite time average with the infinite sum*, we obtain:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} \sum_n \sum_{n'} \{a_n\, a_{n'} \langle X_n(t+\tau)\, X_{n'}(t+\tau)\rangle \cdot 16\sin^2(\pi n f_m \tau)\sin^2(\pi n' f_m \tau) + \text{terms with } X_n \cdot X_{n'} \text{ and } Y_n \cdot X_{n'} \text{ and } Y_n \cdot Y_{n'}\} \tag{B6}$$

Orthogonality relations gives:

$$\langle X_n(t+\tau)\cdot X_{n'}(t+\tau)\rangle = \frac{1}{2}\delta_{nn'}$$
$$\langle Y_n(t+\tau)\cdot Y_{n'}(t+\tau)\rangle = \frac{1}{2}\delta_{nn'}$$
$$\langle X_n(t+\tau)\cdot Y_{n'}(t+\tau)\rangle = 0 \tag{B7}$$

Hence all crossing terms in (B5) vanish and the result is obtained:

$$\sigma_y^2(\tau) = \frac{8}{2\tau^2} \sum_{n=1}^{\infty} (a_n^2 + b_n^2)\sin^4(\pi n f_m \tau) \tag{B8}$$

Finally to obtain eq.(40) one substitutes:

$$a_n = \frac{A_n}{2\pi\nu_0} \qquad b_n = \frac{B_n}{2\pi\nu_0} \tag{B9}$$

* This is allowed because of absolute convergence.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A FREQUENCY-DOMAIN VIEW OF TIME-DOMAIN CHARACTERIZATION OF CLOCKS AND TIME AND FREQUENCY DISTRIBUTION SYSTEMS

David W. Allan, Marc A. Weiss, and James L. Jespersen

Time & Frequency Division
National Institute of Standards & Technology
Boulder, Colorado 80303

Abstract

An IEEE standard (No. 1139-1988) now exists for "Standard Terminology for Fundamental Frequency and Time Metrology. As defined in this standard, the time-domain stability measure, $\sigma_y(\tau)$ has evolved into a useful means of characterizing a clock's frequency stability. There exists an ambiguity problem with $\sigma_y(\tau)$ for power - law spectral densities, $S_y(f)$, proportional to f^α, where $\alpha \geq +1$. For example, white noise phase modulation (PM) and flicker noise PM appear the same on a $\sigma_y(\tau)$ plot. Because of this ambiguity, Mod$\sigma_y(\tau)$ was developed.

More recently, it has become apparent there is no accepted measure for the performance of time and frequency distribution systems. At the current time, there is an important need for a good method for characterizing time and frequency transfer links in telecommunication networks.

Last year at this symposium suggestions were given for ways to characterize time and frequency distribution systems. Because of the above ambiguity problem, $\sigma_y(\tau)$ was shown to be a less useful measure than Mod$\sigma_y(\tau)$ for such systems. It was shown that $\sigma_x(\tau) = \tau \cdot \text{Mod}\sigma_y(\tau)/\sqrt{3}$ is a useful measure of time stability for distribution systems. For the case of white noise PM, $\sigma_x(\tau)$ is simply equal to the standard deviation for τ equal to the data spacing, τ_0 and is equal to the standard deviation of the mean for τ equal to the data length (T).

In this paper, we recast these measures into the frequency-domain. We treat each of these measures as a digital filter and study their transfer functions. This type of measure is easily related to the passband characteristics of a given system, a particularly useful engineering approach.

Introduction

This paper concerns the characterization of frequency standards, clocks and associated systems. These associated systems may include: time and frequency measurement systems, time and frequency transmissions systems, time and frequency comparison systems, and telecommunication networks. As we shall see no single characterization is suitable. However, in this paper we discuss three statistical characterizations which cover most of the situations encountered in actual practice.

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

The characterizations that we require can be approached from two points of view: the time domain and the frequency domain. In this paper we describe in general terms the time and frequency domain approaches and then explain in some detail how the time domain approaches can be interpreted from the frequency domain point of view. We feel it is important to make this interpretation because of the significance of frequency domain approaches--particularly in engineering environments.

We will next explain why the mean and the standard deviation don't work for frequency standards. The simple mean, the standard deviation, and its square the classical variance, are well known statistical measures of a set of data points. It seems natural therefore that we should apply such measures to characterize various kinds of clock associated processes. However, application of these quickly reveals some significant problems.

Let's begin by considering the computation of the mean frequency output of a frequency standard. Normally, when we compute the mean of some process we suppose that including more data points in the computation brings us ever closer to the true mean of the process and that an infinite number of data points yields the true mean. Of course in the real world we must be content with a finite number of points always leaving some uncertainty in our attempt to find the mean. Nevertheless we assume that we can approach the true mean as nearly as we like if we are willing to collect enough data points.

Is this true of frequency standards? The answer surprisingly enough is "No!". How can this be? The answer is simple enough. Frequency standards do not generate a constant frequency output contaminated only by white noise. If they did we could average the output to get rid of the noise. That is, white noise is the kind of noise that can be averaged away. This is because for every phase advance it produces in the output of the frequency

standard, it eventually produces a compensating phase retardation, so that the two cancel in the averaging process.

Much research has shown that a number of noise processes, in addition to white noise, afflict frequency standards. we can qualitatively say that these other kinds of noise represent trends--not necessarily linear trends--in the output frequency of a standard. In later sections of this paper, we will quantitatively identify these "trends" but for now we will stay with the generic "trend" since this notion is enough to demonstrate why standard statistical methods don't work for frequency standards.

As an example, consider a frequency standard whose frequency output is contaminated with white noise and also increases linearly with time--a very simple kind of trend. What can we say about the mean output frequency of such a standard? Not much, because there is no average; we have defined the standard as producing a monitonically increasing frequency. The frequency we find by averaging is a function of when we start the measurement and the length of time over which the average is made.

What can we do when confronted with such a situation? An obvious answer is to remove the trend, by whatever means, and then compute the mean in the normal way from the modified data. In this way we would expect to converge on the true mean as we average away the white noise by using ever more data points.

The process we have just described is the essence of a number of approaches that have been developed to produce useful statistical measures for the output signals of frequency standards and related devices.[1] We can think of it as a two step process: First we remove the offending trends, and then we compute the statistical measures in the standard way. This process is not always evident when we look at the statistical measures in common use. More often than not, the two steps are combined into one obscuring the underlying process.

We will next describe the frequency- and time-domain measures of frequency and time stability. The standard deviation and the mean are examples of what are called time domain measures. That is, we collect a number of data points, one after the other, and then use these points to construct some useful statistical measure. There are also frequency domain measures. The power spectral density of a set of data points is an example. It provides us with a picture of the deviations in the data having a particular (Fourier frequency) spectral component.

As you might suspect, frequency- and time-domain measures are related. In later sections of this paper we explore in some detail the relationships between time- and frequency-domain measures. However, as a simple example consider the following. Suppose we want to remove a long term trend from the output signal of a frequency standard. There are a number of ways we might proceed. We might, for example, fit a polynomial to the data and subtract this polynomial from the data to generate a new set of data which we could then treat in standard statistical fashion. If the trend were strictly linear, then the polynomial curve would simply be a straight line whose slope revealed the magnitude of the frequency drift.

Another approach is to pass the data through a high pass digital filter to remove the low frequency components--which is what a trend looks like to such a filter.[2]

A particularly simple high pass filter can be constructed by taking what are called "first differences" of the data. That is we subtract the kth data point from the kth + 1 data point, for all data points. This process effectively removes long term trends from the data.

As we shall see later the "first differencing" process corresponds to a digital filter in the frequency domain whose characteristics can be defined precisely. We should also add that the polynomial fitting procedure also corresponds to a particular digital filter. However the emphasis in this paper is on "difference" type procedures since they are easy to implement and are commonly used by the time and frequency community.

We now describe three different variances. One is particularly useful for characterizing the frequency stability of clocks and oscillators. The next is most useful for characterizing the frequency stability of time and frequency measurement systems, distribution and comparison systems as well as for distinguishing between white noise PM and flicker noise PM. The last is most useful for characterizing the time stability of any of the above as well as for network synchronization; e.g. telecommunications network. All three of these variances are built upon taking finite differences of the data.

The notion of taking differences to remove trends in data is an old one. We quote von Neuman et. al. from a 1942 paper [3]:

"There are cases, however, where the standard deviation may be held constant, but the mean varies from one observation to the next. If no correction is made for such variation of the mean, and the standard deviation is computed from the data in the conventional way, then the estimated standard deviation will tend to be larger than the true population value. When the variation in the mean is gradual, so that a trend (which need not be linear) is shifting the mean of the population, a rather simple method of minimizing the effect of the trend on dispersion is to estimate the standard deviation from differences."

Perhaps the most important part of this quote is the parenthetical "which need not be linear." As it turns out taking first differences and second differences--repeating the differencing process twice--is sufficient to remove most of the kinds of noise that one encounters in clocks. Much of the work in the last few decades in the statistical characterization of frequency standards has been directed toward understanding in detail the implications of using the differencing approach. Today this approach is the mainstay of time domain approaches to characterizing time related processes. In this paper we will focus on three such characterizations with their associated interpretations in the frequency domain. We briefly introduce them here with details following in later sections.

The first statistical measure we want to introduce, also historically the oldest, is called the "two-sample variance," the "pair variance" or the "Allan Variance" [1, 4-6] It is denoted $\sigma_y^2(\tau)$ and referred to herein as AVAR. It is defined as follows:

$$\sigma_y^2(\tau) = \frac{1}{2} <(\Delta y)^2>, \qquad (1)$$

where the brackets "< >" denote expectation value, Δ is the first finite difference operator and y is the relative frequency offset as defined below.

AVAR was developed to address the problem of finding a suitable measure of the variability of the output frequency of a frequency standard. As we know, the computation of the standard variance will not work when applied to frequency standards because they contain noise processes which cannot be averaged out. The core idea of the Allan variance has already been introduced--the differencing method. Here the data to be differenced consists of a number of samples of the frequency of the standard taken over some period of time. The differencing procedure filters the noise processes that make the normal variance computation unsuitable. The actual formula, discussed later, accomplishes both the filtering and the variance computation in the same step.

Before we leave AVAR in this introductory section we should point out one thing. The output of a frequency standard is actually a signal whose phase advances in time with respect to some reference. We use phase or time difference almost interchangeably. This is so because they are directly proportional: $x(t) = \phi(t)/2\pi\nu_0$, where $\phi(t)$ is the phase difference reading in radians between two standards. The dimensions of x(t) are time. In practice, the frequency is derived by measuring the time or phase difference x(t) of the signal between the standard in question and the reference at two different times say t and $t+\tau$ giving us phases x(t) and $x(t+\tau)$.

Let $\nu(t)$ be the output frequency of the standard in question, and let ν_0 be the frequency of the reference. We will assume, without loss of generality, that ν_0 is perfect. The average relative frequency offset, $y(t) = (\nu(t)-\nu_0)/\nu_0$ of the standard in question over the time interval t to $t+\tau$ is then

$$y(t) = \frac{(x(t+\tau) - x(t))}{\tau} \qquad (2)$$

If we think of AVAR from the point of view of phase measurements x(t) instead of frequency measurements y(t), then AVAR is constructed in terms of the second differences in phase but first differences in frequency since frequency by definition is obtained from first differences in phase. An alternative and very useful definition of AVAR is as follows:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} <(\Delta^2 x)^2>, \qquad (3)$$

where "Δ^2" is the second finite-difference operator.

The second statistical measure we want to introduce is called the modified Allan variance or from here on "MVAR." [5, 7-9] It is defined as follows:

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} <(\Delta^2 \bar{x})^2>, \qquad (4)$$

where $\bar{x}$ denotes phase averages being used in the second difference. We note that equations (3) and (4) are identical except for the phase averages. The three sequential phase averages are each taken over an interval τ. As τ changes, this changes the bandwidth in the software in just the right way to remove the ambiguity problem in AVAR. In other words, MVAR can distinguish between white noise PM and flicker noise PM, whereas AVAR cannot. In a later section this distinction will be more evident when we compare MVAR and AVAR from the frequency domain point of view.

Both AVAR and MVAR are particularly suited to characterizing the frequency instabilities of frequency standards. However there are situations where the emphasis is not on frequency but on time measurements. This brings up the final measurement we want to introduce, TVAR, where the "T" emphasizes the fact that we are focusing on time rather than frequency measurements. It is defined as follows:

$$\sigma_x^2(\tau) = \frac{1}{6} <(\Delta^2 \bar{x})^2>. \qquad (5)$$

We see that $\sigma_x(\tau)$ is just $\tau \mathrm{Mod}\sigma_y(\tau)/\sqrt{3}$ and has many of the advantages of $\mathrm{Mod}\sigma_y(\tau)$, but is now a time stability measure.

How does the change in emphasis come about? [10-11] The notion of studying frequency instabilities has a local flavor to it in the sense that frequency is defined by a certain resonance frequency of the Cesium atom or quartz

resonator while epoch time is an arbitrary manmade concept requiring coordination over time and space. Thus if we want to compare the frequencies of two remotely located standards we need to introduce some communication link which allows us to compare the phase or time difference between our two clocks. By measuring the change in the time or phase difference between these two standards over time we can determine the frequency offset between the two clocks.

Furthermore we might also want to determine the actual time offset between the clocks which again leads us to making time or phase difference measurements. Both of these examples point up the need for some statistical characterization where the focus is on time or phase rather than frequency, hence TVAR. So there is no confusion, we should point out that the time or phase difference between two clocks is a measurement. Whereas, the finite-difference operators, as in the above variance definitions, operates on a time series of measurement data. The advantages and disadvantages of these three variances will be more apparent later. [11]

We shall also look at the TVAR transfer function from the frequency domain point view This view and some other considerations reveal why TVAR is a more suitable measure for time related measurements than AVAR and MVAR.

Transfer Function Approach to Variances

A variance can be viewed from either the time domain or the frequency domain. [12-14] We intend to look at variances from both perspectives to aid in understanding what certain variances measure. We begin by describing the relationship between the variances in the two domains.

A. Variances in the Time Domain: Convolution

If, in the time domain, we have a time series of observables, $x(t)$, we may define a particular variance with the help of a convolving function, $h(t)$, where $h(t)$ is the impulse response function. We use the convolution, g, of x and h:

$$g(\lambda) = \int x(t) \cdot h(t-\lambda)\, dt. \qquad (6)$$

The variance corresponding to h, of the time series $x(t)$ is the infinite time average of g squared,

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \cdot \int_{\lambda=-T/2}^{T/2} g(\lambda)^2 d\lambda. \qquad (7)$$

The convolving function h, here, is the important definition. It determines how the variance selects data in the time domain, which is then squared and averaged. For example, the convolving function, h, for the classical variance is the step pulse: [2]

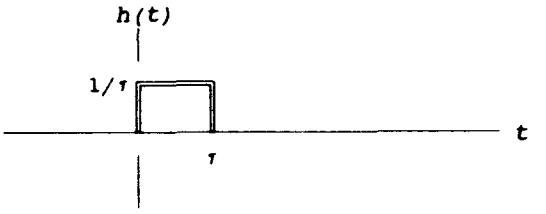


Figure 1. Impulse response function for classical variance. Each sample is taken over with an averaging time τ. Each sample is differenced with the mean, squared and averaged to obtain the classical variance.

This makes

$$g(t) = \overline{x(t)}_\tau, \qquad (8)$$

the average value of x from t to $t+\tau$. Thus the variance here is simply the second moment of $\bar{x}$.

For the Allan Variance (AVAR), h is the double pulse:

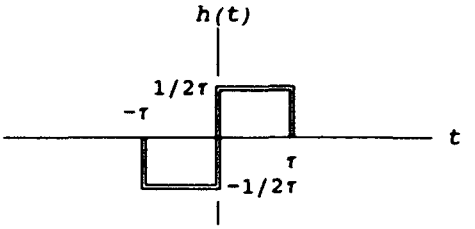


Figure 2. Impulse response function for the Allan or two-sample variance. Adjacent measurements--each averaged over an interval τ--are differenced. This change from one interval τ to the next is squared and averaged across the data, then divided by 2 for an AVAR estimate.

This makes the AVAR the expected value of the first difference squared. The normal use of AVAR is to characterize frequency stability. Thus, if $x(t)$ is a time series of clock time differences, the first difference of these divided by τ, $y(t)$, is the corresponding time series of frequency differences, averaged over the interval τ. Then convolution for the usual AVAR is

$$g(t) = \int y(\lambda) \cdot h_\tau(\lambda - t)\, d\lambda, \qquad (9)$$

giving a first difference of frequencies averaged over a time interval τ, or a second difference of time values. The integral of the square of this g results in the Allan, or two-sample variance

$$\sigma_y^2(\tau) = \lim_{T\to\infty} \frac{1}{T} \cdot \int_{\lambda=-T/2}^{T/2} g(\lambda)^2 d\lambda = \frac{\langle (\bar{y}_t - \overline{y_{t-\tau}})^2 \rangle}{2} \qquad (10)$$

For the Modified Allan Variance (MVAR), we first note that in the Allan variance the time interval for averaging frequency, τ, is a multiple of the basic sampling

interval τ_0. Thus, it is possible to make n shifts of the pulses in figure 2 by τ_0, where $\tau = n \cdot \tau_0$. The modified Allan variance averages several first differences of frequency in this way, thus adjusting the software bandwidth to exploit the bandwidth dependence of white phase noise. As an example we show the convolving function for MVAR as the sum of two functions, as in figure 2, displaced by $\tau/2$.

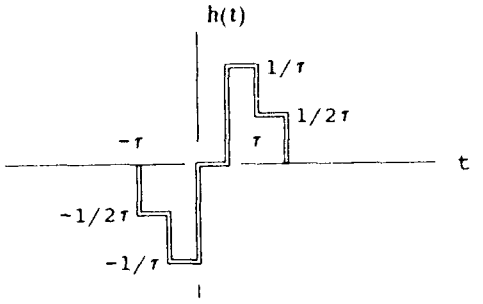


Figure 3. Impulse response function for the modified Allan or two-sample variance. This results from forming a second finite-difference from three contiguous intervals. Each interval contains the phase or time averaged over an interval τ. These second differences are squared and averaged across the data, then divided by $2\tau^2$ for an MVAR estimate.

The new variance, $\sigma_x^2(\tau)$ (TVAR), is simply $\tau^2 \cdot \mathrm{Mod}\sigma_y^2(\tau)/3$. Thus, the convolving function has the same shape as in figure 3, but the vertical scaling needs to be multiplied by $\tau^2/3$.

B. Variances in the Frequency Domain; Transfer Function

We now examine the impulse response functions of these variances as transformed into the frequency domain. We will see that the convolving function h again is the important definition. In this domain, the Fourier transform, H of h, becomes a kind of transfer function for defining the variance.

There are two steps to understanding the passage from the functions we've discussed in the time domain to the frequency domain. First we use the fact that an infinite time average of a function squared equals the integral of the spectrum of that function:

$$\lim_{T\to\infty} \int_{\lambda=-T/2}^{T/2} g(\lambda)^2 d\lambda = \int_0^\infty S_g(f)\,df \qquad (11)$$

Since the variance is an infinite time average of the square of the function g, it also equals the area under the power spectral density of g, the square of the Fourier transform of g.

Second, we use the mathematical relation that the Fourier transform of a convolution is the product of the Fourier transforms. Let us put these two facts together. For the general variance σ^2, defined as the integral of the square of the time series $x(t)$ convolved with $h(t)$, we have

$$\sigma^2 = \int_0^\infty S_x(f) \cdot |H(f)|^2 df, \qquad (12)$$

where $S_x(f)$ is the spectrum of x, and $H(f)$ is the Fourier transform of h.

This $H(f)$ is the transfer function of the variance. This differs from the usual use of the term "transfer function" in that instead of producing a signal sculpted by the shape of H, we produce a variance which is sensitive to frequencies according to the shape of H. [2]

This last equation, then, gives the relationship between the definitions of variance in the time and frequency domains. We see that the convolving function h that defines how the variance selects data in the time domain, also, via its Fourier transform, defines which frequencies the variance is sensitive to.

C. Transfer Functions of AVAR, MVAR, and TVAR

The transfer function for AVAR is shown in figure 4, a linear plot, for two different values of n, where $\tau = n \cdot \tau_0$. We see that the variance selects a band of frequencies for a given τ, and that the width of this band decreases as τ increases. Also note that the function goes to 0 at the origin. Indeed it goes to 0 fast enough that it remains integrable when multiplied by an f^α spectrum with α greater than -3. [6]

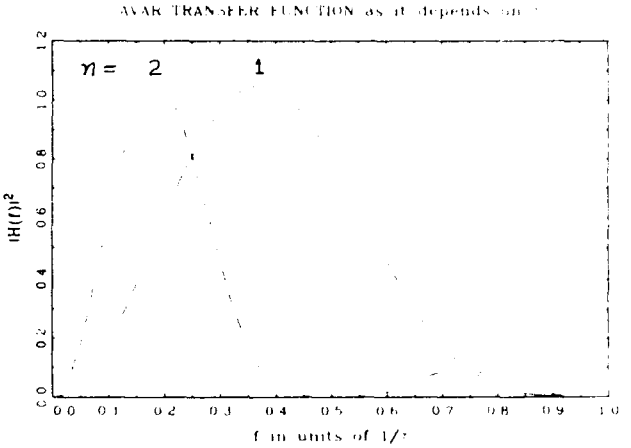


Figure 4. This is a plot of the squared transfer function of the impulse response function shown in figure 2. It is the function that multiplies the spectrum of the frequency deviations to obtain $\sigma_y^2(\tau)$. It is plotted hear for two values of τ (τ_0 and $2\tau_0$ where $1/\tau_0 = 1$). Note, the abscissa is linear and that the bandwidth of H(f) decreases as τ increases.

In figure 5, we see the transfer functions plotted on a logarithmic horizontal axis for n taking on the first eight powers of 2. We see that logarithmically, the bandwidth remains constant, and that the power-of-two transfer functions scan more-or-less independent frequency bands. Figure 6 shows the sum of these transfer functions. We see here that this sum yields a flat band-pass filter. The interpretation here is that this band pass represents the sum of information presented in $\sigma_y(\tau)$ versus τ plot. That is, the Allan variance plot of points chosen with n equal to a range of powers-of-two shows the stability of the data due to a certain band of frequencies. The sensitivity of AVAR to the different frequencies in this band is nearly constant. The band extends from $1/(2n\tau_o)$ to $1/(2\tau_o)$, where n is here the highest power of 2 chosen for the $\sigma_y(\tau)$ plot.

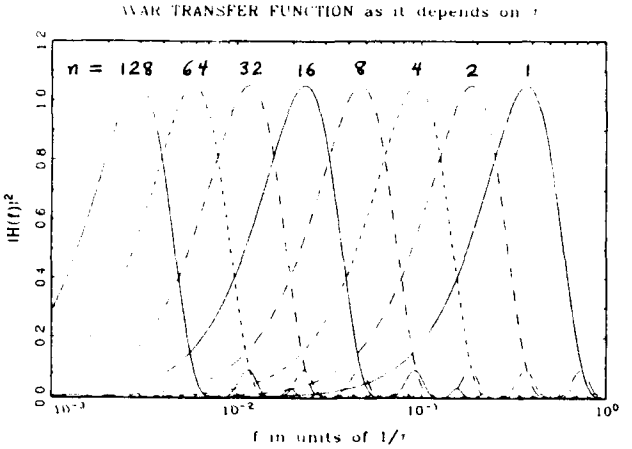


Figure 5. This is a plot of squared transfer functions of the impulse response function shown in figure 2 for 8 values of τ (1, 2, 4, 8....128 x τ_o). Note, the abscissa is logarithmic, and the apparent width of each transfer function is the same. They also appear distributed uniformly across a certain span of Fourier frequencies.

Next we look at the transfer function for MVAR. Analogous to AVAR, we see in figure 7 the MVAR transfer function in a linear plot for two different τ values. Here also, we see that the bandwidth decreases as τ increases, but in addition, the amplitude decreases also. This comes from the additional software filter in MVAR, the phase averaging, allowing MVAR to distinguish white phase noise from flicker phase noise. In figure 8 the MVAR transfer functions for powers-of-two τ values are summed as in figure 6; again, we see a flat band-pass. Thus, an MVAR power-of-two plot also presents the stability information due to a range of frequencies, with nominally equal sensitivity to frequencies within the range. For the same range of τ values, the MVAR cumulative transfer function is a little wider than that of AVAR and the high frequency end is slightly steeper.

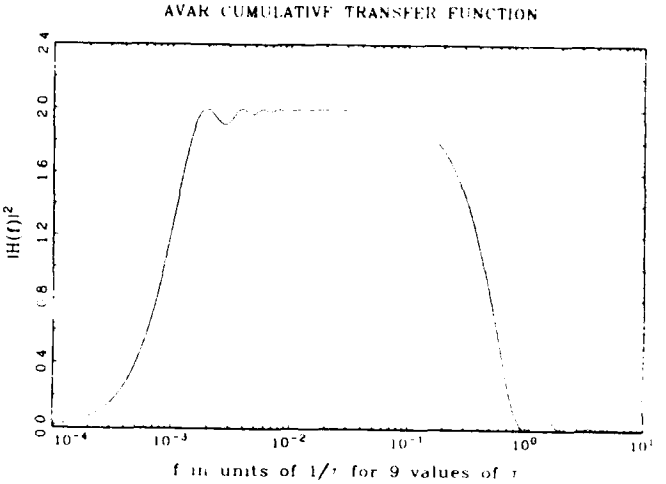


Figure 6. This figure shows the sum of the squared transfer functions for 9 values of τ (1, 2, 4, 8....256 x τ_o) for $\sigma_y^2(\tau)$. We conclude that a $\sigma_y(\tau)$ plot for such a set of τ values gives a nearly constant response to Fourier frequency over about two decades.

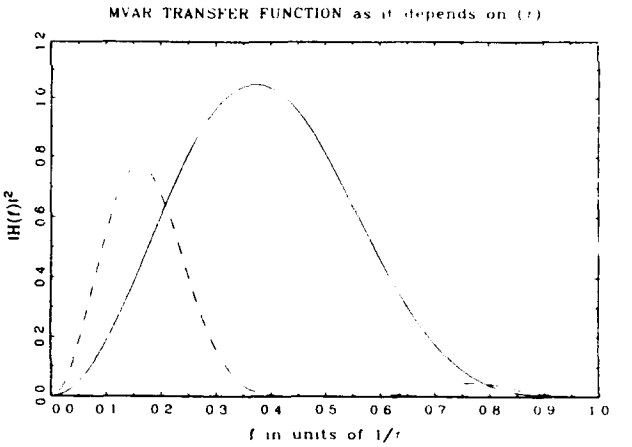


Figure 7. This is a plot of the squared transfer function of the impulse response function shown in figure 3. It is the function that multiplies the spectrum of the frequency deviations to obtain Mod$\sigma_y^2(\tau)$. It is plotted hear for two values of τ (τ_o and $2\tau_o$ where $1/\tau_o = 1$). Note, the abscissa is linear and that the bandwidth of H(f) decreases as τ increases. Notice also, that the amplitude decreases with increasing τ. This is due to the software band-width change brought about by phase averaging.

In figure 9, we see the TVAR transfer function. Note here that the function "rings" forever. That is, neighboring sinusoidal lobes do not die out. Of course with finite data sampling, there is always a high-frequency

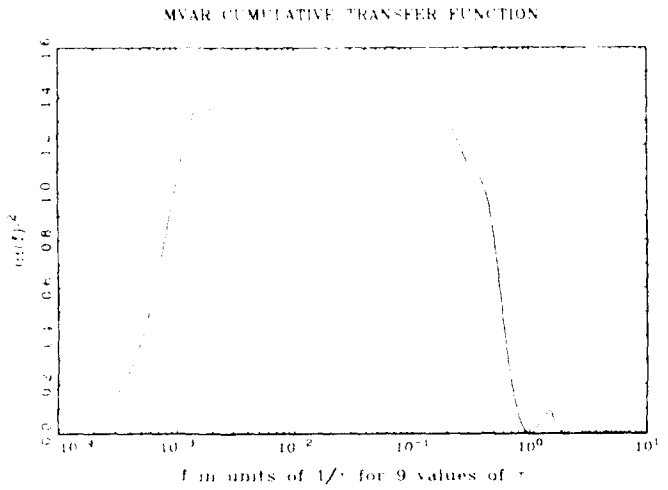


Figure 8. This figure shows the sum of the squared transfer functions for 9 values of τ (1, 2, 4, 8....256 x τ_o) for Mod$\sigma_y^2(\tau)$. We conclude that a Mod$\sigma_y(\tau)$ plot for such a set of τ values gives a nearly constant response to Fourier frequency over slightly more than two decades.

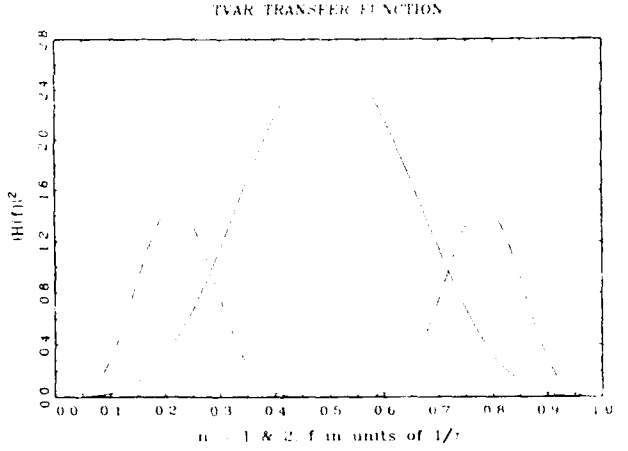


Figure 9. This is a plot of the squared transfer function of the impulse response function for TVAR. It is the function that multiplies the spectrum of the time deviations to obtain $\sigma_x^2(\tau)$. It is plotted hear for two values of τ (τ_o and $2\tau_o$ where $1/\tau_o = 1$). Note, the abscissa is linear and that the bandwidth and the amplitude of H(f) decrease as τ increases--similar to MVAR. As with MVAR, this is due to the software band-width change brought about by phase averaging. This transfer function, for a given τ, has repeat lobes into the higher Fourier frequencies indefinitely.

cut-off given by the Nyquist frequency $1/(2\tau_o)$. The sum of powers-of-two transfer functions, figure 10, shows a fairly flat band-pass up to the high-frequency end where there is greater frequency sensitivity--peaked at the Nyquist frequency, f_{Nyq}. Frequencies higher than the Nyquist frequency can be aliased into a TVAR computation up to the cut-off frequency, f_h. This points out the value of the general rule to have the sampling period equal to or less than $1/(2f_{Nyq})$. Then aliasing will not be a problem.

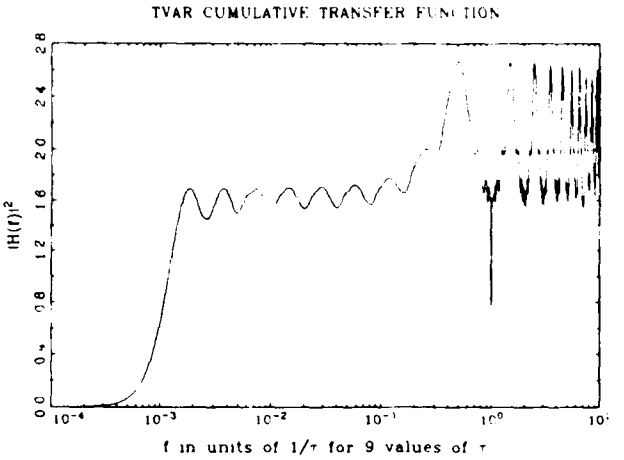


Figure 10. This figure shows the sum of the squared transfer functions for 9 values of τ (1, 2, 4, 8....256 x τ_o) for TVAR. We conclude that a $\sigma_y^2(\tau)$ plot for such a set of τ values gives a nearly constant response to Fourier frequency over slightly two decades but with increased sensitivity to Fourier frequencies at the Nyquist frequency, f_{Nyq}, and at 1/2 f_{Nyq}. Frequencies higher than $2f_{Nyq}$ can be aliased into the computation of TVAR. Frequencies will be aliased up to the measurement system cut-off frequency f_h. Maximum aliasing occurs at 3/2, 5/2, 7/2...x f_{Nyq}. A null occurs at twice the Nyquist frequency. Sensitivities at the non-aliased frequencies above $2f_{Nyq}$ are about the same as they are below the Nyquist frequency.

If the data sampling rate, τ_o, is greater than $1/(2f_{Nyq})$, then a TVAR plot will include Fourier energy from $1/(2n\tau_o)$ up to f_h with about equal sensitivity, except at the aliased values (3/2, 5/2, 7/2... X f_{Nyq} up to f_h). If f_h is equal to f_{Nyq}, then the TVAR cumulative transfer function looks very much like the MVAR cumulative transfer function.

Applications and Discussion

In the previous sections we have learned of three time domain statistical measures that are particularly appropriate for applications involving frequency standards, clocks and their associated measurement, comparison and distribution systems. We have also seen in some detail how these three time domain measures can be interpreted in the frequency domain. In this section we consider where each time domain measure is most appropriately applied. As we shall see, the selection of the appropriate time domain measure is a function of the types of noise which are characteristic of the process we are investigating as well as whether we want to study time stability or the

frequency stability. We want to choose that measure which most clearly reveals the types and levels of noise involved in a particular application.

As we stated in the introduction, AVAR and MVAR were developed first, while TVAR is the newest member of our triad of statistical measures. In general terms, AVAR and MVAR are the measures to use when we are primarily interested in systems and devices where frequency is the quantity of interest, while TVAR is more appropriate where the quantity of interest is primarily time or phase.

Before we begin our discussion of specific applications, let's briefly review the five kinds of noise processes we are likely to encounter for the systems discussed in this paper. Although there are many ways to inventory these noise processes it is common in the time and frequency literature to list them as follows:

1. white noise PM (phase modulation)
2. flicker noise PM
3. white noise FM (frequency modulation)
4. flicker noise FM
5. random walk FM

Mathematically these noise processes have the power-law spectral density relationships shown in the Table 1. Table 2 shows the appropriate mathematical expression for each of the three time domain measures. Table 3 gives the coefficients needed to translate from the time domain to the frequency domain.

Figure 11 displays illustrative examples of the time variation, x(t), of these noise processes. As we proceed from type 1 through type 5 noise we notice that the amplitude variation with time grows increasingly more slowly. Generally speaking, for our applications, the physical explanation for this trend is as follows. The time variations with a f^{-3} and f^{-4} spectrum, for example, are often related to environmental factors such as temperature variations, mechanical shock, and path delay variations while the faster variations represented by f^{0} and f^{-1} processes are more likely related to internal characteristics of the device itself. Here, for example, we think of the noisy electronic components that make up the amplifying stages in a frequency standard.

As we have learned in previous sections AVAR, MVAR and TVAR have different characteristics in the frequency domain so it is not surprising that one time domain measure is better suited for one kind of noise process than another.

AVAR and MVAR are frequency-stability measures, and AVAR is particularly suited for measuring the intermediate to long-term stability of clocks and oscillators. MVAR is generally more suited for electronically

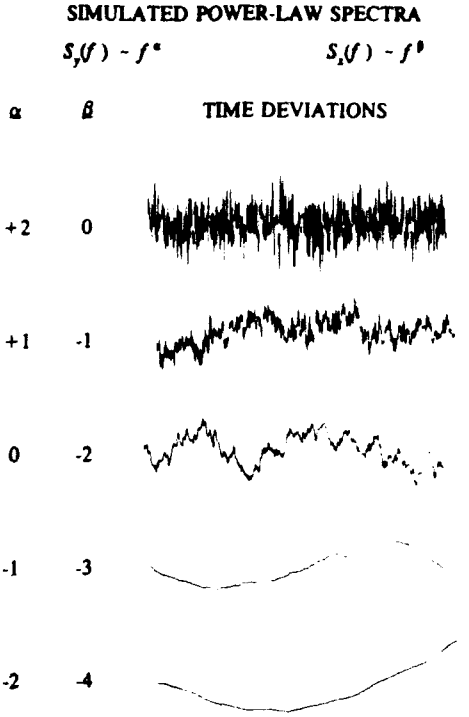


Figure 11. This is a display of the commonly occurring five power-law spectral density processes. These are often used as models for the time and/or frequency deviations in precision frequency sources.

generated noise processes and short-term frequency stability measurements. TVAR is a time-stability measure and is also suited for electronically generated noise processes. Both MVAR and TVAR are also sensitive to low frequency components often determined by environmental factors. TVAR is particularly suited for measuring the stability of time dissemination, comparison or measurement systems. It is also well suited as a measure of synchronization stability in telecommunication's networks.

We can see this from a different perspective by considering how AVAR, MVAR and TVAR vary with τ for our five dominant noise processes. Figures 12, 13 and 14 display the τ dependance for our three time domain measures. If we look at figure 12, we see that AVAR does not discriminate between white PM and flicker PM. This, as we said earlier, was one of the primary reasons for introducing MVAR, which as figure 13 shows, does discriminate between white PM and flicker PM. If we look at figure 14, we see that TVAR displays unambiguously the five noise types, as does MVAR, but that it also more clearly reveals the presence of white PM and flicker PM than does MVAR. This is, of course, the reason that TVAR was introduced since it "focuses" on the noise processes that are of most interest when we are making phase or time measurements.

Table 1. The power-law spectral density relationships for the five kinds of noise processes we are likely to encounter for the systems discussed in this paper.

NOISE TYPE	α	β	μ	μ'	η
White PM	2	0	-2	-3	-1
Flicker PM	1	-1	-2	-2	0
White FM	0	-2	-1	-1	1
Flicker FM	-1	-3	0	0	2
Random Walk FM	-2	-4	1	1	3

Where: $\sigma_y^2(\tau) = a_\mu \tau^\mu$ $\quad S_y(f) = h_\alpha f^\alpha$

$\text{Mod}\ \sigma_y^2(\tau) = b_{\mu'} \tau^{\mu'}$ $\quad S_y(f) = h_\alpha f^\alpha$

$\sigma_x^2(\tau) = c_\eta \tau^\eta$ $\quad S_x(f) = h_\beta f^\beta$

$\sigma_x^2(\tau) = \frac{\tau^2}{3}\ \text{Mod}\ \sigma_y^2(\tau)$ $\quad S_y(f) = (2\pi f)^2 S_x(f)$

Table 2. The appropriate mathematical expression for each of the three time domain measures.

ABBREVIATION	NAME	EXPRESSION
AVAR	ALLAN VARIANCE	$\sigma_y^2(\tau) = \frac{1}{2}\langle(\Delta y)^2\rangle = \frac{1}{2\tau^2}\langle(\Delta^2 x)^2\rangle$
MVAR	MODIFIED ALLAN VARIANCE	$Mod\sigma_y^2(\tau) = \frac{1}{2\tau^2}\langle(\Delta^2 \bar{x})^2\rangle$
TVAR	TIME VARIANCE	$\sigma_x^2(\tau) = \frac{1}{6}\langle(\Delta^2 \bar{x})^2\rangle$

Table 3. The coefficients needed to translate from the time domain to the frequency domain.

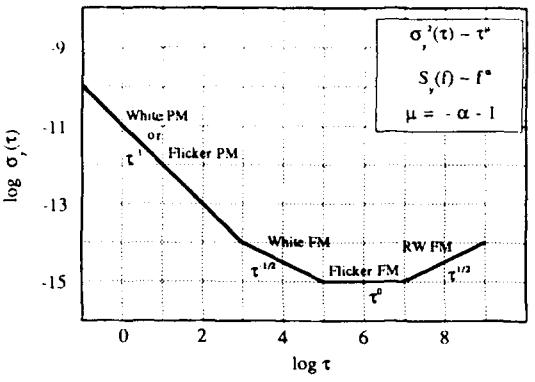
NOISE TYPE	$S_y(f)$	$S_x(f)$
White PM	$\frac{(2\pi)^2}{3fh^2}\ [\tau^2\sigma_0^2(\tau)]\ f^2$	$\frac{1}{\tau_0 f_h}\ [\tau\,\sigma_x^2(\tau)]\ f^0$
Flicker PM	$\frac{(2\pi)^2}{A}\ [\tau^2\sigma_y^2(\tau)]\ f^1$	$\frac{3}{3.37}\ [\tau^0\sigma_x^2(\tau)]\ f^{-1}$
White FM	$2\ [\tau^1\sigma_y^2(\tau)]\ f^0$	$\frac{12}{(2\pi)^2}\ [\tau^{-1}\sigma_x^2(\tau)]\ f^{-2}$
Flicker FM	$\frac{1}{2\ell n2}\ [\tau^0\sigma_y^2(\tau)]\ f^{-1}$	$\frac{20}{(2\pi)^2 9\ell n2}\ [\tau^{-2}\sigma_x^2(\tau)]\ f^{-3}$
Random Walk FM	$\frac{6}{(2\pi)^2}\ [\tau^{-1}\sigma_y^2(\tau)]\ f^{-2}$	$\frac{240}{(2\pi)411}\ [\tau^{-3}\sigma_x^2(\tau)]\ f^{-4}$

$A = 1.038 + 3\ell n(2\pi f_h \tau)$

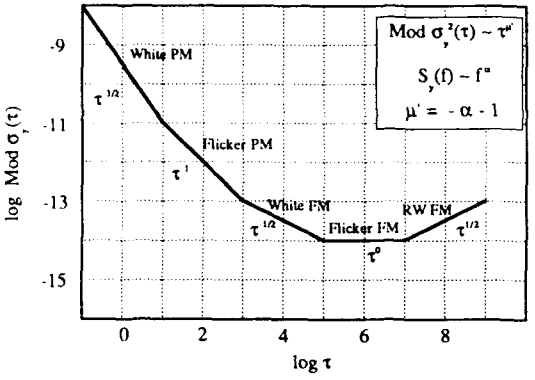
Table 4 shows in some detail the kinds of noise processes that are associated with various areas of application. The table also shows which of the three time domain measures is most appropriate for each particular application. We show the five different types of power-law-spectra and the ranges of applicability. These ranges include those for precision oscillators, for time distribution, network and comparison systems. From this table it is easy to see why MVAR and TVAR are better measures than AVAR for time distribution, network and comparison systems. On the other hand, AVAR estimator of frequency changes for white noise FM processes. This is important for commercial rubidium, cesium and for passive hydrogen masers. AVAR is also simpler to compute and is typically more intuitive than the other two variances. It nicely covers the range of applicability for precision oscillators except for the ambiguity problem in differentiating between white PM and flicker PM. This is only a problem for short-term stability in the case of active hydrogen masers and quartz crystal oscillators. In addition, $k\tau\sigma_y(\tau)$ is unbiased and useful measure of time error of prediction over the interval τ. The constant k depends on the power-law noise type, but is nominally equal to 1. [15]

The five power-law processes for the most part provide adequate modeling for time and frequency metrology. The higher values of alpha typically are used as models for the short-term stability of clocks and oscillators. The lower ends of the ranges are often appropriate models for the long-term stability of clocks and oscillators as well as for the time distribution, comparison, network and measurement systems. These lower values of alpha are often contaminated with diurnal and annual variations in these systems causing them to appear low-frequency dispersive. Some time comparison systems, such as GPS used in the common-view mode, are well modeled by white-noise PM in the day-to-day deviations. The bottom end of the variance ranges are those points where these variances are no longer convergent. If models were needed with lower values of alpha than those shown, then variances with higher order differences could be used, such as the Kolmogorov structure functions. [16] These three measures are convergent for the upper ranges of α and β ($\alpha > +2$ or $\beta > 0$), but only some of the Fourier transform relationships have not been worked out. [2] This is only because these models are not usual.

Sigma Tau Diagram



Modified Allan Variance
distinguishes White PM.



TVAR optimally estimates
time instability with White PM
and distinguishes other noise types.

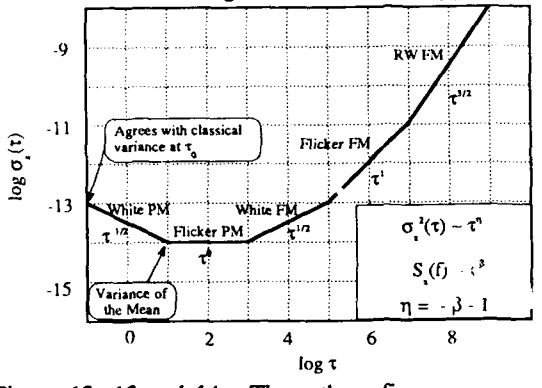
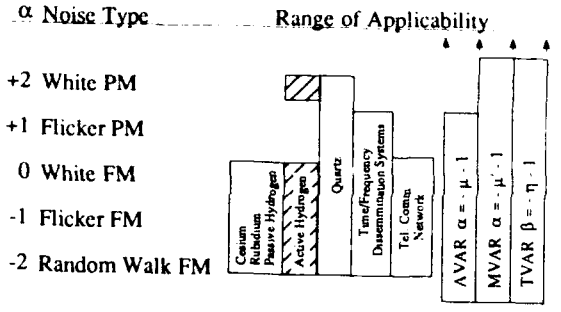


Figure 12, 13 and 14. These three figures are example plots for the square root of AVAR, MVAR, and TVAR, respectively. The sensitivity of these variances to the five power-law spectral density processes illustrated in figure 11 are depicted. Notice, that the slope on the $\sigma_y(\tau)$ plot is the same for white-noise PM as for flicker-noise PM (-1; ζ = -2). Notice also, that it is easier--visually--to distinguish between white phase, flicker phase and random walk phase on a $\sigma_x(\tau)$ plot than from a Mod$\sigma_y(\tau)$ plot. The siope changes are more dramatic to the eye. These three noise processes are particularly useful models for systems where time measurements are important.

Table 4. The kinds of noise processes that are associated with various areas of application. We also see which of the three time domain measures is most appropriate for each particular application.



Conclusion and Summary

Over the last few years, the need for a measure of time stability has become apparent. A search of the literature reveals that the classical measures (standard deviation, mean and variance) have lead to confusing and often misinterpreted conclusions. A measure, TVAR, is shown to have the attributes needed for characterizing the random processes in systems where time stability or phase stability is important. TVAR was compared with and contrasted to the other two previously developed time-domain statistical measures. The need for the three measures is also explained.

In particular, we have discussed how these three measures are appropriate for frequency standards, clocks and their associated measurement and distribution systems. We have shown how these measures can be used to determine the five noise processes that dominate most of the systems of interest in this paper. However, we have also examined these measures from a frequency-domain point of view and have shown how each measure corresponds to a particular transfer function. This procedure reveals the way time domain measures treat the various noise processes from a frequency domain point of view. It also provides a link, for those more accustomed to working in the frequency domain, to the time domain measures which are frequently employed by the time and frequency community.

Acknowledgements

The authors are particularly grateful to the readers, David A. Howe and James A. Barnes, of this paper. We are also very appreciative to the telecommunication industries for their support of the research associated with this paper. We are also grateful for the continued support and encouragement in this project by Donald B. Sullivan.

References

[1] J. Rutman and F. L. Walls, "Characterization of Frequency Stability in Precision Frequency Sources," submitted to Proc. of the IEEE, 1991.

[2] L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Strandards," Proceedings of the IEEE, Vol. 54, No. 2, pp. 136-154, Feb. 1966.

[3] J. von Neuman, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The Mean Square Successive Difference," Ann. Math. Stat. 12, 153-162 (1942).

[4] D. W. Allan, "Statistics of Atomic Frequency Standards," Proc. of the IEEE, 54, 221-231 (1966).

[5] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[6] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M.R. Winkler, Characterization of Frequency Stability, IEEE Transactions on Instrumentation and Measurement, IM-20, No. 2, 105-120, 1971.

[7] D.W. Allan and J.A. Barnes, A Modified "Allan Variance" with Increased Oscillator Characterization Ability, Proceedings of the 35th Annual Frequency Control Symposium,, 470-475, 1981.

[8] P. Lesage and T. Ayi, "Characterization of Frequency Stability: Analysis of the Modified Allan Variance and Properties of Its Estimate," IEEE Trans. on Instrum. & Meas., IM-33, 332-336 (1984).

[9] L. G. Bernier, "Theoretical Analysis of the Modified Allan Variance," Proc. of the 41st Annual Frequency Control Symposium, 1987, pp. 116-121.

[10] D. W. Allan, D. D. Davis, J. Levine, M. A. Weiss, N. Hironaka, and D. Okayama, "New Inexpensive Frequency Calibration Service from NIST," Proc. of the 44th Annual Frequency Control Symposium, 1990, pp. 107-116.

[11] D. W. Allan, "Time and Frequency Metrology: Current Status and Future Considerations," submitted to Proc. of 5th European Frequency and Time Forum, March 13-15, Bescanson, France, 1991.

[12] D. B. Percival, "Reappraisal of Frequency Domain Techniques for Assessing Stability Measures," Proc. of the PTTI, 1987, pp. 69-80.

[13] J. Rutman, "Characterization of Phase and Frequency instabilities in Precision Frequency Sources: Fifteen Years of Progress," Proc. of the IEEE, Vol. 66, No. 9, pp. 1048-1075, September 1978.

[14] J. Rutman, "Characterization of Frequency Stability: A Transfer Function Approach and Its Application to Measurements via Filtering of Phase Noise," IEEE Trans. on I & M, Vol. IM-23, No. 1, March 1974, pp. 40-48.

[15] D.W. Allan, Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 647-654, 1987.

[16] W. C. Lindsey and C. M. Chie, "Theory of Oscillator Instability Based Upon Structure Functions," IEEE 64, 1652-1666 (1976).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

STATUS OF FREQUENCY AND TIMING REFERENCE SIGNAL TRANSMISSION BY FIBER OPTICS*

George F. Lutes and Ronald T. Logan

California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

Abstract

This paper will report recent progress in high stability fiber optic distribution of frequency and timing reference signals. It will give today's state-of-the-art performance at 100 MHz, 1 GHz, and 8.4 GHz for these systems. It will also describe system hardware and discuss cost-performance tradeoffs and future developments.

Introduction

State-of-the-art frequency standards provide accurate and stable frequency and time references needed for location and navigation of distant spacecraft and for precise geographic and geodynamic measurements using Very Long Baseline Interferometry (VLBI), and Connected Element Interferometry (CEI).

The high cost of frequency standards makes it impractical to provide one for each user at a complex. The alternative is to distribute the reference signal generated by a centrally located frequency and timing standard facility to all of the users in a complex.

Over the last two decades engineers have tried, using various technologies and schemes, to develop a means for high stability transmission of frequency references over distances up to a few tens of kilometers[1-4]. For various reasons, none of the technologies were found to be practical until the introduction of fiber optics in the late 1970's.

Since its introduction much progress has been made in the development of high stability analog fiber optic frequency reference distribution systems. Previous papers have reported on past progress in the development of these systems[5-10]. This paper will bring the user community up to date on recent developments.

For comparison, this paper will give the performance of a 1986 fiber optic frequency reference distribution system and the performance of today's state-of-the-art systems. It will also describe a lower cost, lower performance distribution system for the user who does not need full H-maser stability. Finally, the paper will discuss future fiber optic system developments and their potential impact on systems which use high stability frequency reference distribution.

Progress In Fiber Optic Frequency Reference Distribution

Fig. 1 is a block diagram of a typical fiber optic transmitter to be used for high stability applications, such as transmission of reference frequency signals. An electrical signal is applied to the input of the transmitter module where it is added to a constant bias current flowing through the semiconductor laser. The resultant time varying current in the laser diode generates an amplitude modulated (AM) optical signal. This optical signal, emitted by the laser, passes through an integral optical isolator (not shown) which is contained in the transmitter module and then through an optical fiber to an external optical isolator.

From the external isolator the signal passes through a variable optical attenuator which limits the optical output power. This attenuator is needed in short links because the output power of some transmitters may exceed the damage threshold of the receiver. The optical signal having passed through the variable optical attenuator enters another optical fiber and then passes through an optical bulkhead connector and into a fiber optic cable which carries it to the receiver.

* This work represents the results of one phase of research carried out at the Jet Propulsion Laboratory, California institute of Technology, under contract sponsored by the National Aeronautics and Space Administration.

CH2965-2/91/0000-679 $1.00 © 1991 IEEE

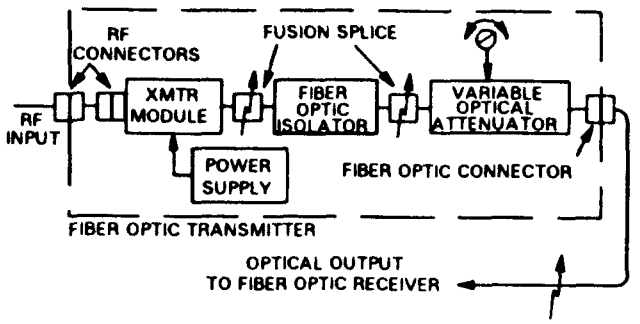


Figure 1. Block diagram of a state-of-the-art analog fiber optic transmitter.

Fig. 2 is a block diagram of a typical fiber optic receiver. The optical signal from the fiber optic transmitter enters the receiver module through a fiber optic bulkhead connector and an optical fiber pigtail. In the receiver module the optical signal is applied to a photodiode detector which converts it back to an electrical signal which is a reproduction of the electrical signal applied to the input of the optical transmitter.

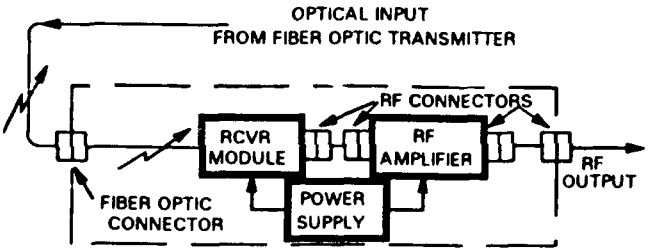


Figure 2. Block diagram of an analog fiber optic receiver.

The electrical signal thus recovered is amplified in an amplifier internal to the receiver module, an external amplifier, or a combination of internal and external amplifiers. In the best optical receivers there is a matching network between the photodiode detector and an internal amplifier. The matching network in effect lowers the equivalent input thermal noise floor of the receiver. The gain of the internal amplifier is chosen such that it will not saturate when large optical signals are applied to the receivers input. For optical signals smaller than the maximum signal an additional external amplifier is needed. The gain of this external amplifier is often adjustable to compensate for various input optical signal levels.

Figure 3 gives the results of measurements of differential frequency stability of a 1986 vintage fiber optic frequency reference distribution link tested over a 14 km distance. The differential stability of this link was 1 X 10^{-15} for 1,000 seconds averaging time, about the same as the stability of a reference signal generated by a H-maser frequency standard[11]. However, the differential frequency stability of the 14 km fiber optic link for 1 second averaging time was about 4 times worse than a H-maser signal for the same averaging time because of inadequate Signal-to-Noise Ratio (SNR) resulting from signal loss over this long distance. When the link length was reduced to 10 meters the SNR improved resulting in improved short term frequency stability about 1.7 times better than a H-maser signal for 1 second averaging time.

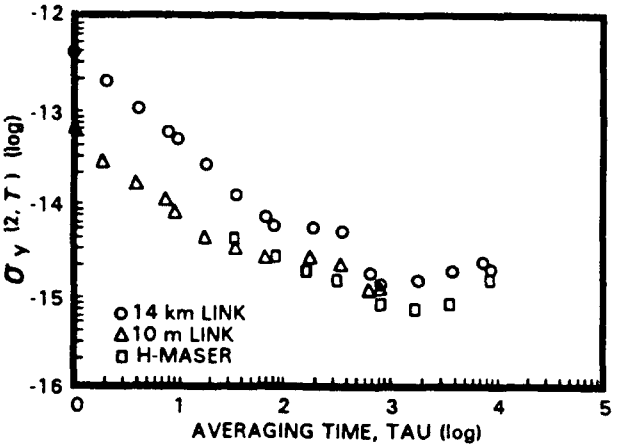


Figure 3. The Allan deviation of a 1986 vintage analog fiber optic link.

The degradation to a H-maser frequency reference signal transmitted over 14 kms with this 1986 vintage fiber optic link was small when the output signal was filtered with a phase locked loop having a 1 Hz bandwidth.

Close-to-carrier (CTC) phase noise was found to be inconsistent in these early analog fiber optic links. This problem was traced to optical reflections in the link from various sources. If reflected light was permitted to enter the laser it resulted in increased (CTC) phase noise as well as increased amplitude noise. Reflections as low as -100 dBmo entering the laser can increase the CTC phase noise.

The abbreviation dBmo refers to optical power level relative to 1 milliwatt, and dBme refers to electrical power level relative to 1 milliwatt. Likewise, dBo is an optical power ratio and dBe is an electrical power ratio.

Prior to 1989 manufacturers of lasers for analog fiber optic systems attempted to minimize noise resulting from reflections by reducing optical reflections from connectors

and other components to less than -60 dBo. Measurements made at JPL proved this to be inadequate for some systems requiring very low CTC phase noise, such as those used for reference frequency distribution. Some manufacturers, responding to the needs of users, began to install integral optical isolators in their laser packages in 1989. Today lasers with integral optical isolators having >30 dBo isolation are the norm for amplitude modulated analog fiber optic systems.

Even though lasers with integral optical isolators improved the laser noise considerably, the 30 dBo isolation they provide is still not adequate for critical applications. An additional external optical isolator having >30 dBo isolation and very low back reflection, <-65 dBo, is needed to achieve the lowest CTC phase noise.

Fig. 4 is a plot of double sideband CTC phase noise versus frequency as a function of reflected optical power into a semiconductor laser transmitter with an integral optical isolator having 30 dBo isolation. The optical power entering the laser itself when a -34.5 dBmo reflection is present is only -64.5 dBmo. The lower line is the measured phase noise when reflections were reduced to a level where further reduction had no effect on the CTC phase noise.

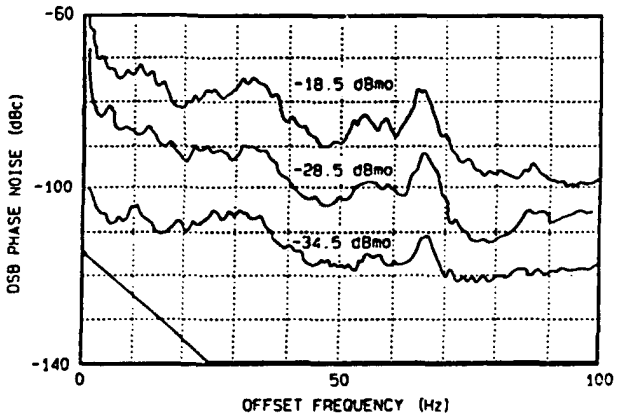


Figure 4. Analog fiber optic link phase noise versus reflected optical power.

Furthermore, it was found that reflections from surfaces internal to the laser package also resulted in increased CTC phase noise. This CTC phase noise is generated by interference fringes which occur in the cavity consisting of the external surface of the laser and the surface of the coupling lens when the laser frequency varies with temperature.

In Fig. 5, a plot of relative group delay in a fiber optic link versus laser diode temperature shows the result of this effect. The cyclical variation is due to reflections

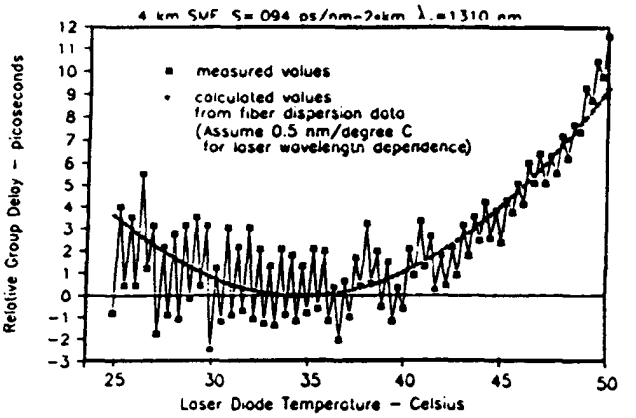


Figure 5. Group delay versus laser diode temperature for a 4 km analog fiber optic link.

internal to the laser package. Based on these findings the suggestions shown in Fig. 6 to reduce CTC phase noise due to these effects were made[12].

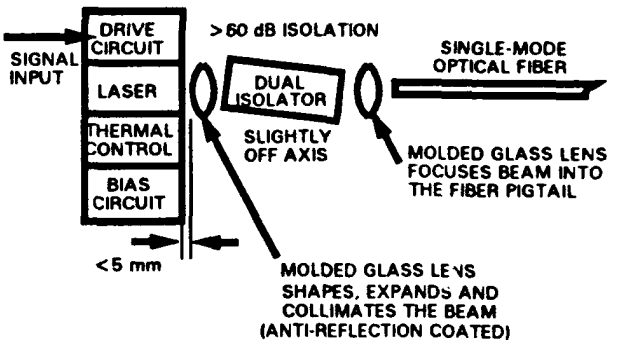


Figure 6. Suggested packaging for an analog fiber optic transmitter.

After this information was reported, Ortel Corporation of Alhambra, California developed a proprietary technique to reduced internal reflections to a very low < -90 dBmo. This substantially reduced CTC phase noise from this source in their transmitters.

In fiber optic systems using narrow linewidth laser sources multiple reflections in the optical fiber also generate increased CTC phase noise. The cause of this noise is interference between the forward signal and reflections propagating in the forward direction. In these

systems optical reflections must be kept to a minimum even if the laser is highly isolated[13].

Ortel Corporation has incorporated low internal reflection and an integral optical isolator into both a low frequency distributed feedback (DFB) laser and a microwave Fabry-Perot semiconductor laser. The DFB laser can be directly modulated up to 1 GHz and the Fabry-Perot laser can be directly modulated at frequencies as high as 12 GHz.

The results of phase noise measurements made at JPL on fiber optic systems using these improved fiber optic transmitters are compared, in Fig. 7, to various frequency sources. The phase noise shown for a fiber optic system using the DFB laser was measured at 100 MHz and normalized to 8.4 GHz. The phase noise shown for a fiber optic system using the Fabry-Perot laser was measured at 8.4 GHz.

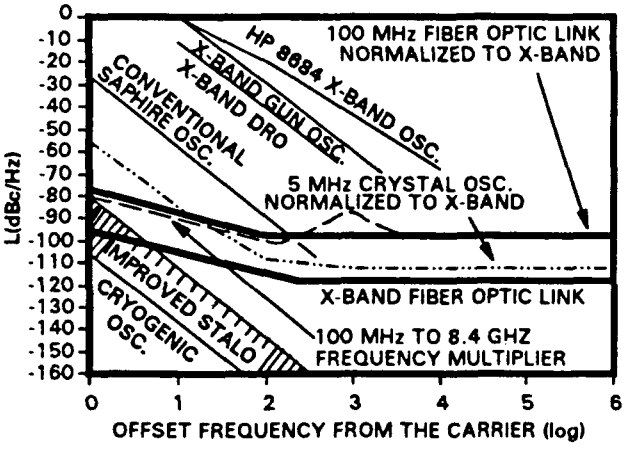


Figure 7. Phase noise of state-of-the-art fiber optic links relative to various frequency sources and a good frequency multiplier.

The Allan deviation for state-of-the-art fiber optic links is shown in Fig. 8. At 100 MHz the measurements were made on a fiber optic link using a DFB laser of the type described above. At 1 GHz and at 8.4 GHz the measurements were made on a fiber optic link using the Fabry-Perot laser described above. The short term Allan deviations shown for these fiber optic systems were calculated from phase noise measurements because the Allan deviations are below the noise floor of existing Allan deviation measurement systems.

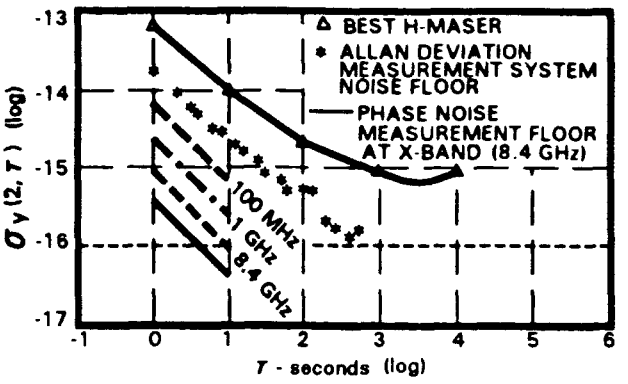


Figure 8. Allan deviation for state-of-the-art fiber optic links at 100 MHz, 1 GHz, and 8.4 GHz.

In most systems, for averaging times longer than a few tens of seconds, frequency instabilities due to thermal effects predominate. Cable delay variations resulting from thermal changes are the major contributor of instability for these averaging times. Low thermal coefficient of delay (0.1 ppm/°C) fiber optic cable, which has recently been developed, reduces cable group delay variations with temperature and greatly improves the long term frequency stability of fiber optic links exposed to large thermal variations[14].

Fiber optic cable stabilizers have been developed which can reduce instabilities below the level achievable with passive means. These stabilizers can virtually eliminate diurnal group delay variations in long analog fiber optic links used to transmit either narrow band reference frequency signals or wide band data signals[15].

A Cost Performance Tradeoff

Table 1 gives a cost breakdown of a state-of-the-art analog fiber optic system with 1 GHz modulation bandwidth which is suitable for applications such as frequency reference distribution. The prices given are small quantity prices and decrease rapidly with volume. The general price trend for this equipment is down. As sales volume picks up, over the next few years, the prices will be reduced drastically.

When these systems are used for frequency reference distribution a phase locked loop filter may be needed to improve short term phase noise. High quality commercial phased locked loop filters are available for about $10,000.

STATE-OF-THE-ART ANALOG FIBER OPTIC LINK		
Name	Quantity	Price
Optical Transmitter Module	1	$12,545
Optical Receiver Module	1	6,495
Optical Isolator	1	1,650
Connectors	4	350
Bulkhead Adapters	2	60
Fiber Optic Cable Organizer Box	2	345
Variable Optical Attenuator	1	432
Enclosure	2	264
Amplifier	1	1,000
Transmitter Power Supply	1	326
Receiver Power Supply	1	140
Miscellaneous Parts		600
Subtotal		$23,107
Phase Locked Loop		$10,000
Grand Total		$33,107
HIGH QUALITY COMMERCIAL ANALOG FIBER OPTIC LINK		
Fiber Optic Transmitter/Receiver Pair		$10,000
Phase Locked Loop		$10,000
Grand Total		$20,000

Table 1. A comparison of the price of a state-of-the-art analog fiber optic system to a commercial fiber optic television transmission system which can be adapted to transmitting 5 MHz frequency references having Cesium stability.

A commercial pulse frequency modulation system sold by Grass Valley Group, Grass Valley, California has been used, with a slight modification, at JPL and in the NASA Deep Space Network (DSN) for simultaneous transmission of a 5 MHz frequency reference and a time code signal[16]. Fig. 9 is a block diagram of this system. Its differential frequency stability (Allan deviation) when used with a clean-up loop is given in Fig. 10. Its phase noise is given in Fig. 11. The cost of this system is $20,000 including the fiber optic link and the phased locked loop filter.

As mentioned previously the optical power of some laser transmitters, when used in a short link, exceeds the maximum input optical power limit of the optical receiver. For links shorter than about 10 km this excess power can be used to reduce the cost of a transmission system by using a single transmitter to send a frequency reference signal to several locations. Such a distribution system is shown in Fig. 12.

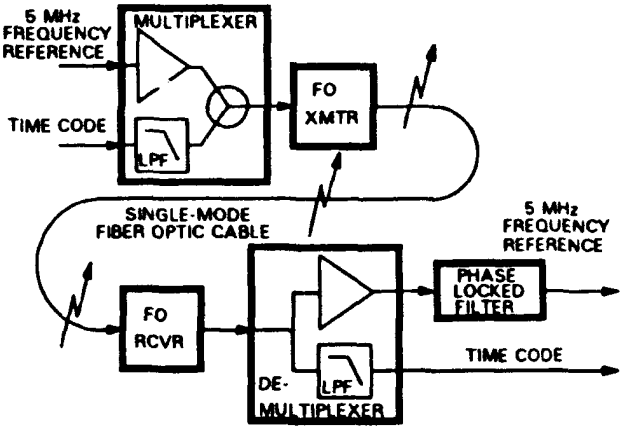


Figure 9. Block diagram of a frequency and timing system which uses a commercial fiber optic transmission system to transmit both a time code and a 5 MHz frequency reference.

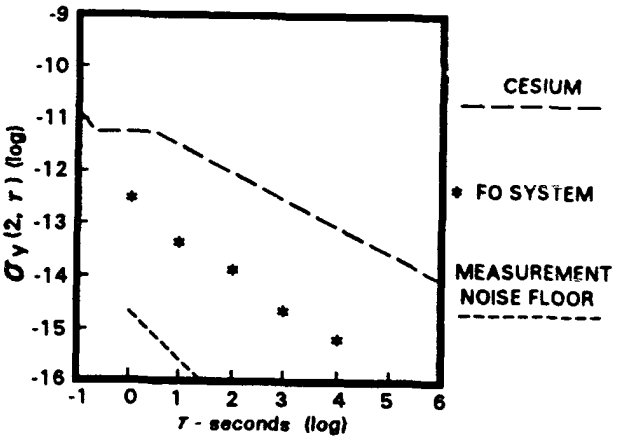


Figure 10. Allan deviation of the system shown in figure 9 compared to the Allan deviation of a Cesium frequency standard.

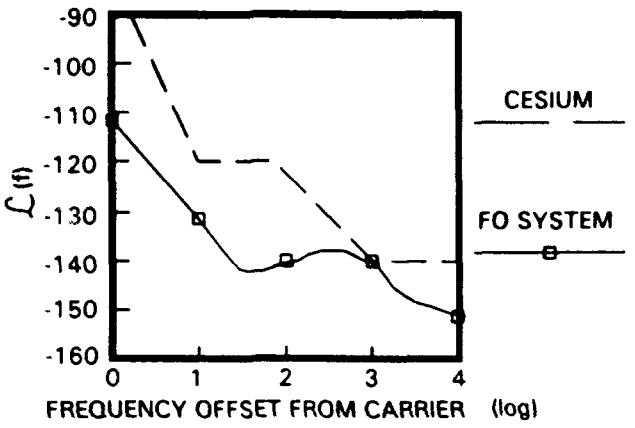


Figure 11. Phase noise of the system shown in figure 9 compared to the phase noise of a Cesium frequency standard.

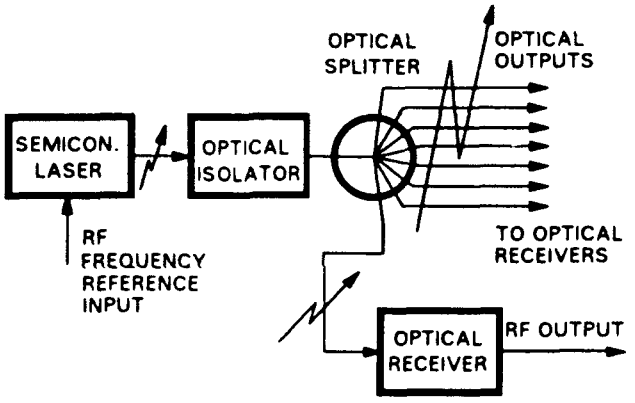


Figure 12. Block diagram showing the use of a single fiber optic transmitter to send a frequency reference to multiple users.

Future Improvements

Semiconductor lasers are predominately used to convert electrical signals to optical signals in today's digital and analog fiber optic transmission systems. The maximum SNR of these systems, about 140 dBe, is usually limited by laser noise. New systems are being developed which use optical transmitters consisting of semiconductor pumped solid state lasers, such as the Nd:YAG laser, with an external electro-optic modulator. These systems will have much higher optical output power and much lower noise.

Fig. 13 shows the optical power, detected RF power, shot noise power density, and thermal noise power density at the input to the receiver versus transmission distance for one of the new systems having +4 dBmo optical output power. Shot noise in this system predominates out to about 25 km. Without considering the effects of optical fiber nonlinearities, the SNR increases linearly with optical power in the shot noise limited region, as shown in Fig. 14.

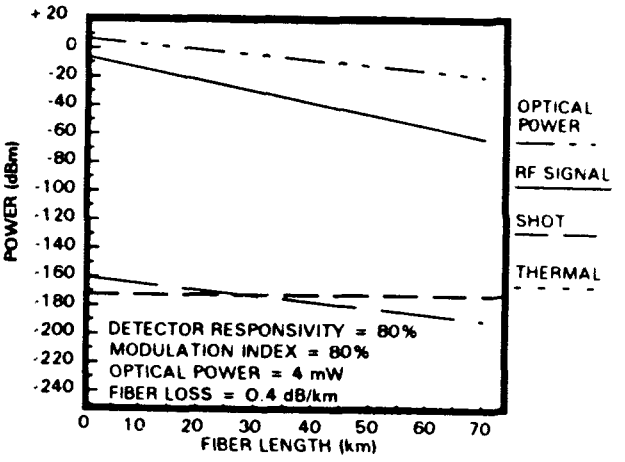


Figure 13. A plot of the predominate noise sources for advanced fiber optic systems now in development.

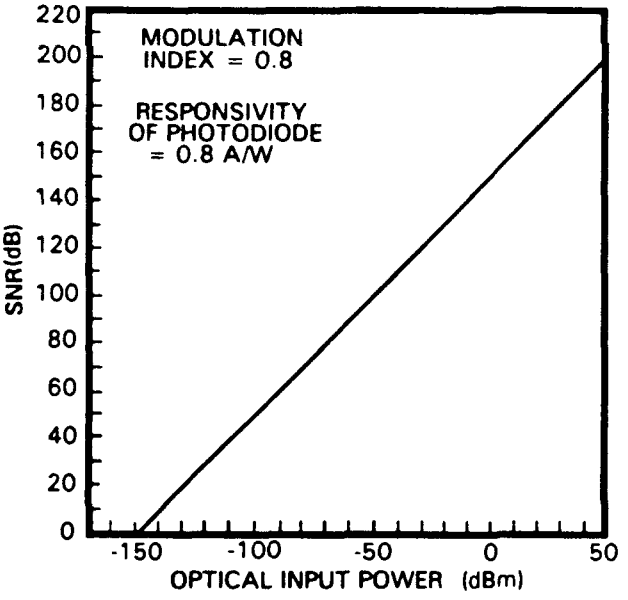


Figure 14. A plot of potential SNR versus optical power for advanced fiber optic links.

The potentially higher SNR of these new systems should result in improved phase noise of analog fiber optic transmission systems to the levels shown in Fig. 15.

Improvements in the 1/f noise region of these systems is uncertain because the mechanisms for this phase noise are not well understood. However, active feedback should give us considerable improvement in this region.

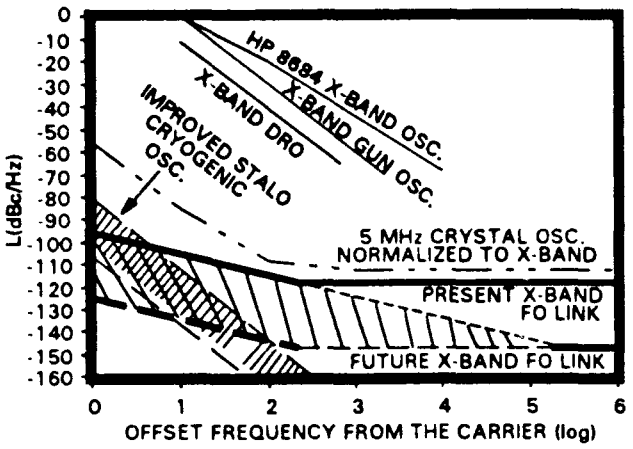


Figure 15. A plot showing the expected phase noise of advanced fiber optic links now in development compared to various frequency sources.

The high stability, dynamic range, and wide bandwidth of these advanced fiber optic systems, which are now in development, could eliminate the need to transmit reference frequencies more than a few feet. These fiber optic systems could transmit the RF and microwave signals directly from the front end of a receiver, for instance, to the vicinity of the frequency and time references for processing. Systems configured in this way will have better stability than those which transmit reference frequencies to a distant location and then send interfrequencies back to a central processing facility for processing.

Conclusion

Recent improvements in analog fiber optic system technology have resulted in substantial reduction of phase noise. The present state-of-the-art phase noise of these systems is -95 dBe, in a 1 Hz bandwidth, 1 Hz from a 8.4 GHz carrier. This is much better than the phase noise of any frequency standard in use today and approaches the phase noise of cryogenic frequency standards being developed[17].

A cost-performance tradeoff has been presented along with a cost savings suggestion to use a single fiber optic transmitter to transmit a signal to several users simultaneously.

The phase noise performance of future fiber optic systems which use a new optical transmitter technology has been predicted to be 25 dB better than today's state-of-the-art fiber optic systems.

Acknowledgements

The authors thank Dr. L. Maleki for his suggestions, inputs and resources in support of this paper. They thank P. Kuhnle for providing test facilities and resources used to collect some of the data used in this paper. And they thank M. Calhoun, L. Primas, and A. Kirk for data they supplied and P. Reder for the assembly of some of the systems which were measured. Finally the authors thank the TDA Technology Development Program Office at JPL and Rome Laboratories for their support of the work described in this paper.

References

1. A. Rogers, "A Receiver Phase and Group Delay Calibrator for Use in Very Long Baseline Interferometry," Haystack Observatory Tech. Note 1975-6.

2. J. W. MacConnell, R. L. Sydnor, "A Microwave Frequency Distribution Technique for Ultrastable Standard Frequencies," The JPL Deep Space Network Progress Report, 42-28, pp. 34-41, Jet Propulsion Laboratory Pasadena, CA, Aug. 15, 1975.

3. G. Lutes, "A Transmission Line Stabilizer," The Deep Space Network Progress Report, 42-51, pp. 67-74, Jet Propulsion Laboratory, Pasadena, CA June 15, 1979.

4. P. A. Clements, "Stable Group Delay Cable," NASA Tech Brief # TSP74-10295, NASA Pasadena Office, 4800 Oak Grove Drive, Pasadena, CA 91109, Year 1975.

5. J. F. Bryant, "Fiber Optic Links for PTTI Dissemination," Proceedings of the Third Annual Department of Defense Precise Time and Time Interval (PTTI) Strategic Planning Meeting, pp. 317-329, U. S. Naval Observatory, Washington, D. C., November 16-18, 1971.

6. K. Y. Lau, "Signal-to-Noise Ratio Calculation For Fiber Optics Links," The Telecommunications and Data Acquisition Progress Report 42-58, pp. 41-48, Jet Propulsion Laboratory, Pasadena, CA, Aug. 15, 1980.

7. G. Lutes, "Optical Fibers for the Distribution of Frequency and Timing References," Proceedings of the Twelfth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 663-680, NASA Conference Publication 2175, Goddard Space Flight Center, Dec. 1980.

8. K. Y. Lau, "Propagation Path Length Variations Due To Bending Of Optical Fibers," The Telecommunications and Data Acquisition Progress Report 42-63, pp. 26-32, Jet Propulsion Laboratory, Pasadena, CA, March, April 1981.

9. L. A. Bergman, S. T. Eng, A. R. Johnston, and G. F. Lutes, "Temperature Dependence of Phase for a Single-mode Fiber Cable," Proceedings of the Third International Conference on Integrated Optics and Optical Fiber Communications, pp. 60, OSA-IEEE, April 27-29, 1981, San Francisco, CA.

10. G. Lutes, "Development of Optical Fiber Frequency and Time Distribution Systems," Proceedings of the Thirteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 243-262, NASA Conference Publication 2220, Naval Research Laboratory, Washington D.C., Dec. 1-3, 1981.

11. G. Lutes, and A. Kirk, "Transmission of Reference Frequencies Over Optical Fiber," Proceedings of the 18th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Proceedings pp. 385-393, Washington D.C., Dec. 1986.

12. R. T. Logan, G. F. Lutes, L. E. Primas, and L. Maleki "Design of a Fiber Optic Transmitter for Microwave Analog Signal Transmission With High Stability," Department of Defense Fiber Optics Conference '90, McLean, Virginia, March 20-23, 1990.

13. R. T. Logan, Jr., Lori E. Primas, G. F. Lutes, L. Maleki,"Modulation Signal Stability Considerations in Analog Fiber Optic Systems," First Annual DARPA/RADC Symposium on Photonics Systems for Antenna Applications, Monterey, California, 13 December 1990.

14. G. Lutes and W. Diener,"Thermal Coefficient of Delay for Various Coaxial and Fiber-Optic Cables,"The Telecommunications and Data Acquisition Progress Report 42-99, pp. 43-59, July-September 1989.

15. L. Primas, G. Lutes, and R. Sydnor, "Stabilized Fiber Optic Frequency Distribution System," Proceedings of the Twentieth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 23-34, Tyson's Corner, VA, Nov. 29 - Dec. 1, 1988.

16. G. Lutes and M. Calhoun,"Simultaneous Transmission of a Frequency Reference and a Time Code Over a Single Optical Fiber," Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Redondo Beach, CA, Dec. 1989.

17. G. John Dick and Jon Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Saphire Dielectric Ring Resonator for Ultra-Low Noise," IEEE Transactions of Ultrasonics, Ferroelectrics, and Frequency Control, vol. 37, No. 5, September 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DETERMINATION OF THE LIMITING FACTORS IN THE ABSOLUTE PHASE NOISE OF AN L-BAND DIELECTRIC RESONATOR OSCILLATOR

M. Mizan, R.C. McGowan, T. Lukaszek, and A. Ballato

US ARMY LABCOM, Electronics Tech & Devices Lab, Ft. Monmouth, NJ

Abstract

After developing and evaluating a state-of-the-art two-stage L-Band Dielectric Resonator Oscillator (DRO), more detailed information about the noise contributions from the individual components needed to be determined. Determination of the noise limiting element in an oscillator is vital to any attempted redesign of the oscillator to reduce its absolute phase noise. In an attempt to understand better the noise performance of our two-stage DRO, a single-stage DRO was constructed for comparison. The results of this investigation are presented in the paper.

Introduction

In any analysis of the predominant noise source in an oscillator, it is important to understand which element is expected to control the noise spectrum in a particular frequency regime and what type of power spectral density is anticipated. Typically, in a well-designed oscillator, at large offset frequencies from the carrier, the noise is limited by the loop amplifier and any of the peripheral circuitry external to the oscillator loop. This region is usually characterized by a f^0 power spectral density. For offset frequencies below this portion of the spectrum, there typically is a region of 1/f power spectral density which is also generated by the loop amplifier and peripheral circuitry and appears as a 10 dB/decade slope in the phase noise plot of the oscillator. At offset frequencies less than the resonator half bandwidth, a $1/f^3$ spectral density is observed. The source of this noise is two-fold. First, the conversion of 1/f and white phase fluctuations in the loop amplifier circuitry will produce a $1/f^3$ spectral density within the resonator half bandwidth. Another possible source of $1/f^3$ noise is short-term 1/f frequency fluctuations of the resonator within its half bandwidth [1]. From this brief discussion it becomes apparent that the performance of the individual components with their $1/f^3$ power spectral density resulting in a 30 dB/decade slope in the half bandwidth of the resonator is critical for the overall performance of an oscillator.

The tool most commonly used for evaluating the noise performance of components in a system is the residual phase noise measurement. This technique allows the investigator to measure the noise added to a signal after it has been processed by any two-port network. Noise added to a signal is the sum of additive and multiplicative noise and has both amplitude and phase modulated components [2]. Using this measurement technique, the actual noise performance of the dielectric resonator and L-Band BJT amplifier can be verified.

Oscillator Design & Temperature Performance

Our discussion will begin with a brief analysis of the design and temperature performance of the one- and two-stage DROs. One can determine from figure 1 that the oscillator is based on a parallel feedback configuration. The amplifier was designed by using S parameters and selecting appropriate input and output matching networks [3]. The resonant cavity structure is shown in figure 2. From the top view of the cavity, one can see that the dielectric resonator is positioned in the center of the cylindrical cavity which corresponds to one-quarter wavelength from the open-circuited end of the 50 ohm microstrip line. This configuration has two advantages. First, by positioning the dielectric resonator one-quarter wavelength away from the open-circuited end of the microstrip line one ensures maximum coupling of the H field into the dielectric resonator. Secondly, by positioning the dielectric resonator in the center of the metallic cavity, the losses due to H field interaction with the walls of the cavity can be minimized.

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

Once the gain of the one- or two-stage amplifier is measured, the height of the low-loss, low dielectric constant spacer can be determined because the height of the spacer is directly proportional to the insertion loss of the dielectric resonator. The height of the dielectric spacer is selected so that a 2-3 dB loop gain margin is maintained to insure operation over a wide range of temperature and biasing conditions. In the case of the resonator for the 2-stage amplifier, figure 3, the insertion loss at the center frequency for the entire resonant structure was 23.5 dB. Given that the gain of the amplifier is 27 dB, there is an excess loop gain of 3.5 dB. Similarly for the single-stage amplifier, the dielectric spacer was chosen to produce an insertion loss of 13 dB in the resonator as seen in figure 4. This provides 2 dB of excess loop gain where the single-stage amplifier has 15 dB of gain.

The loaded quality factor, Q, of the resonator, which is defined as the center frequency divided by the 3 dB bandwidth of the decoupled resonator, can also be measured with the network analyzer. Figure 5 shows the 3 dB bandwidth of the 2 stage resonator as 102 kHz, which corresponds to a loaded Q of 20,500. The results of this two-port measurement for the single-stage resonator shown in figure 6 yielded a 3 dB bandwidth of 165 kHz, which corresponds to a loaded Q of 12,500.

The temperature performance of the two oscillators was examined for completeness. Figures 7 and 8 display the center frequency stability as a function of temperature for the two- and one-stage DROs, respectively. Although very similar, the frequency stability of the single-stage DRO was slightly better with a 100 ppm change over the 100K temperature differential from +55°C to -45°C compared to the 130 ppm change for the two-stage oscillator over the same temperature range. The RF power output response for the two DROs was very different. The two-stage DRO had a constant power output with less than a 1 dBm change in output power over the 100K temperature differential from +55°C to -45°C as seen in figure 9. While the single-stage DRO varied 10 dBm over the same temperature range as seen in figure 10.

Residual and Absolute Phase Noise Performance

After measuring the absolute phase noise of the two-stage oscillator, the question was posed: could the oscillator be improved? This question can be resolved by performing residual phase noise measurements on the 2-stage amplifier and dielectric resonator. However, the data obtained on the two-stage L-Band bipolar junction transistor (BJT) amplifier and the 2 GHz dielectric resonator were inconclusive because both components gave the system floor as their residual phase noise. Therefore, it was decided to build and evaluate a single-stage L-Band DRO based on the two-stage design to determine if the amplifier was the limiting factor in the two-stage DRO and provide more insight into the overall performance of the oscillator.

One of the primary problems in making residual phase noise measurements is inadequate dynamic range of the system. Enhancement of the system's dynamic range or equivalently the system noise floor was achieved by building a half-watt reflection-type DRO. The block diagram of the residual phase noise test setup is shown in figure 11. The noise floor acquired at 2 GHz using the reflection-type 0.5 watt DRO is given in figure 12. Using this high power oscillator as the frequency source in the test setup, the residual noise of the single-stage BJT amplifier was measured and the results of this test are shown in figure 13. Again the amplifier noise measurement is inconclusive because the data is essentially the same as the system noise floor.

When the 0.5 watt reflection-type DRO was used to make the residual noise measurement on the dielectric resonator it was determined that the frequency instability of the source made it impossible to perform the measurement on such a high Q, narrow bandwidth structure. Therefore, the reflection-type oscillator was replaced with an HP 8662A frequency synthesizer and a doubler to provide a stable source for the resonator. The system floor using this source is shown in figure 14. It should be noted that the system floor is only shown out to 10 kHz because after this point the source noise begins to decorrelate as the offset carrier frequency nears the 3 dB bandwidth of the resonator, making the data invalid. Figure 15 shows the residual noise measurement of the dielectric resonator. Again we see that the resonator's residual phase noise is essentially the same as the system noise floor; therefore the data are inconclusive.

The only conclusive data we obtained that indicated that the amplifier might be the primary noise contributor came from the absolute phase noise data. The absolute phase noise measurement test set up seen in figure 16 was used

for both the 1- and 2-stage DROs. The measurement was made by locking the 155 MHz difference frequency signal, derived by downconverting the test DRO with a low noise High-overtone Bulk Acoustic Resonator (HBAR) oscillator, to a Hewlett Packard (HP) 8662A frequency synthesizer driven by an external 10 MHz VCXO. Figure 17 shows the absolute phase noise data for the two-stage DRO and Figure 18 shows the data for the single-stage DRO. In a direct comparison the higher Q two-stage DRO is approximately 5 to 7 dBc/Hz better than the lower Q single-stage DRO.

Conclusions and Future Work

First, it should be noted that the measured data clearly demonstrate the low phase noise capability and excellent frequency stability of the two-stage, 2 GHz DRO. Residual and absolute phase noise measurements were performed in an attempt to determine the noise limiting element of this state-of-the-art 2 GHz DRO [4]. The residual phase noise measurements for the components of the single- and two-stage DROs were inconclusive because they were virtually indistinguishable from the system noise floor. A comparison of the absolute phase noise performance of the two oscillators did provide some insight into the noise performance of the two-stage oscillator. The 5-7 dBc/Hz improvement in the absolute phase noise of the two-stage DRO is in agreement with Leeson's model for an oscillator whose noise is limited by the amplifier. Leeson's model predicts a 6 dBc/Hz improvement in the absolute phase noise of an oscillator if the oscillator is amplifier limited and the loaded Q is increased by a factor of two. This result agrees with our measured data. Our data predict approximately a 4.2 dBc/Hz improvement in the absolute phase noise of the two-stage oscillator for its 1.63 Q enhancement over the single-stage DRO. The source of the 1-3 dBc/Hz difference between the calculated and the measured data is being investigated.

In the near future we hope to build a high power BJT amplifier to boost the signal output from our stable 2 GHz DRO, so that we can improve the system floor of the residual phase noise measurement test set. By doing this, we hope conclusively to determine the noise contribution of the amplifier and the entire resonant structure.

Acknowledgment

The authors would like to thank Elio Mariani for the technical assistance and encouragement that he offered throughout our work. We would also like to thank John Ondria for the guidance he provided with our phase noise measurements while he was at Fort Monmouth.

REFERENCES

[1] G.S. Curtis,"The Relationship between Resonator & Oscillator Noise and Resonator Noise Measurement Techniques", Proceedings of the 41st Annual Symposium on Frequency Control 1987, pp. 420-428, IEEE Catalog No. 87ch2427-3.

[2] T.R. Faulkner, "Residual Phase Noise Measurement", Microwave Journal, 1989 State of the Art Reference, pp. 135-143.

[3] Guillermo Gonzalez, Microwave Transistor Amplifier Analysis and Design, Prentice Hall, 1984.

[4] M.J. Loboda, T.E. Parker, G.K. Montress, "Frequency Stability of L-Band, Two-Port Dielectric Resonator Oscillator", IEEE Transactions on MTT, Vol. MTT35, No. 12, December 1987, pp. 1334-1339.

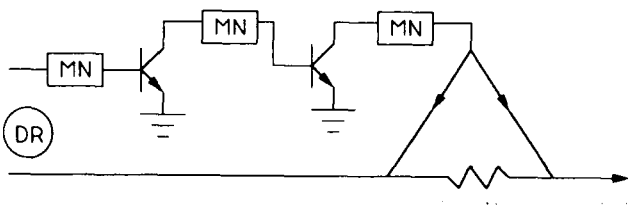
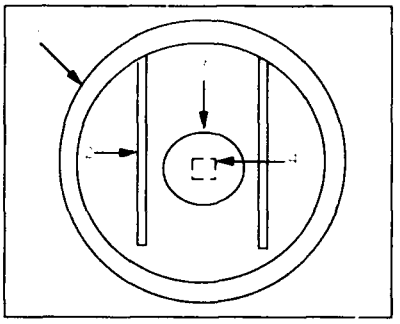
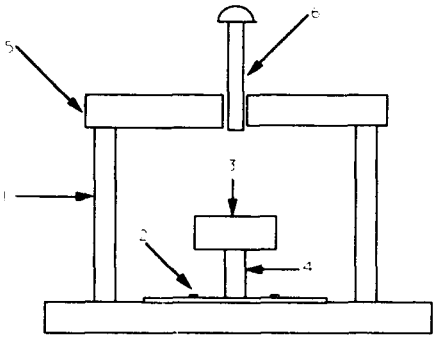


Figure 1 2 GHz Dielectric Resonator Oscillator Block Diagram



Top View, Lid Removed



Cross-Sectional View

1 Metal Cylinder
2 50-ohm Microstrip Line
3 Dielectric Resonator
4 Low-loss, Dielectric Spacer
5 Metal Lid
6 Metal Tuning Screw

Figure 2 Structure of Resonant Cavity

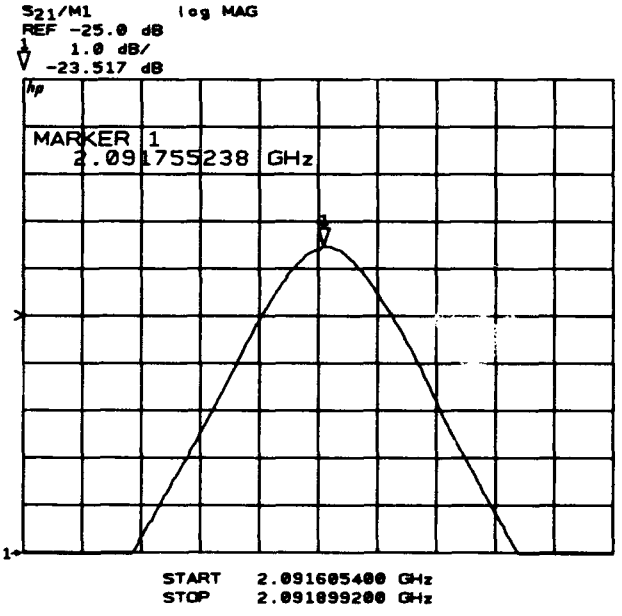


Figure 3 Center Frequency and Insertion Loss Data for Resonator Used with Two-Stage Amplifier

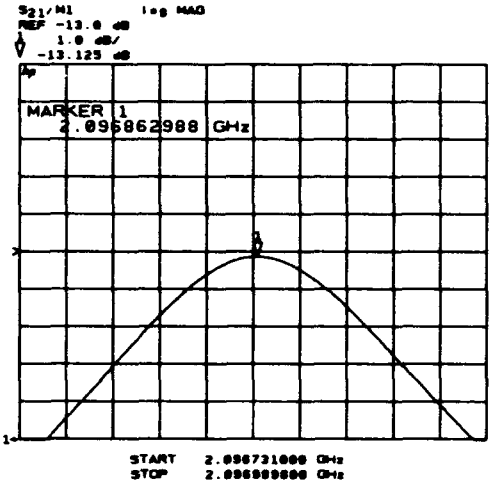


Figure 4 Center Frequency and Insertion Loss Data for Resonator Used with Single-Stage Amplifier

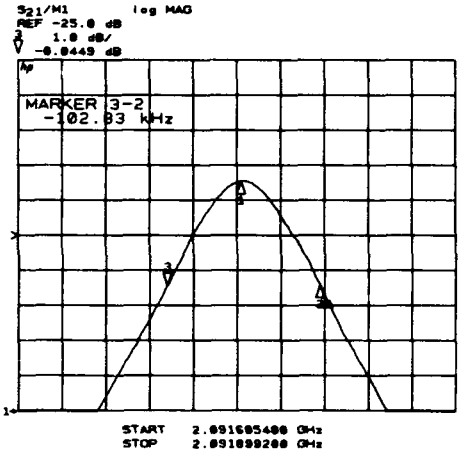


Figure 5 Loaded Q Data for Resonator Used with Two-Stage Amplifier

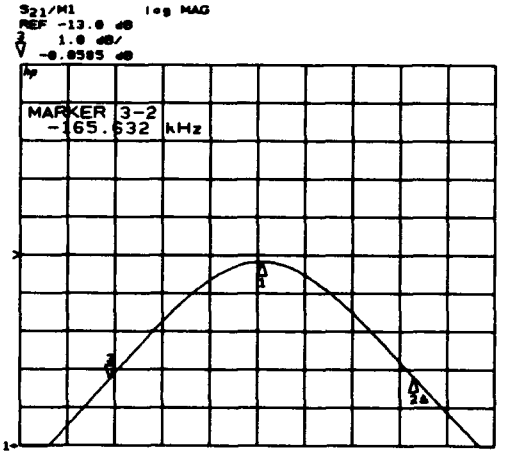


Figure 6 Loaded Q Data for Resonator Used with Single-Stage Amplifier

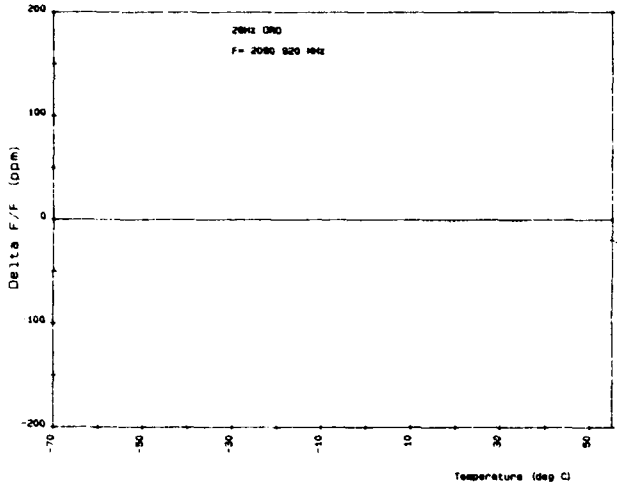


Figure 7 Center Frequency vs Temperature for Two-Stage DRO

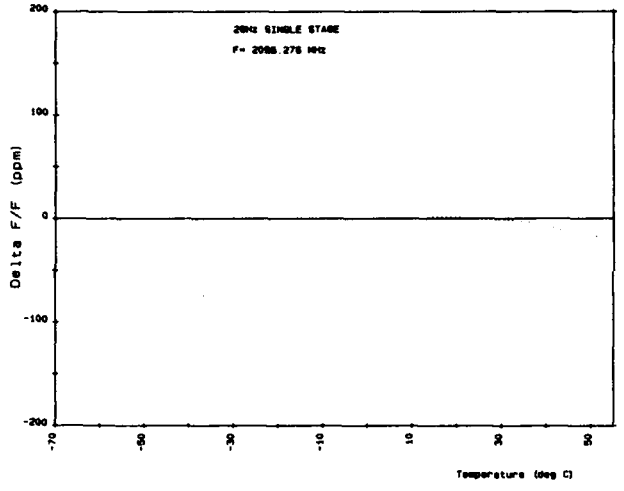


Figure 8 Center Frequency vs Temperature for Single-Stage DRO

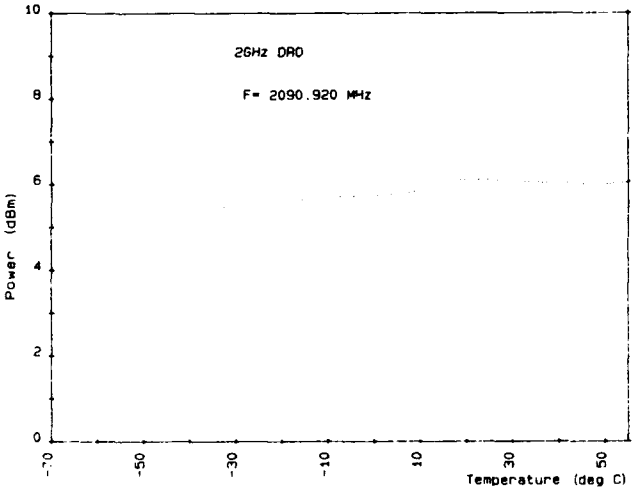


Figure 9 Power Output vs Temperature for Two-Stage DRO

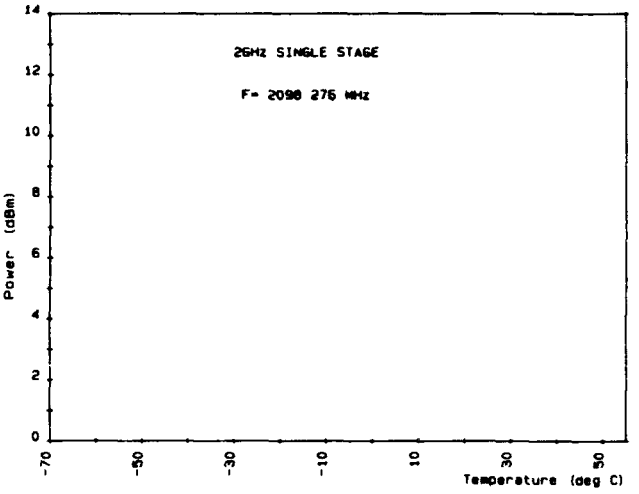


Figure 10 Power Output vs Temperature for Single-Stage DRO

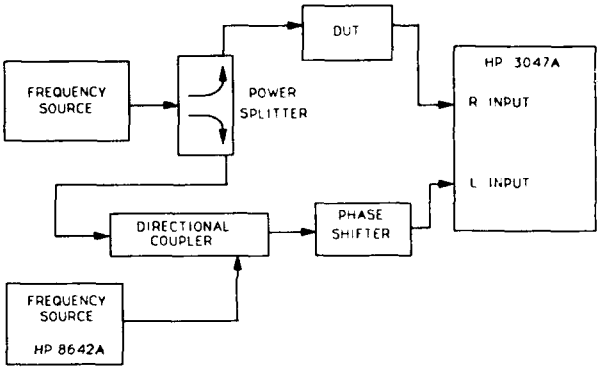


FIGURE 11 RESIDUAL PHASE NOISE MEASUREMENT TEST SETUP WITH SINGLE-SIDED SPURIOUS CALIBRATION

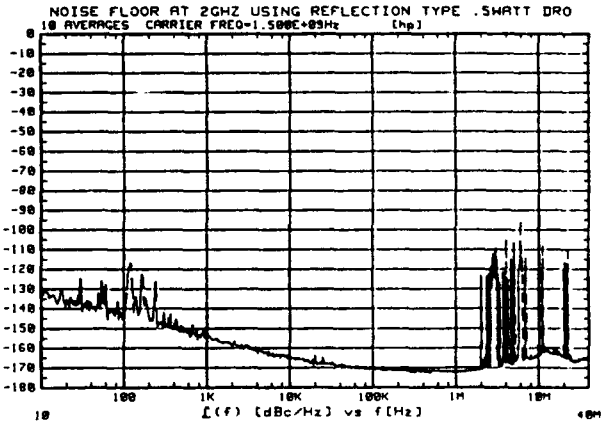


Figure 12 Noise Floor at 2 GHz Using 0.5 W Reflection-Type DRO

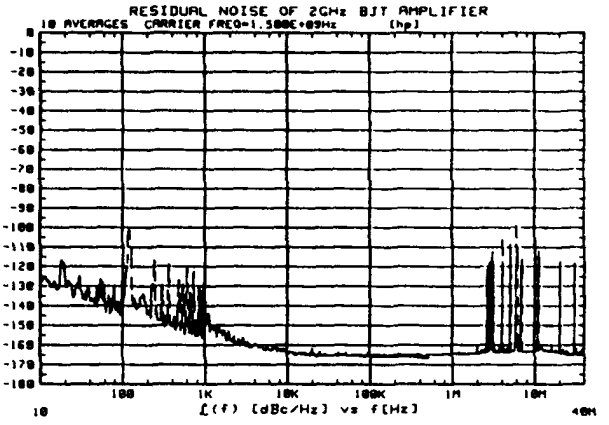


Figure 13 Residual Phase Noise of BJT Amplifier

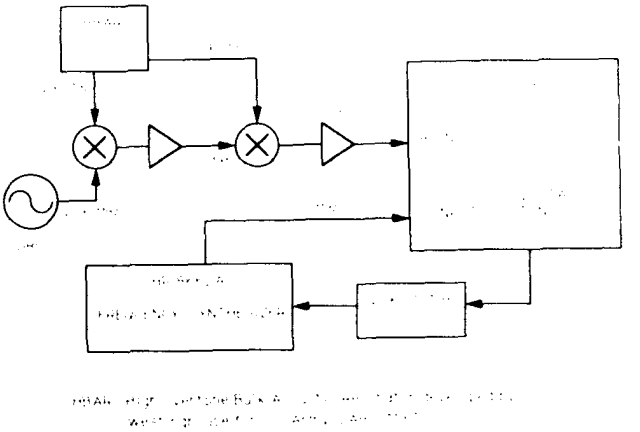


Figure 16 Absolute Phase Noise Measurement Test Setup

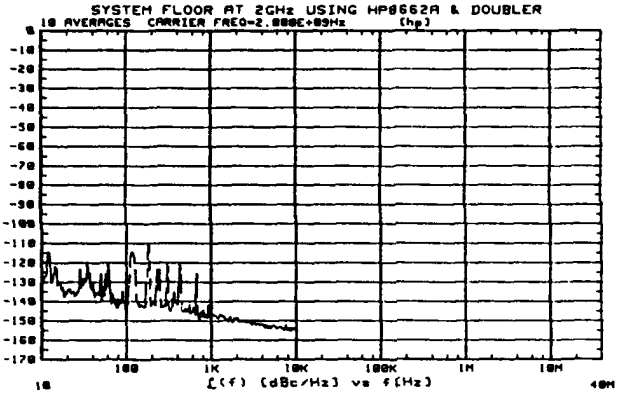


Figure 14 Noise Floor at 2 GHz Using HP8662A and Doubler

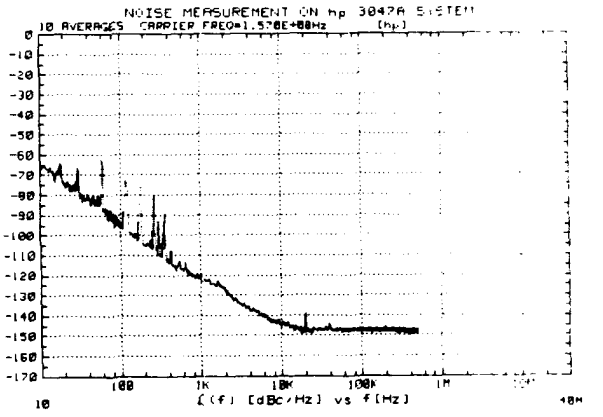


Figure 17 Absolute Phase Noise Data for Two-Stage DRO

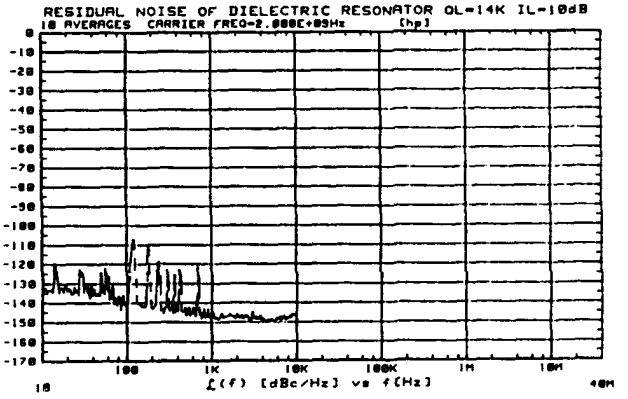


Figure 15 Residual Phase Noise of Dielectric Resonator Used with Single-Stage Amplifier

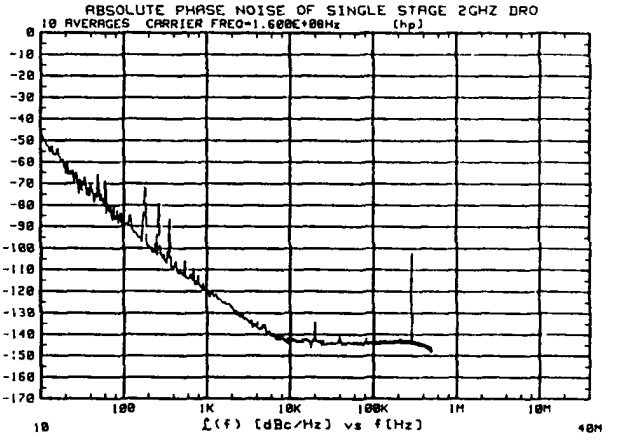


Figure 18 Absolute Phase Noise Data for Single-Stage DRO

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ULTRA-LOW-NOISE, 8.3 GHz DIELECTRIC RESONATOR OSCILLATOR

MUHAMMAD MIZAN and RAYMOND McGOWAN

US Army LABCOM, Electronics Technology & Devices Laboratory
Fort Monmouth, NJ 07703-5601

ABSTRACT

A parallel feedback 8.3 GHz Dielectric Resonator Oscillator (DRO) is designed by evaluating the residual phase noise of the GaAs FET amplifier and the resonant structure (RS). The resonant structure consists of two-port, dielectric resonator coupled to 50 ohm microstrip lines housed inside a copper cavity. The absolute phase noise of the DRO was measured by using a Hewlett Packard (HP) 3047A phase noise measurement system. The oscillator exhibited a single sideband (SSB) phase noise of -67 dBc/Hz at 100 Hz offset frequency. The loaded Q of the RS was measured to be 15,000 at room temperature. The in-house fabricated specially-designed GaAs FET amplifier exhibited a SSB residual phase noise of -150 dBc/Hz at a 1 kHz carrier offset frequency. The loaded Q of the resonant structure and the residual noise of the GaAs FET amplifier was measured again at liquid nitrogen temperature (77K). The loaded Q of the RS at 77K more than doubled but the residual phase noise of the amplifier degraded by as much as 45 dBc/Hz at 10 Hz carrier offset frequency.

INTRODUCTION

The Dielectric Resonator Oscillator (DRO) operating at 8.3 GHz reported in this paper has been shown to have excellent phase noise and temperature stability. Systems operating at X through Ku band frequency ranges with stringent performance requirements can benefit from this specially designed, high-performance DRO. System designers now have the option of selecting fundamentally operated high frequency DROs (5-10 GHz) without sacrificing the critical performance. DROs will play very important roles in future military and commercial system because of their simple construction, small size, high efficiency, low cost, spurious-free RF output spectrum, and good reliability.

OSCILLATOR DESIGN, CONSTRUCTION AND PERFORMANCE

A sinusoidal oscillator circuit can be classified as a two-port active device with a two-port passive network in a feedback configuration or a one-port active device in parallel with a one-port passive network. In order to sustain sinusoidal oscillations, the system transfer function must contain a pair of poles located slightly in the right-half complex plane. When dc power is applied, the noise will give rise to a growing sinusoidal output voltage which will eventually be limited by the circuit's nonlinearities. When the poles are relatively close to the complex axis, the nonlinearities observed will be small. When the poles are placed far into the right-half plane, the nonlinearities observed will be quite large due to large loop gain.

Figure 1 shows a block diagram of a transmission-mode dielectric resonator stabilized feedback-loop oscillator design. The analysis of basic feedback type of circuitry was first given by Leeson [1]. The feedback-loop oscillator configuration allows the circuit designer to isolate faulty components by measuring the residual phase noise of the components before the components are employed in the oscillator circuit. The flicker of frequency noise (30dB/decade) originating as a result of flicker of phase noise in the oscillator resonator and/or amplifier can be estimated using the formula given in ref. [2]. Because of the above circuit advantage, we have chosen the feedback DRO as our approach for fabricating X-band frequency sources. The top view of a two-port resonant structure is shown in Figure 2. The two-port resonant structure and the amplification stages are described in the following paragraph.

The resonant structure shown in Figure 2 consists of a zero ppm/°C dielectric resonator (DR) procured from Murata Erie, Inc. The resonator is housed inside a critically dimensioned

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

copper cavity. Coupling is achieved via 50 ohm microstrip transmission lines. The puck (DR) position and the microstrip line separation inside the cavity was optimized to give the highest loaded Q and minimum insertion loss. The resonant structure has a loaded Q of 15,000 with 13 dB of insertion loss at room temperature. Figure 3 shows the loaded Q measurement data. A mechanical tuning screw was built into the cavity to provide tuning capability. The input and the output 50 ohm microstrip line was constructed on 0.25 mm (10 mil) thick RT/duroid board. The board is soldered to the bottom of the cavity. Residual phase noise measurement was performed on the resonant structure to see the presence of 1/f noise. The test set up for the noise measurement is shown in Figure 4. As shown in Figure 4, an ETDL DRO was used as the frequency source to obtain a decent system floor. The system floor and the residual phase noise of the resonant structure are shown in Figure 5 and 6, respectively.

Two amplifiers were needed to overcome the insertion loss of the dielectric resonator and to provide excess loop gain in the feedback dielectric resonator oscillator. At first a single stage amplifier was designed using a Fujitsu FSX52WF GaAs FET. The input and the output matching network, stability, and the gain performance were analyzed using super compact CAD software. The amplifier circuit was constructed on 0.25 mm (10 mil) RT/duroid 5880 board. The amplifier was biased at 8 volt with 100 mA of drain current. The amplifier was stable and has a gain of 10 dB at 8.3 GHz. The input and the output VSWR was better than 2:1. The residual phase noise of the single stage GaAs FET amplifier was measured using a HP 3047A phase noise measurement system. The measurement test setup is shown in figure 4. The amplifier exhibited a SSB residual phase noise of better than -150 dBc/Hz at a 1 kHz carrier offset frequency. The measured phase noise plot is shown in Figure 7. After characterizing the amplifier and the resonant structure the complete oscillator was fabricated on RT/duroid 5880 board. Phase trimming was necessary to start the oscillator.

Due to lack of a second source with VCO capability, the absolute phase noise of the 8.3 GHz DRO was measured by down converting the test DRO to 95 MHz. The down conversion was achieved by mixing the 8.3 GHz DRO with a High-overtone Bulk Acoustic Resonator (HBAR) oscillator. Then the measurement was made by phase locking the 95 MHz down-converted signal to an HP 8662A frequency synthesizer driven from an external 10 MHz VCXO. The absolute phase noise measurement test set up is shown in Figure 8. The DRO exhibited a SSB phase noise of -67 dBc/Hz at a 100 Hz carrier offset frequency. The phase noise plot is shown in Figure 9. The measured data are 6-8 dBc/Hz better than any other published data.

Frequency stability and RF output power vs temperature were measured by using a computer-controlled temperature chamber. The total frequency shift from -50°C to +20°C is only 25 ppm and from -50°C to +50°C is 65 ppm. The frequency vs temperature plot is shown in Figure 10. The frequency vs temperature stability of our free-running DRO is about five times better than available production DRO. RF output power variation over the temperature range is shown in Figure 11. Typical RF output power at room temperature is +15dBm.

DIELECTRIC RESONATOR AT 77K

Behavior of the resonant structure described in figure 2 was investigated at liquid nitrogen temperature. The loaded Q of the resonant structure at 77K was measured to be about 33,000 which is about 2.5 times better than the room temperature data. The insertion loss of the resonator improved from 13 dB at room temperature to 9.5 dB at 77K. Figure 12 shows the loaded Q measurement data at 77K. The total resonant frequency drift from room temperature to 77K was less than 400 kHz.

GaAs FET AND BJT AMPLIFIERS AT 77K

An attempt was made to measure the absolute phase noise of a 9 GHz DRO at liquid nitrogen temperature. The 9 GHz DRO was built to solve frequency stability problem of a miniature beacon transponder. We were expecting to see 6-8 dBc/Hz improvement in the absolute phase noise at 77K due to loaded Q enhancement of the resonator, but instead we saw spectral degradation during the noise measurement. At this point a decision was made to measure the residual phase noise of the GaAs FET amplifier at 77K to try to determine the source of the degradation. The amplifier was constructed on RT/duroid board using a commercially available hermetic metal/ceramic package unmatched GaAs FET. The amplifier circuit was exposed to the environment during measurement. During room temperature measurement, the amplifier was placed inside a RF shielded oven and during

liquid nitrogen temperature measurement the amplifier was placed inside a RF shielded cryogenic chamber. The system floor, the residual phase noise of the amplifier at room temperature and at 77K are shown in Figure 13, Figure 14 and Figure 15, respectively. The amplifier was biased at 8VDC with 200 mA of drain current. The amplifier was completely immersed in liquid nitrogen and bubbling was observed during measurement. No attempt is being made at this time to explain why the amplifier performance degraded at 77K.

Similar measurements were performed on a MMIC 2-stage 8.6 GHz GaAs FET amplifier, a 8.3 GHz GaAs FET amplifier and a 8.9 GHz internally matched power GaAs FET amplifier. In these measurements the unpackaged amplifier was surrounded by liquid nitrogen without being immersed in it. This arrangement was necessary to avoid bubbling during measurement. Figure 16 and Figure 17 show the residual phase noise of the MMIC amplifier at room temperature and at 77K respectively. Figure 18 shows the residual phase noise of a 8.3 GHz amplifier at 77K. Room temperature data for this amplifier is shown in Figure 7. Figure 19 and Figure 20 show the phase noise of internally matched amplifier at room temperature and at 77K respectively. In each case the residual noise of the amplifiers degraded between 10Hz and 100 Hz offset frequencies.

A 2 GHz BJT amplifier was also designed and tested at 77K. This amplifier circuit was constructed on duroid board which was also exposed to the environment during measurement. The gain of the BJT amplifier dropped from 15 dB at room temperature to 8 dB at 77K. The amplifier was completely immersed in the liquid nitrogen during measurement. Figure 21 shows the system floor at 2 GHz. Figure 22 and Figure 23 show the amplifier residual phase noise at room temperature and at 77K respectively. The 1/f noise of the BJT amplifier at room temperature is masked by the system noise as indicated by the measured system noise floor. The phase noise of BJT amplifier at 77K degraded between 10 Hz and 1 kHz carrier offset frequencies with nominal bias condition. However, the phase noise performance of the amplifier at 77K improved as the bias condition was increased. The result is shown in figure 24.

CONCLUSIONS

The free-running 8.3 GHz DRO reported here exhibits state-of-the-art phase noise and frequency stability. The oscillator is small in size, 2.54 cm x 2.54 cm x 2.54 cm (1 cubic inch) and offers a spurious-free output spectrum. The basic design of the oscillator is of simple construction and employs low cost commercially available electronic components.

The loaded Q enhancement of the dielectric resonator and the noise degradation of unpackaged GaAs FET and BJT amplifiers at 77K has been reported. Investigations are underway to explain the phase noise degradation of GaAs FET amplifier at 77K.

ACKNOWLEDGMENTS

The authors would like to thank Richard Babbitt and Tom Koscica for providing support materials for the DRO and Eric Lenzing for the use of his cryogenic chamber. Special thanks also go out to Frank McBride & Ed Hughes for fabricating the cavities and Ed Maluszczak from EW/RSTA for supporting this project.

REFERENCES

[1] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proceedings of the IEEE, Vol. 54, No. 2, pp. 329-330, February 1966.

[2] G.K. Montress and T.E. Parker, "Design Techniques for Achieving State-of-the-art Oscillator Performance", Proceedings of the 44th Annual Symposium on Frequency Control 1990, pp. 522-535. IEEE Catalog No. 90CH2818-3.

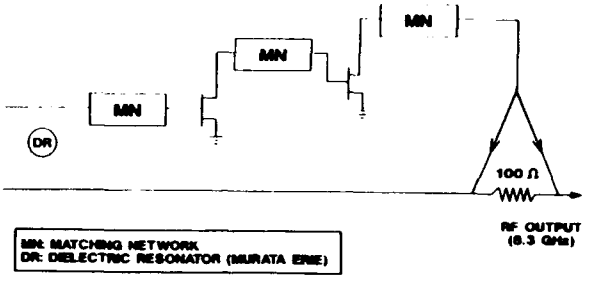


FIGURE 1. FEEDBACK DRO BLOCK DIAGRAM.

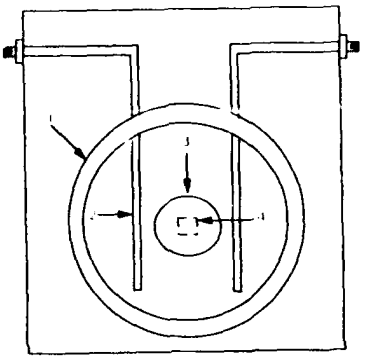


FIGURE 2. TOP VIEW OF TWO-PORT RESONANT STRUCTURE.

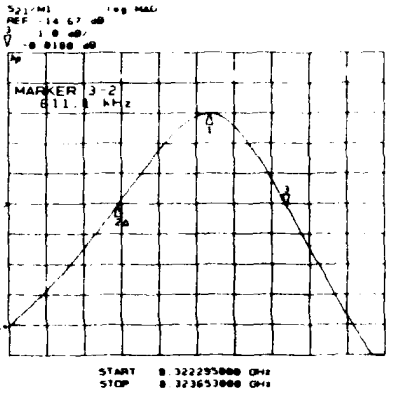


FIGURE 3. LOADED Q MEASUREMENT AT ROOM TEMPERATURE.

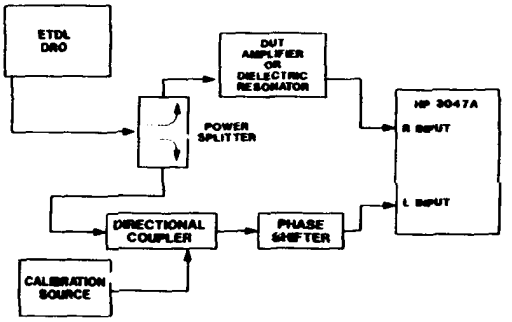


FIGURE 4. AMPLIFIER/RESONANT STRUCTURE RESIDUAL PHASE NOISE MEASUREMENT TEST SET-UP WITH SINGLE-SIDED SPURIOUS CALIBRATION.

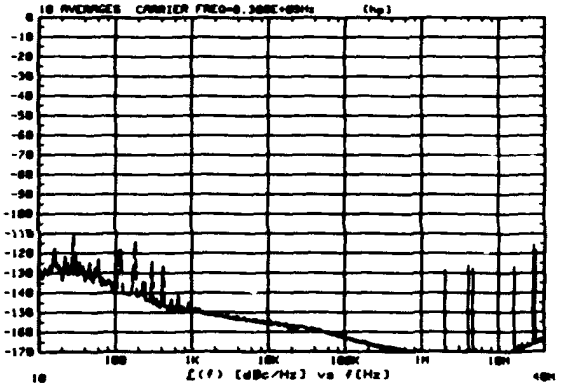


FIGURE 5. SYSTEM FLOOR AT 8.3 GHz USING ETDL DRO AS FREQUENCY SOURCE.

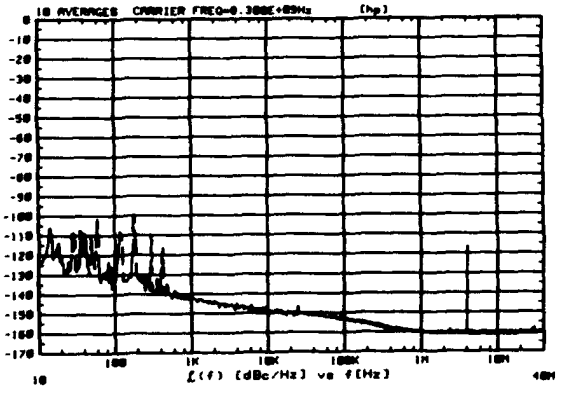


FIGURE 6. RESIDUAL PHASE NOISE OF RESONANT STRUCTURE.

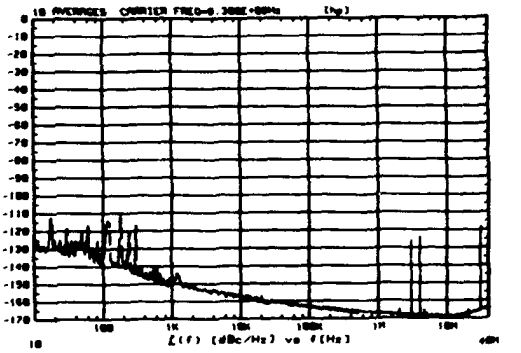


FIGURE 7. RESIDUAL PHASE NOISE OF SINGLE-STAGE 8.3 GHz GaAs FET AMPLIFIER AT ROOM TEMPERATURE.

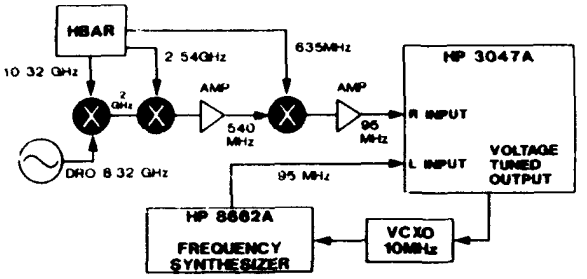


FIGURE 8. 8.3 GHz DRO PHASE NOISE MEASUREMENT TEST SETUP.

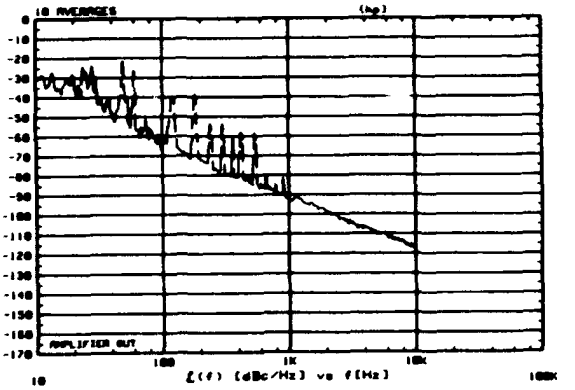


FIGURE 9. ABSOLUTE PHASE NOISE OF 8.3 GHz DRO.

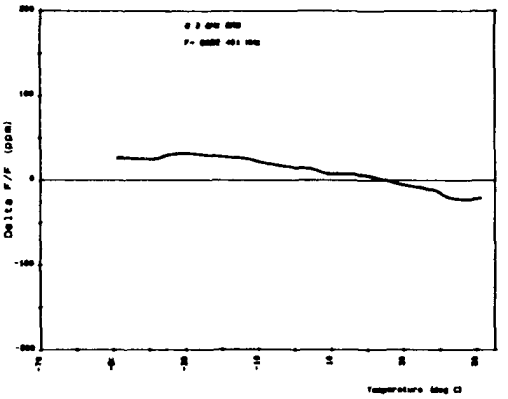


FIGURE 10. FREQUENCY vs TEMPERATURE PLOT.

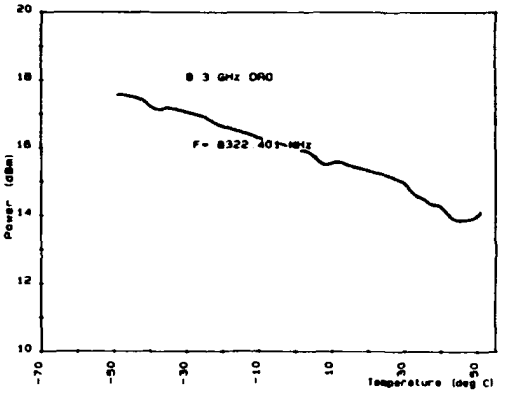


FIGURE 11. RF OUTPUT POWER vs TEMPERATURE.

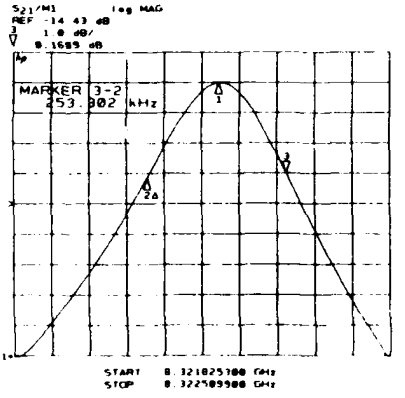


FIGURE 12. LOADED Q AT 77K.

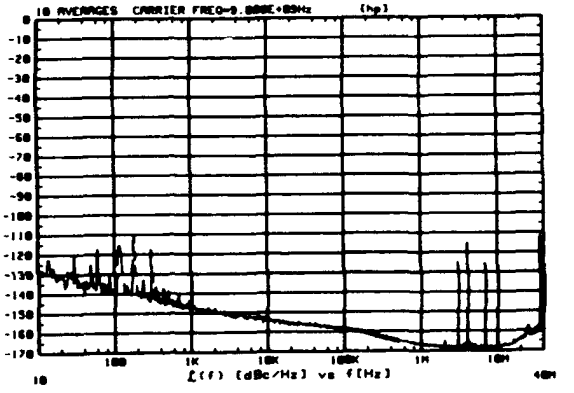


FIGURE 13. SYSTEM FLOOR AT 9 GHz.

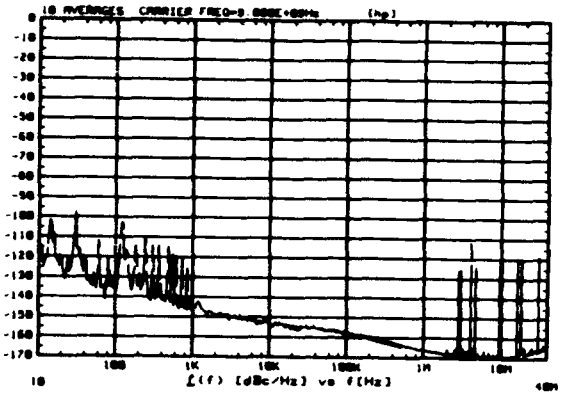


FIGURE 14. RESIDUAL PHASE NOISE OF 9 GHz GaAs FET AMPLIFIER AT ROOM TEAMPERATURE.

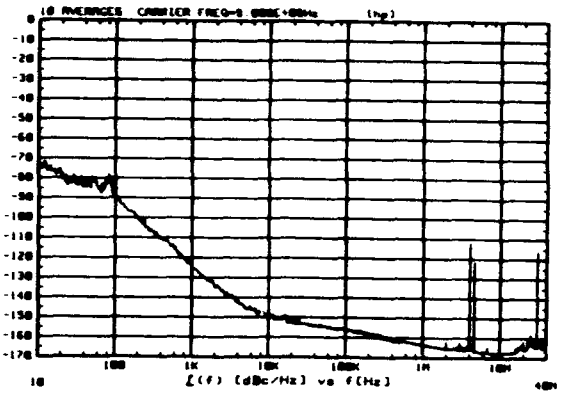


FIGURE 15. RESIDUAL PHASE NOISE OF 9 GHz GaAs FET AMPLIFIER AT 77K.

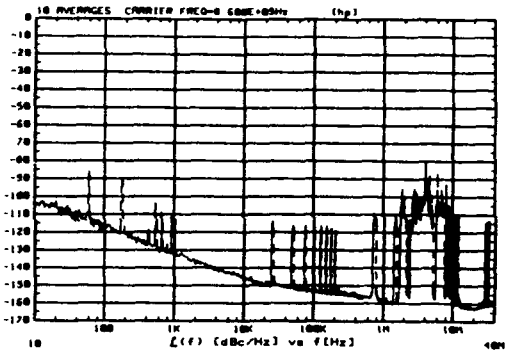


FIGURE 16. RESIDUAL PHASE NOISE OF 8.6 GHz MMIC 2-STAGE GaAs FET AMPLIFIER AT ROOM TEMPERATURE

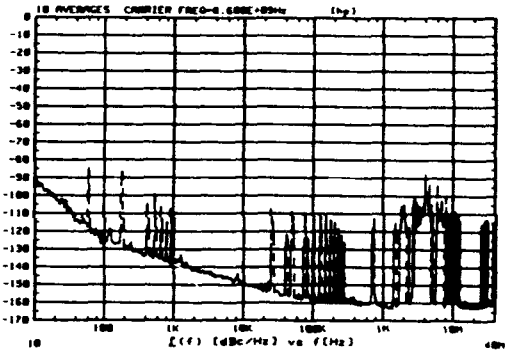


FIGURE 17. RESIDUAL PHASE NOISE OF 8.6 GHz MMIC 2-STAGE GaAs FET AMPLIFIER AROUND 77k.

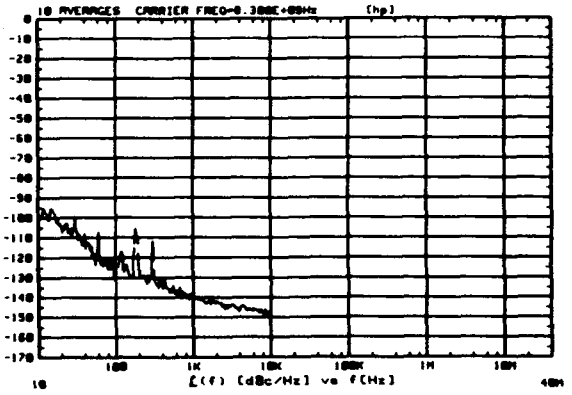


FIGURE 18. RESIDUAL PHASE NOISE OF 8.3 GHz GaAs FET AMPLIFIER AROUND 77K.

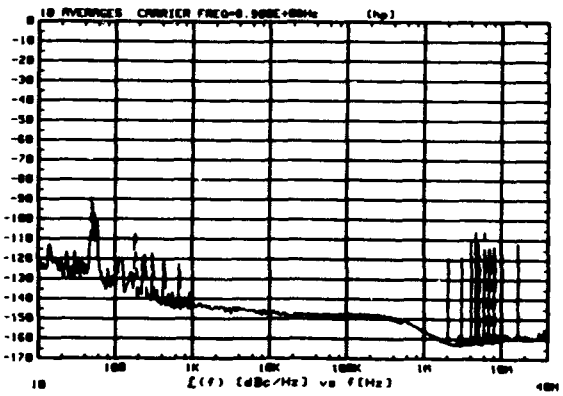


FIGURE 19. RESIDUAL PHASE NOISE OF INTERNALLY MATCHED GaAs FET AMPLIFIER AT ROOM TEMPERATURE.

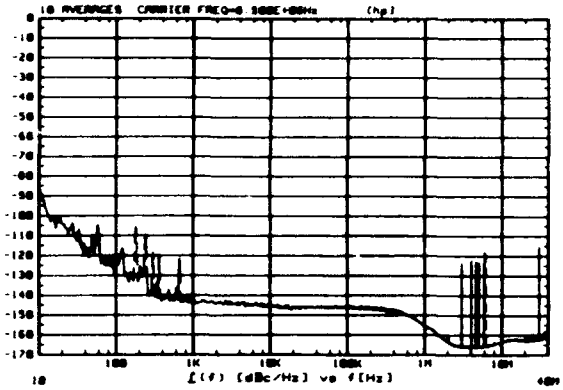


FIGURE 20. RESIDUAL PHASE NOISE OF INTERNALLY MATCHED GaAs FET AMPLIFIER AROUND 77K.

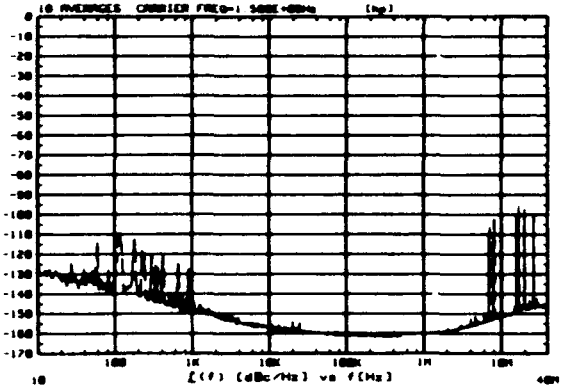


FIGURE 21. SYSTEM FLOOR AT 2 GHz.

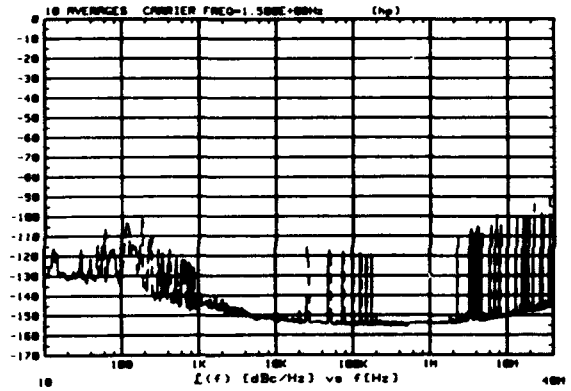


FIGURE 22. RESIDUAL PHASE NOISE OF 2 GHz BJT AMPLIFIER AT ROOM TEMPERATURE.

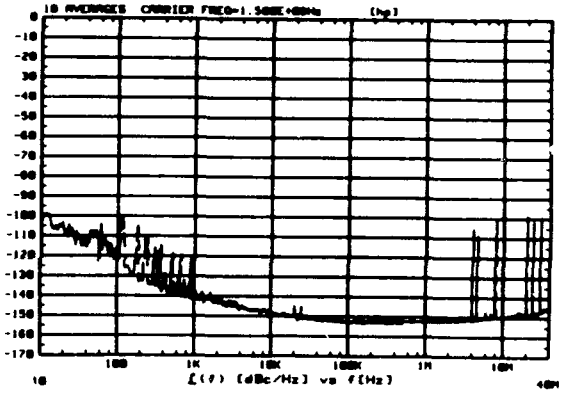


FIGURE 23. RESIDUAL PHASE NOISE OF 2 GHz BJT AMPLIFIER AT 77K.

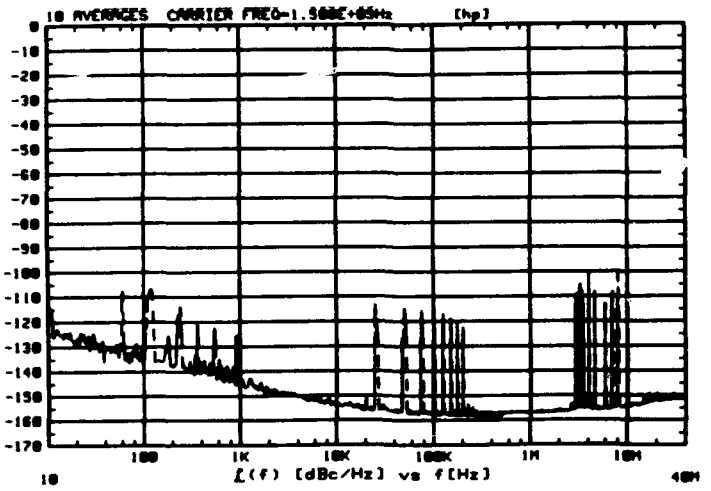


FIGURE 24. RESIDUAL PHASE NOISE OF 2 GHz BJT AMPLIFIER AT 77K WITH INCREASED BIAS CONDITION.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

COOLED, ULTRA-HIGH Q, SAPPHIRE DIELECTRIC RESONATORS FOR LOW NOISE, MICROWAVE SIGNAL GENERATION*

M. M. Driscoll, J. T. Haynes, S. S. Horwitz, R. A. Jelen
Westinghouse Electronic Systems Group
Baltimore, MD 21203

R. W. Weinert, J. R. Gavaler, J. Talvacchio, G. R. Wagner
Westinghouse Science & Technology Center
Pittsburgh, PA 15235

K. A. Zaki, Xiao-Peng Liang
University of Maryland
College Park, MD

ABSTRACT

Ultra-high Q, X-band resonators, used in a frequency discriminator for stabilization of a low noise signal generator, can provide a means of obtaining significant reduction in phase noise levels. Resonator unloaded Qs on the order of 500K can be obtained in a sapphire dielectric resonator (DR) operating on a low order (i.e., TE_{01}) mode at 77K and employing high temperature superconducting (HTS) films installed in the DR enclosure covers.

Rigorous analysis for the determination of resonator frequency, modes, and unloaded Q have been carried out using mode matching techniques. Tradeoff studies have been performed to select resonator dimensions for the optimum mode yielding highest unloaded Q and widest spurious mode separation. Field distributions within the resonator have been computed to enable practical excitation of the required mode.

The results of both analysis and prototype device evaluation experiments are compared for resonators fabricated using enclosures consisting of conventional, metal sidewalls and covers employing high temperature superconducting films as a function of cover (i.e., HTS film) conductivity.

INTRODUCTION

In a low noise, microwave signal generator, output signal phase noise characteristics are largely determined by generator reference oscillator resonator characteristics including loaded Q, insertion loss, operating frequency, drive level, short-term and long-term frequency stability, and frequency sensitivity to environmental stress [1,2].

Figure 1 depicts the X-band microwave signal phase-noise performance attainable using: (1) a VHF quartz crystal oscillator followed by a frequency multiplier and (2) a dielectric resonator oscillator (DRO) operating directly at the microwave frequency. As implied by the figure, a tradeoff exists with regard to near-carrier vs. noise floor spectral performance. The higher, relative noise floor level exhibited by the crystal oscillator-multiplier is an unavoidable consequence of the phase noise sideband level increase cause by the signal frequency multiplication process. On the other hand, the higher, relative near-carrier noise level exhibited by the DRO is primarily a consequence of the dielectric resonator loaded Q value (typically several thousand at X-band) that is much lower than that of the VHF quartz crystal resonator (100,000 at 80 MHz, for example). In order to simultaneously achieve the excellent noise floor performance provided by the DRO and near-carrier noise levels that are significantly superior to that currently attainable using quartz acoustic resonators, it is necessary to utilize resonators operating directly at microwave frequencies that exhibit: (1) loaded Q values in excess of several hundred thousand and (2) acceptably low values of self-noise and environmental stress sensitivities. The predicted phase-noise performance attainable through the use of ultra-high Q microwave resonators is shown in curves 3 and 4 in Figure 1.

The use of cooled, sapphire dielectric, microwave resonators as stable, ultra-high Q, frequency control elements has been reported in the literature. Resonator operation of high order modes has resulted in attainment of unloaded Qs in the range 10^7 to 10^9 at

*This work was partially supported by DARPA, Contract No. 972-88-C-0050.

CH2965-2/91/0000-700 $1.00 © 1991 IEEE

X-band in the liquid nitrogen to liquid helium temperature range [3,4]. The presence of closely frequency-spaced multiple resonant modes can constitute a disadvantage associated with device operation on a high order mode.

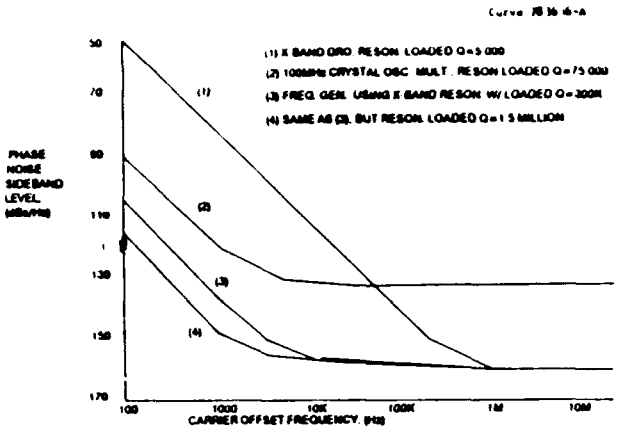


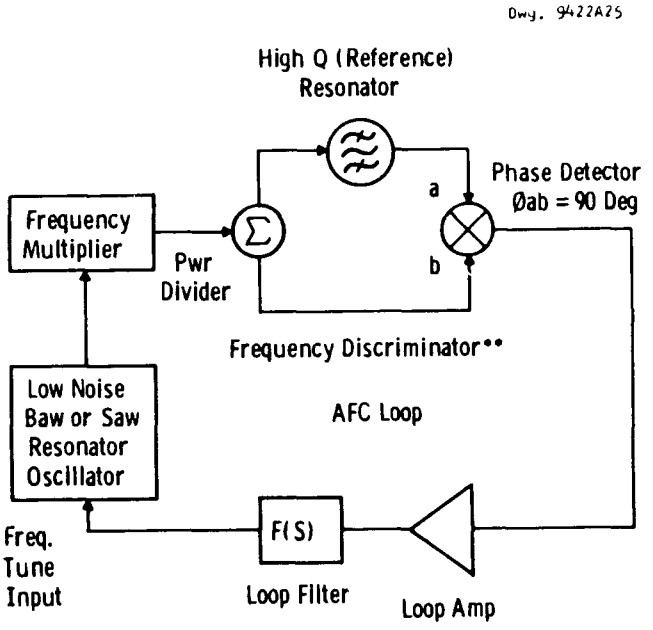
Figure 1. X-Band Signal Spectral Performance as a Function of Resonator Technology.

In a low noise, microwave signal generation application, lower resonator Qs (on the order of 500K) can provide improved signal spectral performance, compared to that currently available using conventional, non-cooled resonator technologies. In the case of resonator utilization in an AFC stabilization feedback loop, the lower (but significant) Q may be preferable from the standpoint of obtaining large loop correction bandwidth. The only resonator technology capable of providing high values, small size, and good mode separation is a cooled sapphire dielectric resonator operating on a low order mode. The loss tangent of sapphire decreases significantly with temperature [5], from ~10^{-5} at room temperature to ~10^{-7} at 77K. Attainable resonator unloaded Q values, for low order modes, at 77°K are limited not by sapphire losses but primarily by the surface losses of the metal enclosure in which the cylindrical, sapphire "puck" is mounted.

For a conventional, metal wall enclosure, resonators designed to operate on a low order mode exhibit unloaded Q values at X-band of several hundred thousand. When high temperature superconducting films are used in the resonator enclosure covers, the resultant decrease in surface resistance results in of unloaded Q values on the order of 1 million. Q values on the order of 0.5 million to 1 million are adequate for significant additional improvement in near-carrier phase noise level. The combined use of high temperature superconducting films and cooled sapphire dielectric technologies is the key to obtaining the requisite resonator performance.

HTS RESONATOR UTILIZATION

The proposed circuit mechanization for utilizing the HTS sapphire dielectric resonator, as shown in Figure 2, employs the resonator in a microwave discriminator. Ultra-high resonator Q results in achievement of high discriminator sensitivity, allowing the detection of FM/PM noise levels in a microwave signal derived from a conventional, low noise acoustic wave oscillator multiplier. As indicated in the figure, the discriminator may be configured to utilize either the resonator transmission or reflection response, or both [5,6]. The reflection and/or transmission plus reflection schemes have the advantage of providing higher values of effective detector gain. An automatic frequency control (AFC) feedback loop is employed for reduction of oscillator-multiplier signal phase noise.



**Resonator May be Operated in Reflection Mode as Well as Transmission Mode, or Both

Figure 2. HTS Resonator Utilization in a Frequency Discriminator for AFC Stabilization of an External Oscillator-Multiplier.

There are several important advantages provided by the AFC feedback loop approach, compared to resonator use directly in a microwave oscillator. With the HTS-sapphire dielectric resonator employed in an X-band oscillator, the oscillator sustaining stage would require a GaAs amplifier. GaAs amplifier flicker-of-phase noise levels (converted to flicker-of-frequency noise in the oscillator) are 20 to 30 dB higher than those exhibited by the Schottky diode phase detector

employed in the discriminator approach. In addition, the discriminator does not dissipate DC power (as an oscillator would), so that integration of the non-resonator discriminator components are part of the cooled resonator assembly is possible, and cryocooler heat removal requirements are minimized. An integrated discriminator assembly is desirable from the standpoint of eliminating the need for critical length RF cables to traverse the cooled/uncooled interface. Finally, the AFC loop results in achievement of signal phase-noise reduction not only in the final microwave frequency output signal, but also in the oscillator and multiple interstage signals.

HTS RESONATOR DESIGN

The structure of the resonator is shown in Figure 3. It consists of a cylindrical dielectric resonator of relative dielectric constant e_{r1}, diameter d, and length t, placed symmetrically inside a cylindrical conducting enclosure of diameter D and length L. The resonator is supported on its ends by two low loss foam discs of relative dielectric constant e_{r2}, diameter D and length G each.

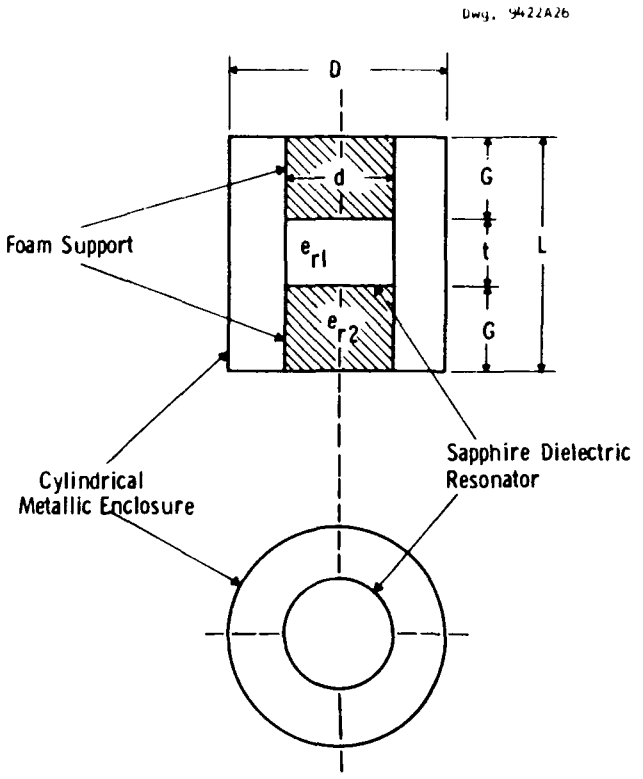


Figure 3. Structure of the Resonator.

Rigorous methods for the analysis of this type of structure have been recently developed [7-11]. These methods are based on the mode matching technique and involve significant amounts of numerical computations. The method divides the region within the conducting enclosure into subregions whose eigenmode fields can be determined analytically. In each of the sub-regions, the electromagnetic fields are expanded as a linear combination of the sub-region eigenmodes that satisfy Maxwell's equations and the boundary conditions on the conducting surfaces. The boundary conditions at the interfaces among the sub-regions are then introduced. By applying a weighted integrated criterion to the equations expressing the boundary conditions, an infinite set of linear homogeneous equations in the unknown coefficients of the eigenmodes is obtained. The characteristic equation for the resonant frequencies of the structure is obtained by equating the determinant of a truncated subsystem of N equations from the infinite set, to zero. Once the resonant frequencies are determined (by numerical solution of the characteristic equation), they are substituted back into the homogeneous system of equations, which is then solved numerically for all the eigenmode coefficients in terms of one single coefficient. The knowledge of the resonant frequencies and the corresponding eigenmode coefficients provide complete characterization of the electromagnetic fields in the structure. Determination of the unloaded Qs involves the calculation of the energy (U) stored in the structure, and the power (W_L) lost in the metallic walls, the dielectric resonator and the surrounding supporting media. The unloaded Q is then calculated from:

$$Q_u = \frac{\omega_o U}{W_L}$$

where ω_o is the resonant frequency.

Contribution to W_L can be separated for each of the enclosure surfaces (i.e., the top and bottom planes of enclosure, and the cylindrical walls), for the resonator loss and for the supporting foam. This helps to understand the loss mechanism and in optimizing the structure parameters to obtain the highest Q, when HTS films are installed in the enclosure end planes (covers).

The resonant electromagnetic field configurations that can exist on the resonant structure shown in Figure 3 are:

(1) Transverse electric modes (TE_{on} modes) which have no axial electric field and its fields have no angular variation.

(2) Transverse magnetic modes (TM_{om} modes) which have no axial magnetic fields and its fields have no angular variation.

(3) Hybrid electromagnetic modes (HE_{mn}) modes which have both electric and magnetic fields in the axial direction, and its fields have angular variations of sin ($n\phi$) and cos ($n\phi$).

Since the objective of this work is to achieve high unloaded Qs in the 10 GHz frequency range, the TE_{on} modes which generally have the highest Q values were considered. In particular the small volume TE_{01} mode DR is considered in detail. Figure 4 shows the computed results of the variation of the resonant frequency at 77K of the TE_{01} with the end wall spacing G of the enclosure, where measured values are indicated in the figure. The total unloaded Q and the contributions of the various parts of the structure to the unloaded Q are shown in Figures 5 and 6 for room temperature and 77K, respectively. In these figures the total unloaded Q is related to the individual contributions by:

$$\frac{1}{Q_T} = \frac{2}{Q_{EW}} + \frac{1}{Q_{SW}} + \frac{1}{Q_{DR}} + \frac{1}{Q_{foam}}$$

where: $1/Q_{EW}$ is the contribution due to one end wall, $1/Q_{SW}$ is the contribution due to the side wall, $1/Q_{DR}$ is the contribution due to the sapphire resonator, and $1/Q_{foam}$ is the contribution due to the foam support.

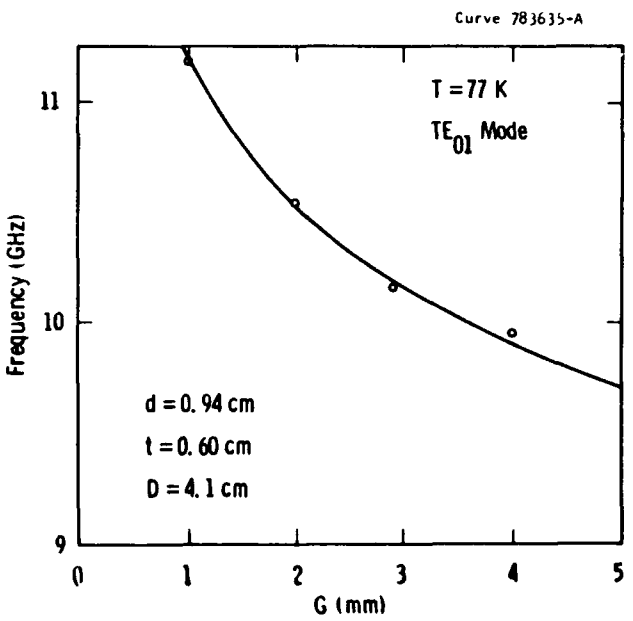


Figure 4. Resonator (TE_{01} mode). Resonant Frequency vs. End Wall Spacing.

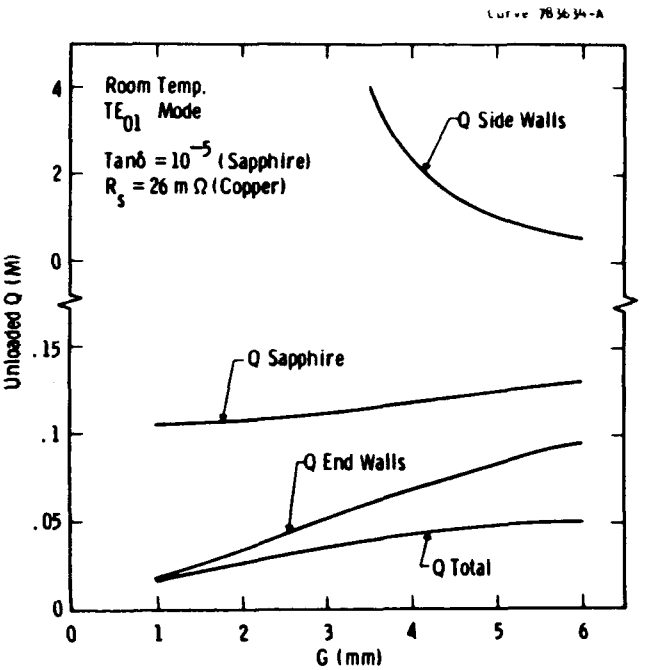


Figure 5. Contributors to Resonator Q-Factor at Room Temperature.

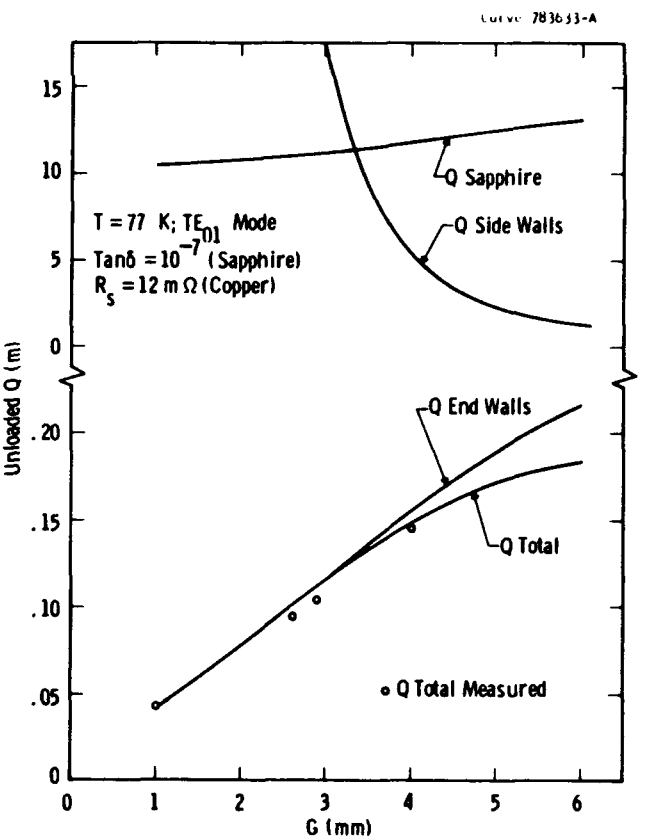


Figure 6. Contributors to Resonator Q-Factor at 77K.

At room temperature the total unloaded Q is controlled by the dielectric loss of the sapphire puck and the ohmic loss of the end walls (covers). At 77K, the dielectric loss in sapphire becomes small, and the total Q is controlled primarily by the ohmic loss in the end walls and loss in the solid foam sapphire puck supports is not considered. The foam supports actually used to obtain the measured Q values (Figure 6) were not a solid, but a "ring" structure, and the measured Q values obtained are close to the calculated "no foam" values. The difference between calculated and measured Q values is taken as a measure of the loss of the foam ring mounting structure.

Examination of the Figure 6 data indicates that, for resonator-to-end plane distances less than 0.5 cm, the primary limitation in attainable Q is due to enclosure end walls losses. Reduction of end plane losses via utilization of HTS films installed in the enclosure covers provides a means of obtaining higher overall resonator Q, with additional Q increase obtainable using a lower loss (compared to foam) DR support structure. Figure 7 shows attainable (TE_{01} mode) resonator Q values for different ratios of HTS film conductivity (compared to copper) values.

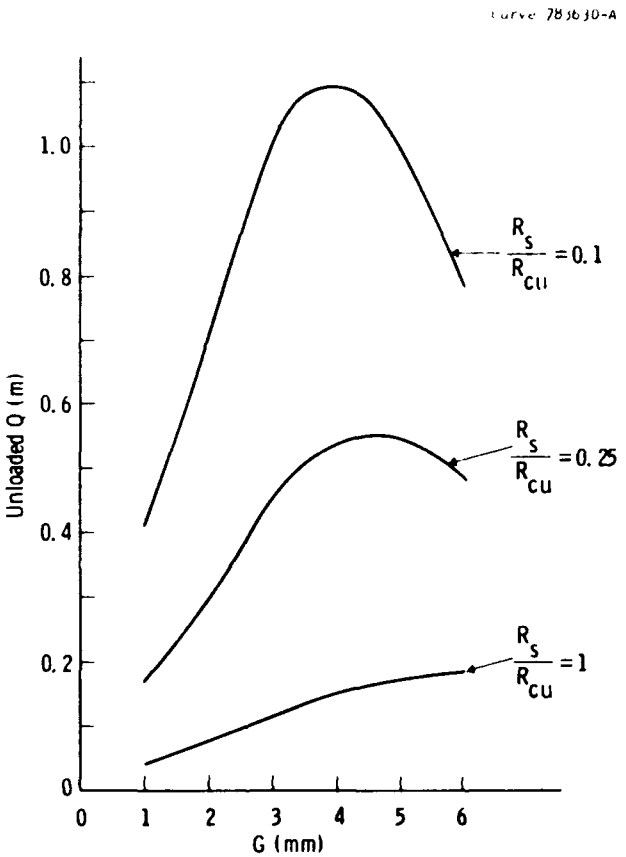


Figure 7. Resonator Unloaded Q as a Function of End Plate Surface Resistance and DR-to-End Plate Dimension.

HTS FILM FABRICATION

The YBCO films utilized were made by a single target sputtering techniques which has been previously described [12]. The substrates were positioned on a quartz holder located 90° off-axis below the sputter gun. There was no bonding or clamping of the substrate to the holder. The substrates were 1.4 inch diameter, 20 mil thick $LaAlO_3$, and the YBCO films were approximately 300 nm thick.

The particular films used in this work were too large for characterization. Films made under nominally identical conditions on large-area substrates were diced for measurement of structural and superconducting properties. The films had a c-axis growth orientation with an x-ray rocking curve width of 0.2 degrees. They grew epitaxially with their c-axis normal to the substrate and with the a and b axes of YBCO parallel to <100> directions in the $LaAlO_3$ substrate. Typical superconducting transition temperatures were 89-91K for resistive transitions (R = 0) and 88-90K for transitions measured by ac susceptibility. The rf surface resistance, R_s was measured on small chips, 1/4 × 1/2 inch, by a parallel-plate resonator technique [13]. At 77K and 10 GHz, R_s varied from wafer to wafer from 0.3 to 1.0 mΩ with variation across a single wafer of ±10% [14].

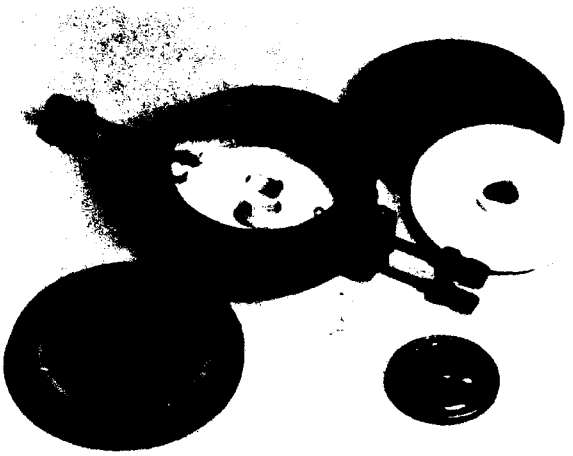
HTS RESONATOR FABRICATION AND TEST

The TE_{01} DR configuration is shown in Figure 8a. The styrofoam ring mounts are compressed between the enclosure covers and hold the sapphire puck firmly in place. The 1.4 inch diameter YBCO/$LaAlO_3$ disks, placed between the foam supports and covers, are held in place with the compressed foam. Loop couplers are used to excite and detect the device resonant response, and the degree of coupling was adjusted by changing the orientation of the loop and/or the loop penetration into the cavity.

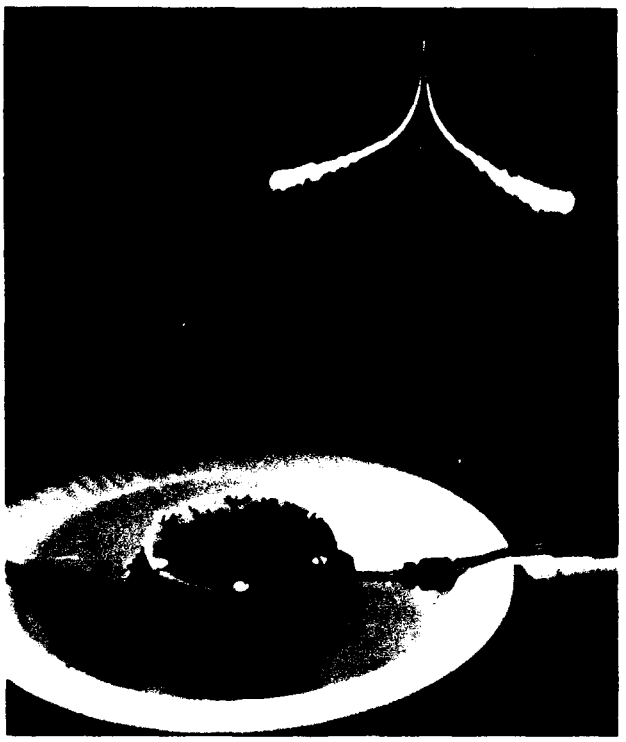
Liquid nitrogen (LN) was used as the coolant, and all measurements were made with the unsealed DR assembly in contact with, but not submerged in, the LN near the bottom of a dewar enclosure. The styrofoam plate of Figure 8b was used as a "Dewar" for the photograph and allows showing the complete assembly.

The magnitude of S21, S11, S22, QL, and frequency were measured using an automatic network analyzer (ANA) and synthesized sweep generator combination. Coupling coefficients and unloaded Q

values were calculated from the ANA loaded Q and reflection loss measurements. Most of our measurements were made at 10.16 GHz (G = 2.9 mm).



(a)



(b)

Figure 8. (a) HTS-Sapphire Dielectric Resonator. (b) Assembled and Operating at T ~ 77K.

Results of initial measurements are listed in Table 1 for copper and for (two different conductivity values) YBCO end walls. Resistance ratios, $R_{s,Cu}$:$R_{s,YBCO}$ (10 GHz, 77K), are calculated from the measured Q_u and foam support ring mount loss. The highest resistane ratio was 4.0. Resistance ratios of greater than 10 are typically obtained at 77K for small area films cut from large wafers.

$\frac{R_s\ (Cu)}{R_s\ (YBCO)}$	Q_u (k)
1.0	102
2.5	200
4.0	320

RESONATOR SHORT TERM FREQUENCY STABILITY

In an effort to verify that the HTS portion of the resonators did not contribute to excess resonator self-noise, prototype resonators were tested at Ft. Monmouth, NJ using an HBAR (higher overtone, bulk acoustic resonator)-based, X-band signal generator designed and built by Westinghouse for the U.S. Army Electronic Device and Technology Laboratory [15]. The HBAR-based source exhibits state-of-the-art phase noise performance that is only 10-15 dB poorer than that predicted for the HTS-sapphire-based source (Figure 1). The results of these measurements are shown in Figure 9 and indicate that, within the resolution of the test equipment, the HTS-sapphire resonator self-noise level is immeasurable. The primary contributor to instrumentation noise level is associated with low test set phase detector drive and resulting sensitivity (70 mv/radian) rather than the HBAR-based source spectral characteristics. This result is significant and encouraging, especially, in light of the relatively high (16 dBm) level of resonator drive employed.

In addition to resonator self-noise measurements, the phase noise performance of low 1/f noise, X-band, Schottky diode phase detectors (double balanced mixers) operated at 77K was also measured. These tests were performed in order to demonstrate the feasibility of cooling both the resonator and detector in an integrated, frequency discriminator assembly in order to avoid the need for critical length, RF coaxial cables traversing the cooled/uncooled interface. The results, shown in Figure 10, are also encouraging and show no increase in detector 1/f noise at 77K.

CONCLUSIONS

We have shown that with high quality, large area, YBCO films and simple, styrofoam mounts, unloaded, X-band Q values of over 250 thousand are

attainable for TE_{01} mode operation of a sapphire dielectric resonator, operated at 77K. With improvement in the yield of HTS films with good uniformity over areas of 10 cm^2, unloaded Q values greater than 500 thousand will be attainable from small volume, TE_{01} shielded, sapphire dielectric resonators.

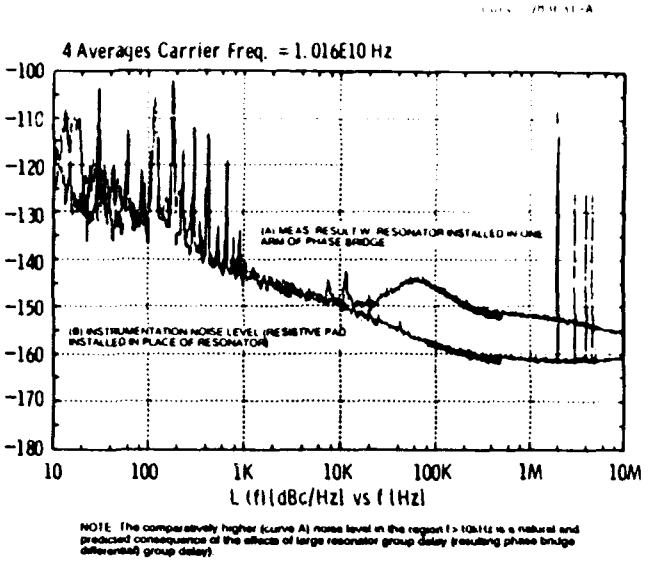


Figure 9. HTS-Sapphire Dielectric Resonator Noise Measurement Results.

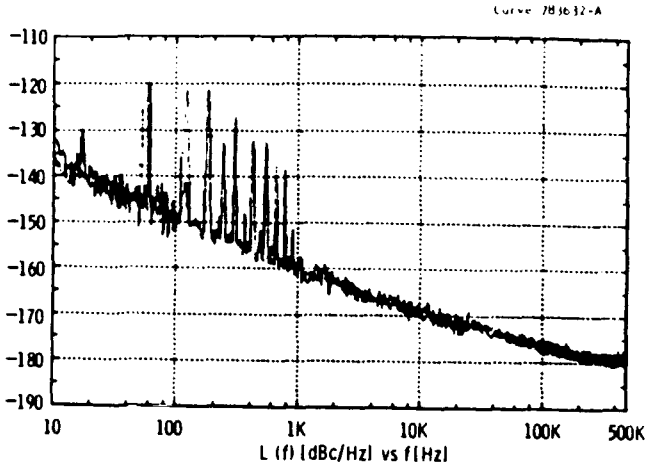


Figure 10. Measured 1/f Noise for WJ M76HC Mixer (Phase Detector) at Room Temperature and 77K.

REFERENCES

1. M. M. Driscoll, "Low Noise Signal Microwave Signal Generation Using Bulk and Surface Acoustic Wave Resonators," IEEE Trans. Ultrason. Ferroel., and Freq. Contr., Vol. 35, No. 3, May 1988, pp. 426-434.
2. M. M. Driscoll, "Advances in Resonator Technology Support Improved Microwave Signal Spectral Performance," Microwave Journal, Vol. 33, No. 6, June 1990, pp. 153-163.
3. A. Giles et al., "A High Stability Microwave Oscillator Based on a Sapphire Loaded Superconducting Cavity," Proc. 43rd Freq. Contr. Symp., May 1989, pp. 89-93.
4. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Ring Resonator for Ultra-Low Noise," Proc. 43rd Freq. Contr. Symp., May 1989, pp. 107-114.
5. V. B. Braginsky et al., Phys. Lett. A, 120, 1987, pp. 300-305.
6. F. L. Walls et al., "High Spectral Purity X-band Source," Proc. 44th Freq. Contr. Symp., May 1990, pp. 542-548.
7. K. A. Zaki and C. Chen, "Intensity and Field Distribution of Hybrid Modes in Dielectric Loaded Waveguides," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-33, pp. 1442-1447, December 1985.
8. K. A. Zaki and C. Chen, "Loss Mechanism Dielectric Loaded Resonators," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-33, pp. 1448-1452, December 1985.
9. K. A. Zaki and C. Chen, "New Results in Dielectric Loaded Resonators," IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-34, pp. 815-826, December 1985.
10. X. P. Liang, H. C. Chang, and K. A. Zaki, "Hybrid Mode Dielectric Resonators on Suspended Substrate," Proceedings of the 19th European Microwave Conference, pp. 1141-1146, September 1989.
11. H. C. Chang and K. A. Zaki, "Unloaded Q's of Axially Asymmetric Modes of Dielectric Resonators," IEEE MTT-S International Microwave Symposium, Digest, pp. 1231-1234, June 1989.
12. J. R. Gavaler and J. Talvacchio, Physics B, 1513, 1990, pp. 165-166.
13. R. C. Taber, "A Parallel Plate Resonator Technique for Microwave Loss Measurements on Superconductors," Rev. Sci. Instrum. 61(8), 2200 (1990).
14. T. T. Braggins, J. R. Gavaler, and J. Talvacchio, accepted for Proc. ICMC, Advances in Cryogenic Engineering (Materials), (Plenum, New York, 1991).
15. M. M. Driscoll et al., "Extremely Low Phase Noise UHF Oscillators Utilizing High Overtone, Bulk Acoustic Resonators," Proc. 1990 IEEE Ultrason. Symp., December 1990.

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY AND TIMING REQUIREMENTS FOR SPACE EXPLORATION*

L. Maleki
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive,
Pasadena, California 91109

Abstract

In this paper the role of time and frequency standards in space missions is briefly discussed. The present status of the frequency and timing capability in NASA's Deep Space Network is presented, together with a review of the anticipated future requirements. Based on the future needs, areas for future development in frequency and timing technology are identified.

Introduction

Frequency and timing standards have traditionally played an enabling role for space navigation. Yet an important application of frequency and timing standards is in science experiments associated with deep space exploration. A number of important radio science experiments on board deep space missions have relied on frequency standards for measurement of the constituents of the planetary atmospheres, and the size of planetary ring particles [1]. Doppler tracking experiments associated with planetary missions have produced important data on the gravity fields of planets, and have made possible experimental search for gravitational waves [1].

This paper will briefly examine the present status of frequency and timing in the NASA Deep Space Complex (DSN). The aim of the paper, however, is to point to the anticipated stability and spectral purity requirements of the future DSN space missions. The paper will explore and identify required technological developments for meeting the future space requirements.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Science Applications of Frequency and Timing

The current applications of ultra-stable frequency and timing standards for space science experiments is predominantly with the use of ground based equipment. In nearly all cases spacecraft standards consist of high quality quartz oscillators that with stability not exceeding parts in 10^{13}. Since the attainable spectral purity is limited by the signal to noise ratio of the spacecraft radio signals at the antenna, this type of standard is nevertheless adequate for some radio science experiments requiring spectral purity.

The space science use of frequency and timing sources may be divided in three categories. In the first group of experiments use is made of the spacecraft oscillator (usually refered to as "USO", or ultra-stable oscillator) to measure fluctuations in phase and frequency of the signals emitted by the spacecraft and propagating through planetary atmospheres or rings. The USO's short term stability (less than 10 s averaging intervals) is also exploited in experiments where the acceleration of the spacecraft is measured to determine the gravity field of planets.

In the second group of experiments Doppler tracking data is used to measure perturbations due to external effects on the velocity or the acceleration of the spacecraft during the cruise period. Notable amongst this type of experiments is the observation of gravitational waves [2] which rely on the "detection antenna" consisting of the earth-spacecraft system and the radio link. In these experiments gravitational waves will be observed with a particular signature on the Doppler link. Observation of gravitational waves with such a scheme will be a major discovery, but is regarded extremely difficult because of the minute size of the expected effect. Thus the gravitational wave detection experiments represent the greatest challenge to the stability of frequency

CH2965-2/91/0000-707 $1.00 © 1991 IEEE

standards by requiring stability as good as 10^{16} or more for averaging intervals corresponding to the round trip light times of the spacecraft-earth system. Since atmospheric propagation delays constitute a large source of uncertainty in Doppler tracking experiments, observation of gravitational waves with ground based standards will probably not require stability exceeding one part in 10^{17}.

The last group of experiments utilize standards for the measurement of other fundamental physical measurements and relativistic effects. These experiments typically require ultra-stable sources on board spacecraft for long term stability (greater than 10^5 seconds). Previous experiments of this type have been limited to a red shift measurement by Vessot et. al [3] which actually involved placing a hydrogen maser in a sub-orbital trajectory with a Scout rocket. A recent experiment utilizing hydrogen masers and a fiber optic distribution link has also examined the size of the possible anisotropy of the speed of light [Ref]. Finally, a measurement of the Sagnac effect utilizing the Global Positioning System (GPS) has also been recently carried out [4].

Present Capability

The details of the capability of frequency and timing subsystem (FTS) in the DSN has been discussed in detail previously [5]. Here we will review briefly this capability in connection with the requirements of current missions, i.e. those recently launched or approved for development. The extension of the Voyager mission to include rendezvous with the outer planets led to an upgrade of the FTS in the DSN. The current level of frequency stability is about one part in 10^{15} for averaging times of 1000 s, and exceeds the requirements of the (recently launched) Galileo mission to explore Jupiter, and for Mars Observer which is expected to be launched in 1992. The requirements for the spectral purity, however, have been more challenging. As mentioned above the spacecraft oscillator (USO) is typically a high quality Quartz oscillator. The ground based oscillator is naturally selected to ensure that the spectral purity of the link is not degraded. This means that ideally the spectral purity of the ground based oscillator should exceed that of the USO by about 10 dB. This requirement has been difficult to meet where the signal from the 5 MHz oscillator is multiplied up to X-band (8.1 GHz), since the multiplication scheme raises the noise level of the signal. The performance of a high quality oscillator in the DSN is given in Table I, representing a 6 dB margin in the specified requirements of spectral purity for Mars Observer mission.

PHASE NOISE	5 MHz	100 MHz
Freq Offset (Hz)	dBc	dBc
0.1	-96	-70
1	-126	-100
10	-139	-113
100	-148	-122
1000	-153	-127
10000	-153	-127
100000	-154	-128

Table I. Spectral Purity Data for a High Quality Oscillator

Current Status of the Technology

While the current capability of the FTS in the DSN essentially can meet the requirements of Galileo and Mars Observer missions, requirements of the Cassini mission cannot be met with the present system. Cassini is a mission for exploring Saturn and its moons with a planned launch date of 1996. The expected requirement of the frequency stability of the Cassini gravitational wave detection mission is in the range of 4×10^{-16} for intervals longer than a few thousand seconds. This level of stability is not achievable with the hydrogen masers. Furthermore, since an upgrade of the communications frequency to K_a band (32 GHz) is anticipated for Cassini, the current spectral purity of multiplied quartz signals is inadequate, if the same requirements of spectral purity of Mars Observer are to be met.

The anticipation of these and similar problems have provided the impetus for the development of advanced frequency standards at JPL. The Superconducting Cavity Stabilized Oscillator (SCMO) [6] has been under development to provide a source with 35 dB more spectral purity (at 10 GHz, 1 Hz from the carrier) than the highest grade quartz oscillator. This standard also has the highest stability of any microwave standard at averaging intervals below 1000 s. The latest measurements of the stability of SCMO yield stability of 3×10^{-15} from 1 to 1000 s. This level of stability enables new science experiments with spacecraft at short light distances. Such an experiment involving a search for gravitational waves is planned for Galileo spacecraft at its second earth conjunction.

A second source for spectral purity has also been under development. The Sapphire Phase Stabilizer (SPS) is expected to perform at -70 db/f^3 at X and

K_a band frequencies [7]. Here f is the Fourier frequency from the carrier. The major attribute of this cryogenic oscillator is the ability to produce a high level of spectral purity at 77 K., rendering it a simpler instrument than SCMO with 1.56 K as the operating temperature.

Improved frequency stability at longer averaging intervals is approached with the development of the trapped mercury ion frequency standard [8]. The adoption of a novel trap geometry [9] for this standard has produced stability of $\sigma = 1 \times 10^{-13}/\sqrt{\tau}$, with τ the averaging interval. This performance has been measured for intervals as long as 10^4 s, where the the limit of stability of the measurement standard (H-maser) available at JPL's Frequency Standards Test Facility is reached [10].

The superior stability performance of SCMO at short averaging intervals can be combined with the long term stability of the trapped mercury ion standard to achieve improved performance at all averaging times. This in fact is the planned upgrade of the FTS at the DSN for meeting the requirements of Cassini. Figure 1 depicts the current and expected performance of the frequency and timing subsystem in the DSN through 1995. The figure also includes the performance of the fiber optic distribution link required for distribution of the reference frequency [11].

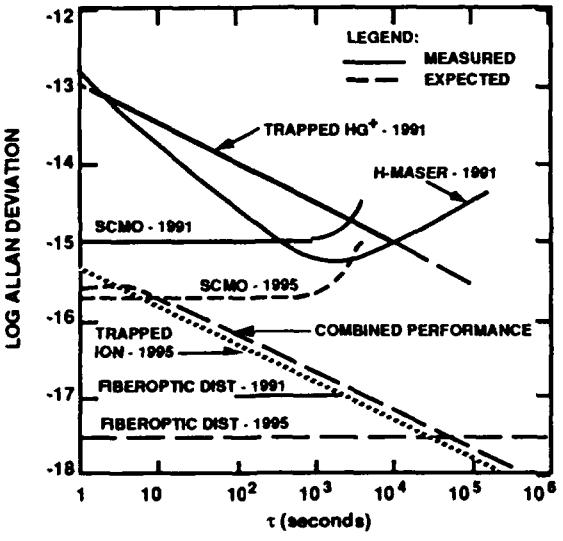


Fig 1 Status of Frequency Stability in the DSN

Future Requirements

A number of space missions for start beyond 1996, and through the first part of the 21st century are currently under study. While the specifics of the planned missions are certainly subject to modifications, and their actual selection subject to various scientific, economic, and political priorities, it is nevertheless possible to project the requirements of frequency and timing for future space missions. These projections will be based on the type of the missions under study, rather than the specific requirements of a particular mission. The projections will also address the needs in two other dimensions, *viz.* the anticipated increase to higher communications frequency, including optical frequencies, and the desirability of higher performance space-borne frequency and time standards.

Future missions may be grouped in three distinct areas with respect to their frequency and timing requirements: those requiring improved navigation, those requiring frequency and time for metrology, and those requiring frequency standards and clocks as science instruments for particular experiments. The first group represent a number of missions that will include several or clusters of spacecraft and/or landers. All three groups require improved frequency and timing stability, but the major challenge to support future missions is the development of ultra stable frequency and timing for spacecraft application, and the development of ultra- stable standards at optical frequencies.

Various missions planned as precursors to manned landing on Mars (Mars sample return missions, MRSR) which involve landing and retrieving of one or several sample gathering rovers on Mars. In many of these missions precise location of the rovers, for example, necessitate improved navigation techniques, some of which require improved clock performance. Clusters of spacecraft also require precise knowledge of spacecraft-spacecraft distance, and distance to earth. Such navigational needs point to techniques which utilize ultra-stable frequency and timing standards, including those deployed on spacecraft and rovers, and those required for optical communication links in space.

A number of future missions will have stringent requirements for frequency and time standards in connection with metrology. In particular, imaging interferometers planned for space deployment and deployment on the moon require knowledge of distance as long as tens of kilometers with a precision of a

fraction of an optical wavelength. These and similar interferometers, such as the space microwave interferometer for the detection of gravitational waves represent major challenges for frequency and time metrology in both microwave and optical wavelength regimes. In the case of lunar interferometer spectrally pure and frequency stable optical sources for hetrodyne measurement schemes will be required for deployment on the Moon. The moon also provides an ideal clock bearing satellite for performing various experiments, including gravitational red shift measurements.

Virtually all radio science experiments are adversely influenced by atmospheric fluctuations and perturbations. Similarly, the quality of Doppler tracking data is diminished by the atmospheric propagation delays. Planned Orbiting VLBI missions require undegraded performance at frequencies as high as 100 to 300 GHz. An obvious solution to this problem, insofar as frequency standards are concerned, is to place a stable standard on board spacecraft. Space- borne standards enable either the implementation of two-or multi-way communication links for calibration of the atmospheric perturbations, or tracking from a space based communications platform.

Observation of gravitational waves with laser interferometers as in the LAGOS mission rely heavily on the enabling role of frequency standards at optical frequencies with stability in the range of parts in 10^{18}. These standards will require deployment on board spacecraft throughout the several years duration of the mission. Other proposed relativity tests also require stability in the range mentioned for standards on board spacecraft.

Conclusions

The frequency and timing capability presently existing in the DSN meets the requirements of missions through 1995. Cassini mission planned for 1996 launch represents the first DSN mission requiring the stability of new standards, the SCMO and the mercury ion standard.

The future need for ultra-stable frequency and timing standards for science experiments is expected to increase with increasing opportunities to perform science from space. Astrophysical studies, planetary studies, and fundamental physics studies are but a few of the myriad of scientific diciplines poised to exploit various space based bases, platforms, and spacecraft planned for the future. Lunar and Mars exploration missions will offer even more opportunity for space based scientific studies. Virtually all space missions rely on the enabling role of frequency and timing standards, and while improved stability has been the major thrust of developments till now, future missions require space borne ultra-stable standards compatible with limitations of size, weight, and power consumption of space missions. The need for references at 100 GHz to optical frequencies are other challenges for future technology developments. Thus the frequency and timing technology will continue to play an important role in space exploration, which will also provide it with significant technological challenges throughout the 1990's and into the twenty-first century.

References

1. V. R. Eshelman, G. L. Tyler, J. D. Anderson, G. Fjeldbo, G. S. Levy and G. E. Wood, "Radio Science Investigations with Voyager," *Space Sci. Rev.*, **21**, 207-232 (1977).

2. S. J. Nelson and J. W. Armstrong, "Gravitational Wave Searches Using the DSN," *TDA Progress Report 42-49*, (JPL Publication), 75-85 (1988).

3. R. F. C. Vessot, M. W. Levine, E. M. Mattison, E. M. Bloomberg, T. E. Hoffman, G. U. Nystrom, B. F. Farrell, R. Decher, P. B. Eby, C. R. Baugher, J. W. Watts, D. L. Tauber, and F. O. Willis, "Test of Relativistic Gravitation with a Space-Borne Hydrogen Maser," *Phys. Rev. Lett.* **45**, 2081-2084 (1980).

4. D. W. Allan, M. A. Weiss and N. Ashby, "Around-the-World Relativistic Sagnac Experiment," *Sci.*, **228**, 69-70 (1985).

5. L. Maleki and P. Kuhnle, "A Review of the Frequency and Timing Activities at the Jet Propulsion Laboratory," *Proc. 18th Annual Precise Time and Time Interval Applications and Planning Meeting*, 89-97 (1986).

6. R. T. Wang and G. J. Dick, "Combined Stability of Hydrogen Maser and Superconducting Cavity Maser Oscillator", in these proceedings.

7. G. J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," in of the 43rd Annual Frequency Control Symposium, 107-114 (1989).

8. L. S. Cutler, R. P. Giffard, P. J. Wheeler. and G. M. R. Winkler, "Initial Operational Experience with a Mercury Ion Storage Frequency Standard," in *Proceedings of the 41st Annual Symposium on Frequency Control,* 12-19, (1987). item J. D. Prestage, G. R. Janik, G. J. Dick and L. Maleki,"Linear Ion Trap for Second-Order Doppler Shift Reduction in Frequency Standard Applications, *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.*, **37**, 535-542, (1990).

9. J. D. Prestage, R. L. Tjoelker, G. J. Dick and L. Maleki, "Ultra-Stable Hg^+ Trapped Ion Frequency Standard," in these proceedings.

10. L. E. Primas, R. T. Logan and G. F. Lutes, "Applications of Ultra-Stable Fiber Optic Distribution Systems," *Proceedings of 43rd Symposium on Frequency Control*, 202-221 (1989).

FORTY-FIFTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A Wide Range (550-700 MHz) Monolithic Timing Recovery Circuit

E. M. Frymoyer and B. Lai

Hewlett Packard, San Jose, CA

ABSTRACT

A retiming technology has been developed which enables timing recovery of a pseudorandom (2^{23}-1) data sequence to be accomplished in a single IC chip at high bit rates (550-700 Mbps). The retiming technology incorporates a digital bang-bang (thermostat like) method to adjust the on board IC ring oscillator to the rate of the incoming data. Both timing extraction and data regeneration (shape and phase) are accomplished on the same chip.

The retiming IC is realized in a high speed (f_t>12 GHz) silicon bipolar process. For a 622 Mbps input rate, the recovered timing has a jitter of less than 2 Degrees phase. The technology provides a recovered timing lock in less than 2 ms. Performance of the chip under a variety of conditions including temperature are presented.

Performance and Environmental Conditions

Input Conditions:

- •Signal level 100 mV Peak to Peak
- •Signal type— Data, pseudorandom to 2^{23}-1

Environment Conditions:

- •Supply Voltage V_{EE} =-5.2 V
- •Reference Voltage V_{TT} = -2.0 V

Process technology:

- •HP10, a 10 GHz min f_t silicon bipolar process
- • Chip size 2.8x2.8 mm

Performance:

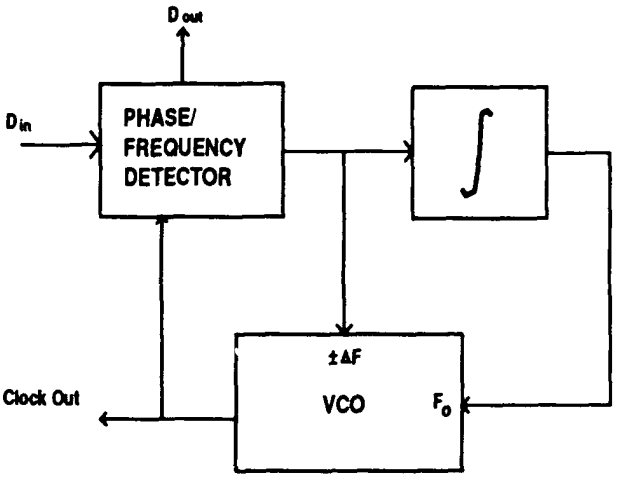
- •Retiming clock and inverse at 600 mV min ECL levels
- •Retimed data and data inverse at 600 mV ECL levels
- •Lockup time < 2 ms
- •Jitter generation < 8 ps at 622 Mb/s with 2^{23}-1 pseudorandom data (test set input jitter ~4ps)

Power Dissipation: < 1.5 W

Test Package: Special 16 lead surface mount

User adjustment required: NONE

Block Diagram of Retiming Circuit

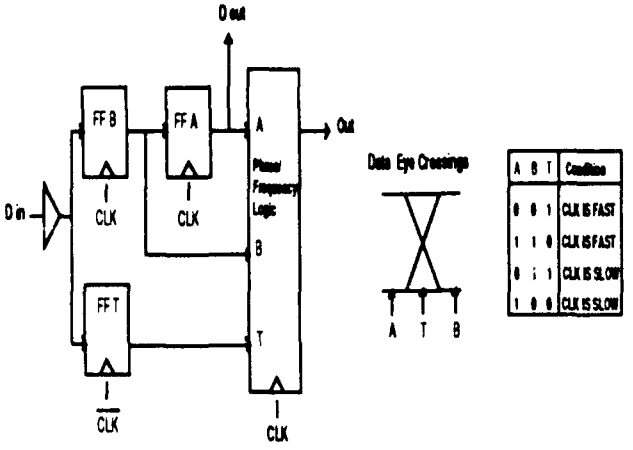


Phase/frequency detector

- Samples incoming data
- Provides digital error signal to speed up or slow down the VCO
- Provides recovered clock and retimed data
- Recovered clock is output by VCO.

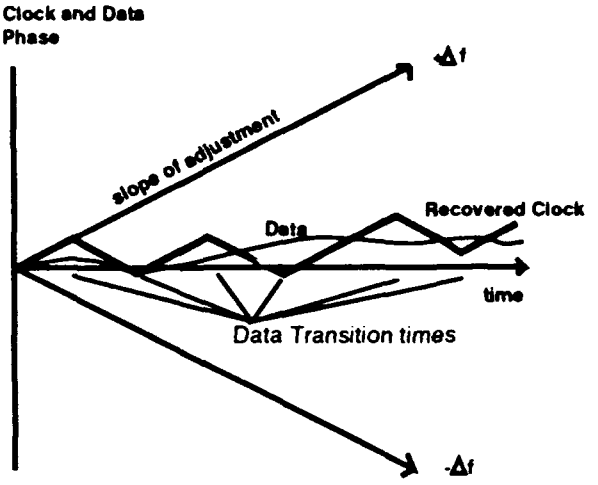
CH2965-2/91/0000-712 $1.00 © 1991 IEEE

Phase/frequency detector operation



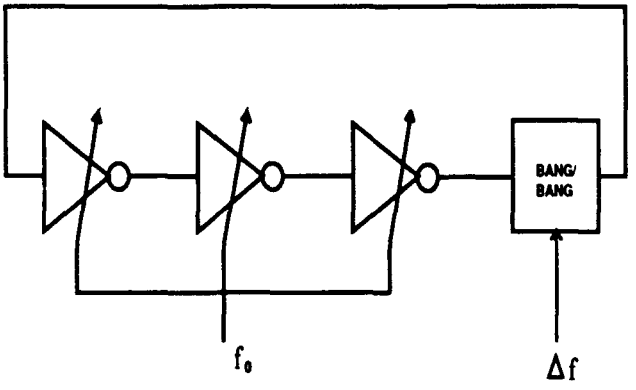
- Previous sample of Din is clocked to FFT (memory cell) and provides retimed data out.
- FFA and FFB contain the current sample of Din with $\overline{CLK}$ and CLK respectively.
- Relative positions of A, T, B provide lookup table information determining slow down or speed up of clock.

Dynamic Clock/Data Behavior



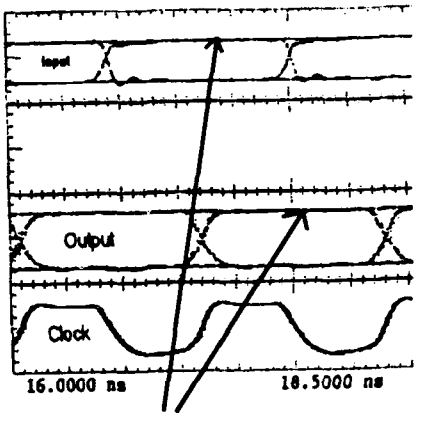
- Recovered clock phase tracks phase variations of incoming data.
- Clock phase continues to change between data transitions and dynamically tracks the phase of the incoming data.

VCO (Ring Oscillator)



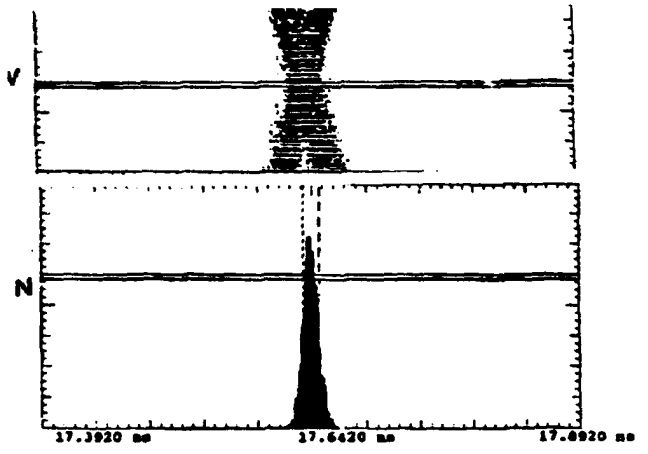
- Error signal is integrated to provide tracking of the frequency, f_0 , to the data rate.
- Error signal directly adjusts small bang/bang time delay providing tracking of the data phase.
- Total loop delay is approximately 800 ps re 622 MHz.
- Bang/bang adjustment time is approximately 3 ps re 622 MHz.
- Settling time after bang/bang is approximately 700 ps re 622 MHz.

Typical Recovered Data Output



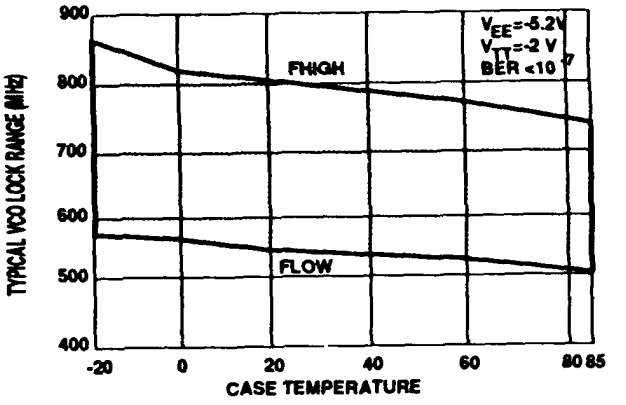
- One half bit time shift of output data and recovered clock relative to input data.
- Output data is phase locked to recovered clock.

Jitter Histogram of the Regenerated Data



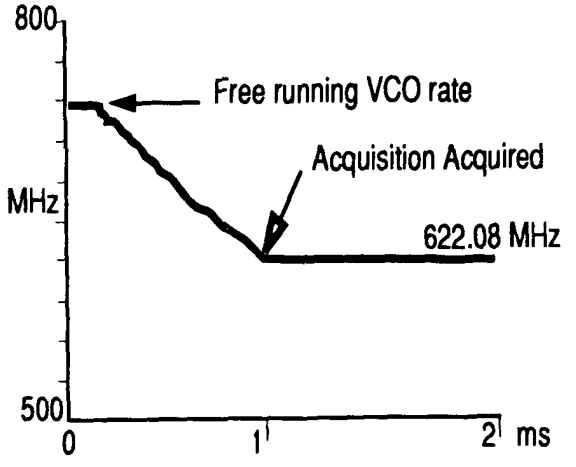
- Histogram of data eye provides measurement of jitter generation
- Measured jitter generation is 7.5 ps

VCO Range with Temperature



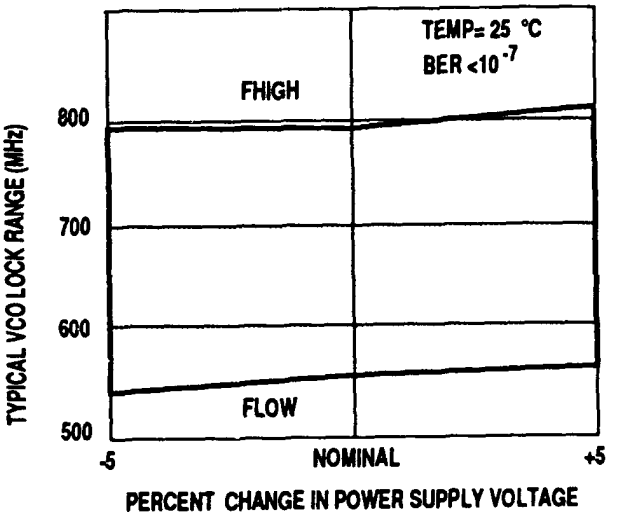
- Usable temperature range will vary with process variations

Acquisition Time of Retiming Circuit



- Acquisition time is less than 2 ms with 39 nF external integration capacitors with 622.08 input rate.
- The circuit corrects at each data transition toward the incoming data rate.

VCO Range with Supply Voltage Variation



- Typical usable range of VCO variation with power supply voltage.

Applications of Digital Retiming Technology

- Digital fiber optic telephone systems
  SONET 622 Mbps retiming with 2^{23}-1 pseudorandom data

- Datacom systems
  High speed links with M B/ N B encoding
  Example: 600 Mb/s with 8B/10B
  Synchronization equipment

- Radio links employing pseudorandom data methods
  QAM systems
  Wireless data links

- In conjunction with secondary retiming technologies
  Lower rate Phase Locked Loop systems (demultiplexed)
  SAW based methods

Acknowledgments:

The authors would like to acknowledge the measurements made by Cynthia Zee, and the packaging provided by Jay Dekhordi.

References:

1. Lai, B. et al, "A Monolithic 622 Mb/s Clock Extraction Data Retiming Circuit", IEEE ISSCC91 Session TPM8.7, February 1991
2. Alexander, J. D. H., "Clock Recovery from Random Binary Signals", IEEE Electron Letters, vol 11, pp 541-542, October 1975

AUTHOR INDEX

Allan, D.W. 667
Andersen, J.W. 239
Anderson, A.C. 460
Asahara, J. 9
Aseev, P.I. 102
Auld, B.A. 222
Avramov, I.D 230

Babbit, R. 482
Bagwell, T.L. 452
Bahadur, H. 37
Ballato, A. 22,217,687
Bartholomew, T. 608
Beard, R. 608
Benjaminson, A. 393
Bergquist, J.C. 534
Besson, R.J. 431
Bezdelkin, V.V. 212
Bigler, E. 222
Blair, D.G. 495
Bloch, M. 330
Border, J.S. 594
Bradaczek, H. 114
Braymen, S.D. 207
Brendel, R. 341
Burch, J.F. 477
Burns, S.G. 207

Cadotte, R. 482
Capelle, B. 58,166
Capelle, N. 137
Carru, H. 166
Charuvy, I. 247
Chen, C.-C. 500
Clark, R.L. 352
Clayton, L.D. 309
Cochet-Muchy, D. 58
Cook, C. 640
Croitoru, N. 247
Cutler, L.S. 452,544

Daly, K.P. 477
DeMarchi, A. 544
Détaint, J. 58,137,166
Dick, G.J. 491,572
Diedrich, F. 534
Djian, F. 341
Dragonette, R.A. 586
Drakhlis, B.G. 447
Driscoll, M.M. 700
Dull, D.B. 554
Dutka, J. 184
Dworsky, L. 148

EerNisse, E.P. 254,309
Egan, W.F. 629
Elsner, F. 534
Eltsufin, E. 567
Erasmus, J. 184

Fel, S. 567
Filler, R.L. 398
Flory, C.A. 544
Folen, V. 467
Frank, A. 467
Frymoyer, E.M. 712
Fujimoto, K. 261
Fujishima, S. 261

Gavaler, J.R. 700
Gevorkyan, A.G. 562,591
Giffard, R.P. 544
Gifford, A. 582,608
Golan, G. 247
Golding, W.M. 467
Gouzhva, Y.G. 591
Greenstein, J. 117
Greer, J.A. 321
Groslambert, J. 636
Grouzinenko, V.B. 102,212
Gualtieri, J.G. 22

Hamaguchi, K. 9
Haynes, J.T. 700
Hein, M. 467
Hirose, Y. 72
Hisadome, K. 513
Hollenhorst, J.N. 452
Horwitz, S.S. 700
Hu, R. 477

Issaev, V.A. 181
Itano, W.M. 534

Jackson, C.M. 477
Jelen, R.A. 700
Jespersen, J.L. 667
Joly, C. 58,166

Kawashima, H. 130,410
Kihara, M. 513
Kinsman, R. 148
Klipov, V.A. 29
Korniyenko, V.V. 591
Kosinski, J. 22,217
Koyama, M. 124
Kursinski, E.R. 594
Kusters, J.A. 554

Lai, B. 712
Lakin, K.M. 201
Larionov, I.M. 181
Lee, A. 477
Lee, P.C.Y. 156
Lewis, L.L. 521
Liang, X-P. 700
Lipeles, J. 117
Logan, Jr., R.T. 508,679
Long, B. 384
Lowe, J. 645
Lu, J-Q. 360
Lukaszek, T. 217,687
Lutes, G.F. 679
Lyons, W.G. 460

Maleki, L. 508,572,707
Mallikarjun, S. 217
Manasevit, H.M. 477
McGowan, R.C. 687,693
Meeker, T.R. 77
Mizan, M. 687,693
Mizuno, S. 124
Montress, G.K. 321
Moss, F. 649
Mourey, M. 431
Müller, G. 467
Mueller, L.F. 554
Murray, R.A. 2

Nagai, K. 9
Nakamura, T. 261
Nakazato, M. 130
Nemesh, A. 626
Newman, N. 452
Norton, J.R. 426

Oates, D.E. 460
Okazaki, M. 124
Olivier, M. 636
Opie, D.B. 467

Parker, T.E. 321
Parzen, B. 368
Peled, E. 626
Petite-Hall, C. 477
Phillips, D.E. 336
Philippot, E. 166
Piel, H. 467
Powers, E. 582
Prestage, J.D. 572

Rachlin, A. 482
Raizen, M.G. 534
Rasmussen, S. 608
Reinhart, M.J. 442
Rejent, J.A. 266
Ritz, E. 222
Robert, E. 341
Rodichev, A. 102
Rose, B. 393
Rose, D. 191
Rubiola, E. 636

Sakharov, S.A. 181
Sakuta, Y. 405
Sang, E. 222
Schneider, H.-P. 467
Schone, H.E. 467
Schreiter, S. 106
Schwartzel, J. 58,166
Seidman, A. 247
Sekine, Y. 405
Shadaram, M. 508
Sherman, Jr., J.H. 184
Shimizu, H. 418
Shmakov, N.N. 29
Smythe, R.C. 117,281
Sone, H. 9
Stebbins, S. 608
Stern, A. 567,659
Stewart, J.T. 137
St. John, D.C. 477
Sunaga, K. 410
Suter, J.J. 586

Taber, R.C. 452
Takemoto, J.H. 477
Taki, S. 9
Talvacchio, J. 700
Tiersten, H.F. 289,298
Tjoelker, R.L. 572
Tobar, M.E. 495
Tsuzuki, Y. 72,360

Ueno, Y. 418

Vahala, K.J. 539
Vianco, P.T. 266
Vig, J.R. 77
Vorokhovsky, Y.L. 447
Vulcan, A. 330

Wagner, G.R. 700
Walls, F.L. 645
Wang, H.T.M. 540
Wang, R.T. 491
Ward, R.W. 254
Watanabe, M. 405
Weaver, G. 384
Weber, R.J. 207,364
Weinert, R.W. 700
Weiss, M.A. 667
Wheeler, P. 582
Wilber, W. 482
Win, M.Z. 500
Wineland, D.J. 534
Wolf, S. 467

Yang, J.S. 156
Yankov, D.Y. 106
Yong, Y-K. 137

Zaki, K.A. 700
Zarka, A. 58,137,166
Zelitzki, M. 626
Zheng, Y. 58,137,166
Zholnerov, V.S. 562
Zhou, Y.S. ^39,298
Zilberstein, M. 247

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS (IEEE)

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1987 (ANSI/IEEE) Standard on Piezoelectricity (SH11270)

177-1966 Standard Definitions & Methods of Measurements for Piezoelectric Vibrators

180-1986 (ANSI/IEEE) Definitions of Primary Ferroelectric Crystal Terms (SH10553)

319-1971 (Reaff 1978) Piezomagnetic Nomenclature (SH02360)

1139-1988 Standard Definitions of Physical Quantities for Fundamental Frequency & Time Metrology (SH12526)

DEPARTMENT OF DEFENSE

Order from: Military Specifications and Standards
700 Robbins Ave., Bldg. 4D
Philadelphia, PA 19111-5094
(215) 697-2667/2179 - Customer Service
(215) 697-1187 to -1195 Telephone Order Entry System

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for

MIL-C-49468 Crystal Units, Quartz, Precision, Gen Spec for

MIL-C-24523 (SHIPS), Chronometer, Quartz Crystal

MIL-F-15733 Filters & Capacitors, Radio Interference, Gen Spec for

MIL-F-18327 Filters, High Pas, Band Pass Suppression and Dual Processing, Gen Spec for

MIL-F-28861 Filters and Capacitors, Radio Frequency/Electromagnetic Interference Suppression, Gen Spec for

MIL-F-28811 Frequency Standard, Cesium Beam Tube

MIL-H-10056 Holders (Encl), Crystal Gen Spec for

MIL-O-55310 Oscillators, Crystal, Gen Spec for

MIL-O-39021 Oven, Crystal, Gen Spec for

MIL-S-49433 Surface Acoustic Wave Devices, Bandpass Filter

MIL-S-49433(ER) Surface Acoustic Wave Devices, Gen Spec for

MIL-STD-683 Crystal Units, Quartz/Holders, Crystal

MIL-STD-188-115 Interoperability & Performance Standards for Communications, Timing & Synchronization Subsystems

MIL-STD-1395 Filters & Networks, Selection & Use of

MIL-T-28816(EC) Time Frequency Standard, Disciplined, AN/URQ-23, Gen Spec for

MIL-W-46374D Watch wrist: General Purpose

MIL-W-87967 Watch wrist: Digital

GENERAL SERVICES ADMINISTRATION

Order from: Naval Publication & Form Center or General Services Administration Business Service Centers in major U.S. cities

FED-STD-1002 Time & Frequency Reference Information in Telecommunication Systems

ELECTRONIC INDUSTRIES ASSOCIATION

Order from: Electronic Industries Assoc.
2001 Eye Street, NW
Washington, DC 20006
(202) 457-4900

(a) Holders and Sockets

EIA-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder Types).

EIA-367, Dimensional & Electrical Characteristics Defining Receiver Type Sockets (including crystal sockets).

EIA-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

EIA-186-E, (All Sections), Standard Test Methods for Electronic Component Parts.

EIA-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

EIA/IS-17-A, Assessment of Outgoing Nonconforming Levels in Parts Per Million (PPM).

EIA-IS-18, Lot Acceptance Procedure for Verifying Compliance with the Specified Quality Level (SQL) in PPM.

(c) Application Information

EIA Components Bulletin No. CB6-A, Guide for the Use of Quartz Crystal Units for Frequency Control, Oct 1987.

(d) EIA-477, Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985).

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC)

Order from: American Nat'l. Standard Inst.
(ANSI), 1430 Broadway
New York, NY 10018
(212) 354-3300

IEC PUBLICATIONS

122: Quartz crystal units for frequency control and selection.

122-1 (1976) Part 1: Standard values and test conditions. Amendment No. 1 (1983).

122-2 (1983) Part 2: Guide to the use of quartz crystal
units for frequency control and selection.

122-3 (1977) Part 3: Standard outlines and pin connections. Amendment No. 1 (1984)

122-3A (1979) First supplement.

122-3B (1980) Second supplement.

122-3C (1981) Third supplement.

283 (1986) Methods for the measurement of frequency & equivalent resistance of unwanted resonances of filter crystal units.

302 (1969) Standard definitions & methods of measurment for piezoelectric vibrators operating over the frequency range up to 30 MHz.

314 (1970) Temperature control devices for quartz crystal units. Amendment No. 1 (1979)

314A (19) First supplement.

368: Piezoelectric Filters.

368-1 (1982) Part 1: General information, standard values and test conditions.

368A (1973) First supplement. Amendment No.1 (1977). Amendment No.2 (1982).

368B (1975) Second supplement.

368-2 Part 2: Guide to the use of piezoelectric filters.

368-2-1 (1988) Section One-Quartz crystal filters

368-3 (1979) Part 3: Standard outlines.

368-3A (1981) First supplement.

444: Measurement of quartz crystal unit parameters by zero phase technique in a π-network.

444-1 (1986) Part 1: Basic method for the measurement of resonance frequency and resonance resistance of quartz crystal units by zero phase technique in a π-network.

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

444-3 (1986) Part 3: Basic method for the measurement of two-terminal parameters of quartz crystal units up to 200 MHz by phase technique in a π-network with compensation of the parallel capacitance Co.

444-4 (1988) Part 4: Method for the measurement of the load resonance frequency fL, load resonance resistance RL and the calculation of other derived values of quartz crystal units, up to 30 MHz.

483 (1976) Guide to dynamic measurements of piezoelectric ceramics with high electromechanical coupling.

642 (1979) Piezoelectric ceramic resonators and resonator units for frequency control and selection. Chapter I: Standard values and conditions. Chapter II: Measuring and test conditions.

679: Quartz crystal controlled oscillators.

679-1 (1980) Part 1: General information, test conditions & methods.

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators. Amendment No. 1 (1985).

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.

758 (1983) Synthetic quartz crystal. Chapter I: Specification for synthetic quartz crystal. Chapter II: Guide to the use of synthetic quartz crystal. Amendment No. 1 (1984).

862: Surface acoustic wave (SAW) filters.

862-1-1 (1985) Part 1: General information, test conditions and methods.

862-3 (1986) Part 3: Standard outlines.

CONSULTATIVE COMMITTEE ON INTERNATIONAL RADIO (CCIR)

Order from: International Telecommunications Union
General Secretariat - Sales Section
Place des Nations
CH-1211 Geneva
SWITZERLAND

Ask for CCIR 17th Plenary Assembly, Volume VII, "Standard Frequencies and Time Signals (Study Group 7)", which contains all of the following documents:

RECOMMEN)ATION 457-1	Use of the Modified Julian Date by the Standard-Frequency and Time-Signal Services
RECOMMENDATION 458-1	International Comparisons of Atomic Time Scales
RECOMMENDATION 460-4	Standard Frequency and Time Signal Emissions [Note: defines the UTC system]
RECOMMENDATION 485-1	Use of Time Scales in the Field of Standard-Frequency and Time Services
RECOMMENDATION 486-1	Reference of Precisely Controlled Frequency Generators and Emissions to the International Atomic Time Scale
RECOMMENDATION 535-1	Use of the Term UTC
RECOMMENDATION 536	Time Scale Notations
RECOMMENDATION 538	Frequency and Phase Stability Measures
REPORT 580-2	Characterization of Frequency and Phase Noise
RECOMMENDATION 583	Time Codes
RECOMMENDATION 685	International Synchronization of UTC Time Scales
RECOMMENDATION 686	Glossary

Annual Symposium on Frequency Control
PROCEEDINGS ORDERING INFORMATION

NO.	YEAR	DOCUMENT NO.	OBTAIN FROM*	PRICE MICROFICHE / HARDCOPY
10	1956	AD-298322	NTIS	$8.00 / $53.00
11	1957	AD-298323	NTIS	8.00 / 60.00
12	1958	AD-298324	NTIS	8.00 / 60.00
13	1959	AD-298325	NTIS	8.00 / 67.00
14	1960	AD-246500	NTIS	8.00 / 45.00
15	1961	AD-265455	NTIS	8.00 / 39.00
16	1962	AD-285086	NTIS	8.00 / 45.00
17	1963	AD-423381	NTIS	8.00 / 60.00
18	1964	AD-450341	NTIS	8.00 / 60.00
19	1965	AD-471229	NTIS	8.00 / 60.00
20	1966	AD-800523	NTIS	8.00 / 60.00
21	1967	AD-659792	NTIS	8.00 / 53.00
22	1968	AD-844911	NTIS	8.00 / 60.00
23	1969	AD-746209	NTIS	8.00 / 39.00
24	1970	AD-746210	NTIS	8.00 / 39.00
25	1971	AD-746211	NTIS	8.00 / 39.00
26	1972	AD-771043	NTIS	8.00 / 39.00
27	1973	AD-771042	NTIS	8.00 / 45.00
28	1974	AD-A011113	NTIS	8.00 / 45.00
29	1975	AD-A017466	NTIS	8.00 / 45.00
30	1976	AD-A046089	NTIS	8.00 / 53.00
31	1977	AD-A088221	NTIS	8.00 / 60.00
32	1978	AD-A955718	NTIS	8.00 / 53.00
33	1979	AD-A213544	NTIS	8.00 / 67.00
34	1980	AD-A213670	NTIS	8.00 / 53.00
35	1981	AD-A110870	NTIS	16.50 / 67.00
36	1982	AD-A130811	NTIS	8.00 / 53.00
37	1983	AD-A136673	NTIS	8.00 / 53.00
38	1984	AD-A217381	NTIS	11.00 / 39.00
39	1985	AD-A217404	NTIS	17.00 / 53.00
40	1986	AD-A235435	NTIS	15.00 / 53.00
41	1987	AD-A216858	NTIS	15.00 / 53.00
42	1988	AD-A217275	NTIS	15.00 / 53.00
43	1989	89CH2690-6	IEEE	68.00 / 68.00
44	1990	90CH2818-3	IEEE	68.00 / 68.00
45	1991	91CH2965-2	IEEE	68.00 / 68.00

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161, U. S. A.
Tel: 703-487-4650

*IEEE - Inst. of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854, U.S.A.
Tel: 800-678-4333 or 201-981-0060

An index to all the papers in the Proceedings since 1956, and other Symposium information are available as ASCII files on a 5 1/4" (13 cm) MS-DOS format floppy disk, for $5-, from Synergistic Management, Inc., 3100 Route 138, Wall Township, NJ 07719.