

EFTF 91

Cinquième Forum Européen Temps-Fréquence
Fifth European Frequency and Time Forum

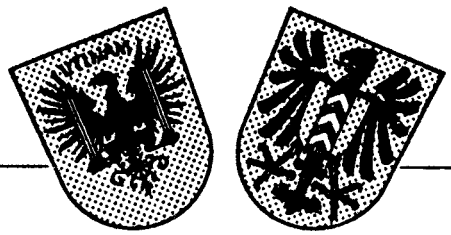
ACTES

PROCEEDINGS

12,13, 14 MARS 1991

12, 13, 14 MARCH 1991

BESANCON
FRANCE



EFTF 91

Cinquième Forum Européen Temps-Fréquence
Fifth European Frequency and Time Forum

ACTES

PROCEEDINGS

12, 13, 14 MARS 1991
12, 13, 14 MARCH 1991

BESANCON
FRANCE

5e FORUM EUROPEEN

TEMPS-FREQUENCE

5th EUROPEAN FREQUENCY

AND TIME FORUM

Organisé par les Communautés Scientifiques de Besançon (FRANCE) et de Neuchâtel (SUISSE)

Organised by the Scientific Communities of Besançon (FRANCE) and Neuchâtel (SWITZERLAND)

12, 13, 14 mars 1991

12, 13, 14 march 1991

Le KURSAAL

Place Granvelle

25000 - BESANCON (FRANCE)

Ces actes ont été imprimés avec l'aide du CONSEIL GENERAL DU DOUBS et par les soins de son imprimerie.

These proceedings have been printed by courtesy of "CONSEIL GENERAL DU DOUBS" in its printing office.

5e FORUM EUROPEEN TEMPS-FREQUENCE
5th EUROPEAN FREQUENCY AND TIME FORUM

Président - Chairman ..... R.J. BESSON
Secrétaire Général - General Secretary ..... A. REMOND
Organisation/Secrétariat - Organization/Secretariat ..... F. BARONCHELLI

E.N.S.M.M. - L.C.E.P., route de Gray, La Bouloie, 25030 BESANCON CEDEX, FRANCE
Tél : 81 66 66 32 - 81 66 66 27 - Fax : 81 88 57 14

COMITE EXECUTIF - EXECUTIVE COMMITTEE

R. BESSON
Laboratoire de Chronométrie, Electronique et Piézoélectricité, Ecole Nationale Supérieure de Mécanique et des Microtechniques, Besançon, France.

M. ECABERT
Fondation Suisse pour la Recherche en Microtechnique, Neuchâtel, Switzerland.

J.J. GAGNEPAIN
Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, Besançon, France.

P. KARTASCHOFF
Section Recherche et Développement de la Direction Générale des PTT, Bern, Switzerland.

A. REMOND (E.R.)
Société Française de Microtechnique et de Chronométrie, Besançon, France

B. SCHLUETER
Keres Engineering Inc., Neuchâtel, Switzerland.

Avec le soutien :

Ministère de la Recherche et de la Technologie, Région de Franche-Comté, Conseil Général du Doubs, Ville de Besançon

With support of :

Société Française de Microtechnique et de Chronométrie, Ecole Nationale Supérieure de Mécanique et des Microtechniques, Observatoire de Besançon, et Laboratoires : Laboratoire de Chronométrie, Electronique et Piézoélectricité (ENSMM) et Laboratoire de Physique et Métrologie des Oscillateurs (CNRS)

COMITE SCIENTIFIQUE - SCIENTIFIC COMMITTEE

AUDOIN C.	CNRS - LHA, Orsay	France
BALLATO A.	US Army Electronics Techn. & Devices Lab., Fort Monmouth	USA
BESSON R.	ENSMM - LCEP, Besançon	France
BOLOIX R.	Observatoire de San Fernando (ROA), San Fernando	Spain
BRUNET M.	CNES, Toulouse	France
BUSCA G.	Observatoire Cantonal, Neuchâtel	Switzerland
CUTLER L.	Hewlett-Packard Co, Palo Alto	USA
DE BOER H.	Physikalisch-Technische Bundesanstalt Braunschweig	FRG
DE MARCHI A.	Universita di Ancona, Ancona	Italy
DEYZAC F.	ONERA, Chatillon	France
DOWSETT J.	STC Components LTD, Harlow Essex	UK
DUCHAUSSOY H.	DRET, Paris	France
GAGNEPAIN J.J.	CNRS - LPMO, Besançon	France
GRAF E.	Oscilloquartz SA, Neuchâtel	Switzerland
GRANVEAUD M.	Observatoire de Paris, Paris	France
GUINOT B.	Bureau National des Poids et Mesures, Sèvres	France
HANDREK Horst	Quarzkeramik GmbH, Stockdorf	FRG
HARTL Ph.	Universität Stuttgart, Stuttgart	FRG
HENAFF J.	CNET - DIT - ASC, Issy-les-Moulineaux	France
HRUSKA C.	York University, Downsview	Canada
KALLIOMAKI K.	University of Oulu, Oulu	Finland
KARTASCHOFF P.	PTT, Bern	Switzerland
KIRCHNER D.	Technische Universität Graz	Austria
KNIGHT D.	National Physical Laboratory, Teddington	UK
KREMPL P.	AVL List GmbH, Graz	Austria
LESCHIUTTA S.	Scuole Polytechnico, Torino	Italy
MAERFELD C.	Thomson-Sintra-DTAS, Valbonne	France
NAKAGIRI K.	Radio Research Laboratory, Tokyo	Japan
NARD G.	Sercel, Carquefou	France
NEUBIG B.	Tele Quartz GmbH, Neckarbischofsheim	FRG
NIEUWEJAAR P.	Norwegian Naval Materiel Command, Haakonsvern	Norway
NOTTARP K.	Institut für Angewandte Geodaesie, Frankfurt-am-Main	FRG
OEHMAN B.	Swedish Telecom Radio, Haninge	Sweden
PROST L.	Office Fédéral de Métrologie, Wabern	Switzerland
RUTMAN J.	FIEE, Paris	France
STARKER S.	DLR, Wessling	FRG
STEELE J.	National Physical Laboratory, Teddington	UK
THOMANN P.	Oscilloquartz SA, Neuchâtel	Switzerland
UNVERDROSS R.	Unverdross Technik, Starnberg	FRG
VANIER J.	National Research Council, Ottawa	Canada
VON WILLISEN F.K.	MDC Max Dätwyler AG, Bleienbach	Switzerland
WALLS F.	National Inst. of Standards & Technology, Boulder	USA
WINKLER G.	US Naval Observatory, Washington	USA
YANNONI N.	Rome Air Development Center, Bedford	USA
ZINGG W.	Micro Crystal, Grenchen	Switzerland

EXPOSANTS - EXHIBITORS

C.E.P.E., CIE D'ELECTRONIQUE ET DE PIEZO-ELECTRICITE	44, avenue de la Glacière	F-95105 ARGENTEUIL
KINEMETRICS/TRUE TIME	3243 Santa Rosa Avenue	USA-SANTA ROSA, CA 95407
Q K PIEZOELECTRONIQUE	30, Boulevard Galliéni	F-92390 VILLENEUVE LA GARENNE
RACAL-DANA	18, avenue Dutartre	F-78150 LE CHESNAY

PRESIDENTS DE SEANCE - SESSION CHAIRMEN

Sessions

David W. ALLAN - Séance d'ouverture / Opening session
National Institute of Standards and Technology, Boulder, Colorado, USA

Claude AUDOIN - Horloges Atomiques I / Atomic Clocks I — 1A
LHA-CNRS, Bâtiment 221, Université Paris-Sud, F-91405 Orsay

Hugues DUCHAUSSOY - Résonateurs I / Resonators I — 2A
DRET, 26 boulevard Victor, F-00460 Paris-Armées

Daniel HAUDEN - Ondes de Surface / Saw Devices — 3A
LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon

Michel BRUNET - Temps-Fréquence I / Time and Frequency I — 2B
CNES, 18 avenue E. Belin, 31055 Toulouse

Michel GRANVEAUD - Temps-Fréquence II / Time and Frequency II — 3B
LPTF - OBSERVATOIRE DE PARIS, 61avenue de l'Observatoire, F-75014 Paris

Albert ZARKA - Matériaux / Materials — 4A
LABORATOIRE DE MINERALOGIE CRISTALLOGRAPHIE
Université Pierre et Marie Curie, 4, place Jussieu, F-75252 Paris Cedex 05

Jacques DETAINT - Résonateurs II / Resonators II — 5A
CNET, 196 avenue Henri Ravera, F-92220 Bagneux

Nickolaï KOSHELYAEVSKY - Horloges Atomiques II / Atomic Clocks II — 4B
VNIIFTRI, Time and Frequency Service, USRR-41570 Moscow Region

Giovanni BUSCA - Temps-Fréquence III / Time and Frequency III — 5B
OBSERVATOIRE CANTONAL, rue de l'Observatoire 58, CH-2000 Neuchâtel

Philippe DEFRANOULD - UHF et Ondes de Surfaces / UHF and Saw Devices — 6A
THOMSON SINTRA ASM DTAS, 399 route des Crêtes, BP 138, 06561 Valbonne Cedex

Jean-Jacques GAGNEPAIN - Oscillateurs I / Oscillators I — 6B
LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon

Wolfgang PFLAUM - Résonateurs III / Resonators III — 7A
BALL CORP. EFRATOM, Irvine, California, U.S.A.-92718-1605

Marcel OLIVIER - Oscillateurs II / Oscillators II — 8A
LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon

Andreas BAUCH - Horloges Atomiques III / Atomic Clocks III — 7B
PHYSIKALISCH-TECHNISCHE BUNDESANSTALT
Bundesallee 100, D-3300 Braunschweig

Sigfrido LESCHIUTTA - Temps-Fréquence IV / Time and Frequency IV — 8B
SCUOLE POLITECHNICO, I-10129 Torino

EXPOSANTS - EXHIBITORS

C.E.P.E., CIE D'ELECTRONIQUE ET DE PIEZO-ELECTRICITE	44, avenue de la Glacière	F-95105 ARGENTEUIL
KINEMETRICS/TRUE TIME	3243 Santa Rosa Avenue	USA-SANTA ROSA, CA 95407
9 K PIEZOELECTRONIQUE	30, Boulevard Galliéni	F-92390 VILLENEUVE LA GARENNE
RACAL-DANA	18, avenue Dutartre	F-78150 LE CHESNAY

PRESIDENTS DE SEANCE - SESSION CHAIRMEN

	Sessions
David W. ALLAN - Séance d'ouverture / Opening session National Institute of Standards and Technology, Boulder, Colorado, USA	
Claude AUDOIN - Horloges Atomiques I / Atomic Clocks I LHA-CNRS, Bâtiment 221, Université Paris-Sud, F-91405 Orsay	1A
Hugues DUCHAUSSOY - Résonateurs I / Resonators I DRET, 26 boulevard Victor, F-00460 Paris-Armées	2A
Daniel HAUDEN - Ondes de Surface / Saw Devices LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon	3A
Michel BRUNET - Temps-Fréquence I / Time and Frequency I CNES, 18 avenue E. Belin, 31055 Toulouse	2B
Michel GRANVEAUD - Temps-Fréquence II / Time and Frequency II LPTF - OBSERVATOIRE DE PARIS, 61avenue de l'Observatoire, F-75014 Paris	3B
Albert ZARKA - Matériaux / Materials LABORATOIRE DE MINERALOGIE CRISTALLOGRAPHIE Université Pierre et Marie Curie, 4, place Jussieu, F-75252 Paris Cedex 05	4A
Jacques DETAINT - Résonateurs II / Resonators II CNET, 196 avenue Henri Ravera, F-92220 Bagneux	5A
Nickolaï KOSHELYAEVSKY - Horloges Atomiques II / Atomic Clocks II VNIIFTRI, Time and Frequency Service, USRR-41570 Moscow Region	4B
Giovanni BUSCA - Temps-Fréquence III / Time and Frequency III OBSERVATOIRE CANTONAL, rue de l'Observatoire 58, CH-2000 Neuchâtel	5B
Philippe DEFRANOULD - UHF et Ondes de Surfaces / UHF and Saw Devices THOMSON SINTRA ASM DTAS, 399 route des Crêtes, BP 138, 06561 Valbonne Cedex	6A
Jean-Jacques GAGNEPAIN - Oscillateurs I / Oscillators I LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon	6B
Wolfgang PFLAUM - Résonateurs III / Resonators III BALL CORP. EFRATOM, Irvine, California, U.S.A.-92718-1605	7A
Marcel OLIVIER - Oscillateurs II / Oscillators II LPMO - CNRS, 32 avenue de l'Observatoire, F-25000 Besançon	8A
Andreas BAUCH - Horloges Atomiques III / Atomic Clocks III PHYSIKALISCH-TECHNISCHE BUNDESANSTALT Bundesallee 100, D-3300 Braunschweig	7B
Sigfrido LESCHIUTTA - Temps-Fréquence IV / Time and Frequency IV SCUOLE POLITECHNICO, I-10129 Torino	8B

TABLE DES MATIERES - TABLE OF CONTENTS

INTRODUCTION

Executive Committee, Support iii
Scientific Committee iv
Exhibitors v
Session Chairmen vi

Opening session

1. **Time and frequency metrology : current status and future considerations** 1
D.W. Allan - NIST, Boulder, Colorado, USA

Session 1A Plenary
ATOMIC CLOCKS I

2. **Frequency Standard at 3.39 μm based on linear optical ramsey resonance in a methane molecular beam◆** 10
G. Kramer - PTB, Braunschweig, Germany

3. **Frequency shifts in cesium clocks due to majorana transitions : Towards a theoretical understanding** 11
A. Bauch, R. Schröder - PTB, Braunschweig, Germany

4. **Recent progresses in the Mg beam frequency standard** 18
A. Godone, C. Novero - Istituto Elettrotecnico Nazional "Galileo Ferraris", Torino, Italy

5. **Determining the effects of microwave power variations on the output frequency of a high-performance cesium frequency standard** 21
S.K. Karuza, W.A. Johnson, J.P. Hurrell, F.J. Voit and A. Leong - Communications Electronics Dept., The Aerospace Corporation, Los Angeles, CA, USA

Session 2A
RESONATORS I

6. **A new approach to modeling quartz resonator and oscillator acceleration sensitivity** 25
J.A. Kosinski, A. Ballato*, T.J. Lukaszek*, S. Mallikarjun***
* U.S. Army Labcom Elect. Tech. & Devices Lab. , Ft. Monmouth, N.J., USA
** Dept. of Chemistry and Physics, Monmouth College, N.J., USA

◆ Abstracts only

7. **Optimization of the design of thickness shear devices using the AT cut of berlinite** 30
*J. Détaint, H. Carru, J. S. Schwartzel, C. Joly, E. Philippot**
CNET, Bagneux, France
* Univ. de Montpellier, Lab. Physico-Chimie des Solides, Montpellier, France

8. **Observation of defects and vibration modes in lithium tantalate** 41
B. Capelle, J. Détaint**, J. Schwartzel**, A. Zarka*, Y. Zheng*, C. Joly***
* LMC, Paris, France
** CNET, Bagneux, France

9. **Sensitivity of baw devices to radial in-plane stress distribution, comparison between analytical and finite element results** 50
S. Ballandras - LPMO - CNRS, Besançon, France

Session 3A
SAW DEVICES

10. **Saw filters for mobile radio◆** 55
P. Defranould - Thomson Sintra Activités Sous-Marines, Valbonne, France

11. **Low loss saw filters using multiple transducers** 56
B.P. Abbott , J. Heighway, D.P. Morgan & T. Thorvaldsson
ASCOM Microsystems, Bevaix, Switzerland

12. **Saw sensitivity to stresses and temperature gradients. New theoretical and experimental results** 62
S. Ballandras, E. Bigler - LPMO - CNRS, Besançon, France

13. **Propagation of piezoelectric waves in simply- and doubly- periodic metallic gratings** 66
A.R. Baghai-Wadji, Ch. Mecklenbräuker - Techn. Univ. Wien, Inst. für Allegemeine Elektrotechnik & Elektronik, Austria

Session 2B
TIME AND FREQUENCY I

14. **Time transfer by passive use of direct broadcasting satellite signals** 74
J.A. Davis, P.R. Pearce, P. Menary, M. Hutchins**, M. Sagin****
NPL, Middlesex, United Kingdom
* BTI, Rugby, UK
** HP Customer Support Centre, Winnersh, UK
*** Radiocode Clocks Ltd, Cornwall, UK

15. **Time transfer using a geostationary direct TV satellite** 82
F. Meyer, M. Brunet**, M. Granveaud***, F. Vernotte*, M. Vincent**
* Observatoire de Besançon, France
** CNES, Toulouse, France
*** LPTF, Observatoire de Paris, France

16. **A direct sequence spread-spectrum modem design for time transfers via geostationary satellites** 89
D.D. Davis, D.A. Howe,
National Inst. of Standards and Technology, Boulder, USA

17. **Preliminary comparison of time transfers via Lasso, GPS and two-way satellite** 96
P. Uhrich, R. Tourde, M. Granveaud
LPTF, Paris, France
P. Grudler, F. Baumont, Ch. Veillet, D. Féraudy, J.M. Torre, J.F. Mangin, J. Gaignebet, J.L. Hatat
Observatoire de la Côte d'Azur, Grasse, France
W. Hanson, A. Clements, J. Jespersen, M. Lombardi
National Inst. of Standards and Technology, Boulder, USA
D. Kirchner, U. Thyr
Technical University, Graz, Austria
G. Kirchner, H. Ressler
Institut für Weltraumforschung, Graz, Austria
W.J. Klepczynski, P.J. Wheeler, W. Powell, A. Davis
US Naval Observatory, Washington, USA
W. Lewandowski
Bureau International des Poids et Mesures, Paris, France

Session 3B
TIME AND FREQUENCY II

18. **Very fast direct digital synthesizer with DRO time base for instrumentation applications** 105
M.T. Diep, J.F. Zürcher*, F. Gardiol*, P. Denervaud***
* Laboratoire d'Electromagnétisme et d'Acoustique, Lausanne, Switzerland
** Centre Suisse d'Electronique et de Microtech., Neuchâtel, Switzerland

19. **An ultra low phase noise synthesizer for radar applications** 110
B. Fognini
Siemens Albis AG, Zurich, Switzerland

20. **Phase noise of regenerative frequency dividers** 115
E. Rubiola, M. Olivier**, J. Groslambert***
* Dipt. di Elletronica-Politecnico, Torino, Italy
** LPMO - CNRS, Besançon, France

21. **A narrow band frequency synthesizer with a NCO circuit** 123
P. Schumacher
Observatoire Cantonal de Neuchâtel, Switzerland

Session 4A

MATERIALS

22. **Synthetic quartz material characterizations. Connexions with the quality of AT-cut resonator** 129
J.J. Boy, P. Zecchini**, B. Viard***
* LCEP/ENSMM, Besançon, France
** LCCM, Besançon, France

23. **Insitu X-ray topographic study of sweeping in quartz crystals◆** 136
M.T. Sebastian, H. Klapper***
* Reg. Research Lab., Trivandrum, India
** Inst. für Kristallographie, Aachen, Germany

24. **The contribution of the transit-time and the resonator method. Towards determination of the nonlinearities in quartz** 137
C.K. Hruska
Piezoelectricity Research Laboratory, Ontario, Canada

25. **Present state of $GaPO_4$ - Research** 143
P. Krempl, J. Stadler, W. Wallnöfer, W. Ellmeyer, R. Selic
AVL LIST Ges.m.b.H., Physics Division, Graz, Austria

26. **Electrodiffusion (sweeping) and irradiation effects in crystalline quartz** 148
H. Bahadur
NPL, New Delhi, India

Session 5A

RESONATORS II

27. **On the influence of a fabrication imperfection on the in-plane acceleration sensitivity of contoured quartz resonators with rectangular supports** 157
H.F. Tiersten, Y.S. Zhou
Rensselear Polytechnic Inst., Troy, NY, USA

28. **A general model for quasi-thickness vibrations in contoured quartz plates** 164
B. Dulmet
LCEP/ENSMM, Besançon, France

29. **Enhanced-selectivity, intermediate-bandwidth, crystal filters** 173
P.K. Wall
Peter Wall Electroacoustics, UK

Session 4B

ATOMIC CLOCKS II

30. **Incoherent and coherent optical pumping in Rubidium standards** 181
P. Willemin, G. Busca, J.T. Liu, P. Thomann**
Observatoire Cantonal de Neuchâtel, Switzerland
* Oscilloquartz SA, Neuchâtel, Switzerland

31. **Experimental study of two-laser. Optical pumping in a cesium clock** 187
N. Dimarcq, V. Giordano, C. Jacques, G. Théobald, P. Cerez*
LHA, Orsay, France
* Inst. for National Measurements Standards, Ottawa, Canada

32. **Laser cooling of Cs atomic beam at the NRLM** 190
T. Ikegami, S.I. Ohshima, Y. Nakadan
National Research Lab. of Metrology, Ibaraki, Japan

33. **Frequency synthesis in primary cesium clocks** 194
R. Schröder
Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

34. **The optically pumped cesium beam frequency standard at Peking University** 201
Y-Q Wang, D-H Yang***
* Depart. of Radio-Electronics, Peking University, Beijing, P.R. China
** Beijing Institute of Modern Physics, Peking University

Session 5B

TIME AND FREQUENCY III

35. **Atomic frequency standards for different systems and aids of Positioning and Timing Service : state-of-the art of developments and prospects for international cooperation** 205
Y.G. Kouzhva, A.G. Gevorkyan, A.M. Moskovsky
Leningrad Scientific Research Radiotechnical Institute, Leningrad, USSR

36. **ESA center of excellence in time and frequency** 207
S.J. Feltham
European Space Research and Technology Centre, Noordwijk, The Netherlands

37. **SA and control of local oscillator in GPS receivers** 211
K. Kalliomäki, T. Mansten***
* Univ. of Oulu, Dept. of Electrical Engin., Oulu, Finland
** Technical Research Centre of Finland, Espoo, Finland

38. **GPS time closure around the world using precise ephemerides, ionosphere measurements and accurate antenna coordinates** 215
W. Lewandowski, G. Petit*, C. Thomas*, M.A. Weiss***
* BIPM, Sèvres, France
** Nat. Inst. of Standards and Technology, Boulder, USA

Sessions A et B

POSTERS SESSION

39. **Determination of a complete fully consistent data set of the fundamental properties of high purity syntheric quartz** 221
B.J. James
GEC-Marconi Limited, Hirst Research Centre, Middlesex, UK

40. **An alternative and economic conception for digital compensated crystal oscillators** 224
W. Balzer
KVG GmbH, Neckarbischofsheim, Germany

41. **A laser cooled cesium atomic fountain : Towards a high performance clock** 228
A. Clairon, Ch. Salomon**, S. Guellati***, W.D. Phillips*****
*LPTF, Paris, France
** Lab. de Spectroscopie Hertzienne, Ecole Normale Sup., Paris, France
*** Lab. Aimé Cotton, Orsay, France
**** NIST, Gaithersburg, MD, USA

42. **Environmental sensitivity and long term stability limitations induced by C-field variations in commercial cesium beam frequency standards** 237
A. De Marchi, E. Rubiola
Dip. Elettronica, Politecnico di Torino, Italy

43. **Majorana transitions : their effect on the velocity distributions of cesium atoms in an atomic beam** 243
A. Bauch, T. Heindorff
Physikalisch-Technische Bundesanstalt, Braunschweig, Germany

44. **The expansion of optical locking range of a spectral-narrowed laser diode system using a current feedback technique** 249
*T. Maeno, H. Saito, J. Umezu, Y. Ohta, R. Hayashi, P. Thomann**
Communications Research Lab., Tokyo, Japan
* Observatoire de Neauchâtel, Switzerland

45. **A new kind of hydrogen-maser time and frequency standard for chinese VLBI network** 252
Z.C. Zhai, C.F. Lin, H.X. Huang, J.W. He, W.Q. Zhang
Shanghai Observatory, Academia Sinica, China

46. **TF-4010A, TFL's compact rubidium frequency standard** 255
A. Stern, Y. Eltzufin
TFL, Holon, Israel

47. **Computer controlled short term frequency stability measurement system with the application of covariance method** 259
B. Gniewinska
Tele and Radio Research Inst., Warsawa, Poland

48. **Systems of convolution equations, deconvolution and Gabor and Wavelet signal analysis** 262
S.D. Casey
Dept. of Mathematics and Statistics, American Univ., Washington, USA

49. **Realization of the belgian time scale** 275
F. Collin, P. Pâquet
Royal Observatory of Belgium, Brussels, Belgium

50. **Comparison of RNS -and SYLT- Lorans with GPS as a reference** 281
T. Mansten(1), K. Kalliomäki(2), B. Selin(3)
(1) Technical Research Centre of Finland (MTI), Espoo, Finland
(2) University of Oulu, Finland
(3) Swedish Telecom Radio, Haninge, Sweden

51. **European computer time service** 284
E. Brännström
Swedish Telecom Radio, Haninge, Sweden

52. **Optical pumping of the Rubidium maser by the use of a frequency-modulated laser** 288
P. Tremblay, A. Michaud, J. Beaubien, E. Cauchon*
Centre d'Optique Photonique et Laser, Université Laval, Québec, Canada
* BIPM, Sèvres, France

53. **Portable atomic clock on the basis of an active hydrogen maser, "Sapphir"** 293
B.A. Gaygerov, V.P. Sysoev, F.S. Rusin
NPO VNIIFTRI, Mendeleevo, Moscow Region, USSR

54. **Hydrogen time and frequency standard "Sphere"** 296
A.N. Boyko, B.A. Gaygerov, G.A. Yelkin, N.D. Jestkova, G.N. Kurnikov, V.P. Sysoev
NPO VNIIFTRI, Mendeleevo, Moscow Region, USSR

55. **State of the art polishing and lapping for VHF quartz blanks** 300
L. Morawiec
Quartz Works, USA

Session 6A

UHF AND SAW DEVICES

56. **Dielectric resonator whispering gallery modes. Application to millimeter wavelength oscillators** 304
D. Cros, P. Guillon
Inst. de Recherche en Communications Optiques et Microondes, Limoges, France

57. **NO_2 - Surface acoustic wave Gas sensor** 309
M. Hoummady, D. Rebière, D. Hauden, J. Pistré*, P. Cunin** and R. Planade***
LPMO - CNRS, Besançon, France
* IXL, Univ. Bordeaux I, Talence, France
** Centre d'Etudes du Bouchet, Vert le Petit, France

58. **Quartz crystal rotation rate sensors** 314
J.J. Gagnepain, G. Marianneau*, L. Gaudriot***
* LPMO - CNRS, Besançon, France
** Sté Métravib, Ecully, France

59. **Visco-elasticity, phase transistors and adhesivity measurements by shear horizontal acoustic waves** 318
M. Hoummady, F. Bastien
LPMO - CNRS, Besançon, France

60. **The state-of-art electrostatic field calculation of surface acoustic wave devices** 323
A.R. Baghai-Wadji, H.P. Zidek, H.P. Reichinger, B.H. Jakoby
Technische Univ. Wien, Inst. für Allgemeine Electrotechnik und Elektronik, Autria

Session 6B
OSCILLATORS I

61. **Time and frequency in the tomorrow road traffic** 331
S. Leschiutta
Politecnico di Torino, Italy

62. **A novel, low-power, fast-warmup, miniature oscillator** 337
W. Pflaum, P. Naepflin, J. Fellner, W. Chen, J. Kusters, P. Stoermer
Ball Corporation, Efratom Division, California, USA

63. **Compact micro processor controlled DTCXO** 343
R. Olbert
Tele quartz GmbH, Neckarbischofsheim, Germany

64. **Design and application of an ultra stable oscillator for space applications** 347
E. Gérard, R. Molle
CEPE, Argenteuil, France

Session 7A
RESONATORS III

65. **New results concerning quartz and berlinite resonators obtained by synchrotron radiation X-ray topography** 358
A. Zarka, B. Capelle*, Y. Zheng*, J. Detaint**, J. Schwartzel**, E. Philippot****
* LMC Univ. P. et M. Curie, Paris, France
** CNET, Bagneux, France
*** LPCMI Univ. de Montpellier, France

66. **Quality asessment of quartz crystal units acc. to ISO 9000 with implementation of statistic process control** 365
L. Schulzke
KVG GmbH, R & D Crystals, Neckarbischofsheim, Germany

67. **New possibilities offered by silicon structures for resonators and sensors design◆** 371
N. de Rooij
Inst. de Microtechniques, Univ. de Neuchâtel, Switzerland

68. **Nonlinear structure of the phase motion in oscillators** 372
M. Planat
LPMO - CNRS, Besançon, France

Session 8A
OSCILLATORS II

69. **Nonlinear modelling techniques for quartz crystal oscillators** 377
R. Brendel, G. Marianneau*, E. Robert***
*LPMO - CNRS, Besançon, France
**CNES, Toulouse, France

70. **Crystal controlled oscillators in the UHF-range by PLL-technique** 385
B. Adler, B. Neubig
Tele quartz Groupe, Neckarbischofsheim, Germany

71. **Magnetic sensitivity of oscillator components** 390
R. Brendel, C. El Hassani*, P. Krux*, E. Robert***
LPMO - CNRS, Besançon, France
** CNES, Toulouse, France

72. **First measurements on ocir controlled ovens** 396
B. Abboud, S. Galliou, M. Mourey, J.P. Valentin
LCEP/ENSMM, Besançon, France

73. **A new way to control the phaselocked frequency standards** 401
J. Mannermaa, K. Kalliomäki
Univ. of Oulu, Dept. of Elect. Eng., Finland

Session 7B
ATOMIC CLOCKS III

74. **Recent progress in the development of the EFOS-B hydrogen maser at Neuchâtel Observatory** 404
L.G. Bernier, G. Busca
Observatoire Cantonal de Neuchâtel, Switzerland

75. **Fast cavity auto-tuning system for H-masers : experimental results** 407
R. Barillet, J. Viennet, P. Petit, C. Audoin
LHA, Orsay, France

76. **Preliminary investigations with the NIST optically pumped primary frequency standard** 412
R.E. Drullinger, J.H. Shirley, D.J. Glaze, J. Lowe, HS Lee, A.S. Zibrov
NIST, Boulder Colorado, USA

77. **H- masers in USSR** 415
N. Koshelyaevsky
NPO "VNIIFTRI", Mendeleevo, Moscow region, URSS

Session 8B - Chairman : S. Leschiutta

TIME AND FREQUENCY IV

78. **Study of the clock-ensemble correlation in algos using real data** 435
P. Tavella, C. Thomas**, J. Azoubib***
* IEN Galileo Ferraris, Torino, Italy
** BIPM, Sèvres, France

79. **Promise into practice : Implementing TA2 on real clocks at NIST** 442
M.A. Weiss, T. Weissert
Time and Frequency Div., NIST, Boulder, Colorado, USA

80. **Practical calculation methods of spectral density of instantaneous normalized frequency deviation from instantaneous time error samples** 449
F. Vernotte, J. Groslambert**, J.J. Gagnepain***
* Observatoire de Besançon, France
** LPMO - CNRS, Besançon, France

81. **A simple temperature control for semiconductor lasers** 456
G.D. Rovera, A. De Marchi**, G. Santarelli**
* LPTF, Paris, France
** Dipt. Elletronica Politecnico, Torino, Italy

82. **Ultralinear small-angle phase modulator** 461
J. Lowe, F. Walls
National Inst. of Standards and Technology, Boulder, Colorado, USA

OPENING SESSION



5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session OPENING SESSION Number 1

TIME AND FREQUENCY METROLOGY: CURRENT STATUS AND FUTURE CONSIDERATIONS

By

David W. Allan
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado

ABSTRACT

Over the last three decades we have seen methods evolve for the characterization of clocks and oscillators. The principal progress has been made in the time domain, but significant clarification has occurred in the frequency domain as well. We now have a CCIR recommendation and an IEEE standard for clock characterization. However, as we see the frontiers pushed forward, we can see a demand for even more exacting characterization of oscillators and clocks. The effect of environmental perturbations on clocks and oscillators is likely to become even more important as it effects the long-term stability. In addition, we see an important need to characterize measurement systems. To date, there is no standard for the proper characterization of measurement systems. Yet often the techniques for comparing either laboratory or remotely separated clocks may limit their frequency or time stability. As we look a decade ahead, these measurement concerns become even more important. Hence, there is a clear need to arrive at measures for characterizing measurement systems. In addition, we need to refine our ability to characterize the advanced clocks we anticipate in the coming years. Stronger ties with the telecommunications industry, where system timing and synchronization needs are high, will be mutually beneficial to both fields. This paper reviews some of the highlights of time and frequency metrology, makes recommendations for some needed standardization, and calls attention to certain unresolved problems.

INTRODUCTION

There is a need to be still more definitive in the measures we use to describe time and frequency (T/F) devices, and comparison systems. There are now a useful IEEE standard (No. 1139-1988) and a consistent CCIR recommendation for characterizing clocks and oscillators. [1] There is not, however, a similar measurement standard for time and frequency measurement systems (which might include clocks and oscillators).

Since the construction of the first atomic clock in 1948 we have seen about a factor-of-10 improvement in accuracy of primary standards every seven years. We expect this trend to continue. This will place new demands on metrology within the lab as well as on comparisons of clocks remote to each other.

In addition, as we design servo electronics, we cannot optimize system design unless we have properly characterized both the frequency standards involved and the measurement electronics. Within the laboratory environment this is often done well. However, when servo-controlling clocks remote from each other, serious mistakes are often made.

FOUR AREAS OF T/F METROLOGY

There are four general categories of time and frequency metrology as shown in Table 1. The first category involves frequency sources. Many advances, large and small, over the years have produced devices of remarkable stability and accuracy. A recent example is the work of Andrea De Marchi [2], who showed how Rabi pulling in cesium-beam frequency standards degrades the long-term frequency stability of these devices. Understanding these effects, he then showed how to reduce environmental perturbations on cesium standards. A significant improvement in long-term frequency stability and accuracy is now available.

The second category is turning a frequency standard into a reliable clock. At first thought, counting cycles from a frequency standard seems straightforward. But there are problems knowing absolute delays through critical parts of the measurement equipment. At the sub-nanosecond scale and over long distances, knowing these delays poses significant challenges. In addition, there is promise for a frequency standard in the optical region of the spectrum. We are probably decades away from being able to count optical frequencies without degradation. A breakthrough is needed.

Included within this second category is the important problem of properly using combining algorithms. When there is more than one clock, how should the readings be combined? Here again characterizing both the clocks as well as the measurement systems is essential for algorithm optimization. Caution is necessary because algorithms can make matters worse. Properly used, algorithms can provide improved reliability and performance. Past work has demonstrated that for intermediate and long-term frequency stability, the output of an algorithm can be better than the best physical clock in the system. Now,

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

Table 1. Areas of Time & Frequency Metrology

Frequency Sources	Time Keeping Metrology	Measurement Systems	Analysis and Modeling
ISSUES			
State preparation	Counting	Time interval & frequency counters	Systematics
Interrogation	Frequency dividers		Stochastic
Particle detection	Delays absolute stable	Dual-mixer time difference	Frequency-domain
Frequency multiplication		Direct	Time domain
Servo electronics	Trigger points	Heterodyne	Optimum estimation
Spectral purity	Balanced zero-crossing detectors	Homodyne	Time & frequency stability & accuracy
Signal-to-noise	Distortion	Dead time	Abnormal behavior detection
Internal & external perturbations	Dispersion	Hardware Filters Bandpass High pass Low pass	Digital filters
	Combining algorithms		Distribution of error
		Dissemination & comparision systems	

the short-term stability of a good individual clock is better than can be achieved with an algorithm. This is due to digitization noise and long servo response times. As digitization speeds improve, the short-term stability of an algorithm's output may eventually be better than any of the contributing clocks.

Because of the high stability and accuracy of time and frequency devices, direct measurement methods almost always add significant measurement noise. A common example is the use of time-interval counters. If, for example, a counter has 1 ns resolution, it may require hours or even days of integration before the digital measurement noise is less than that of a precise clock. Such a measurement arrangement cannot be used to determine the intermediate or short-term stability of good clocks. There are more cost-effective, stable and accurate ways to characterize precise clocks. In general, it is desirable that the measurement system noise be less than the noise of the clock or oscillator. This is especially important for those regions of integration or sample time of interest. [1]

I cannot do full justice to analysis methods here, but I will present a few problems as examples. Consider the 1 ns counter mentioned above. If this is the measurement equipment used, and we want to measure the frequency difference between the clocks, then the data should be taken at the maximum rate possible; this is so because if the counter's measurement noise is well modeled as white noise PM (phase modulation), then the uncertainty of the frequency estimate is

$$s = \frac{\sqrt{12}\,\sigma}{\tau_0 N^{3/2}}, \qquad (1)$$

where σ is the standard deviation of N time measurements spaced τ_0 apart. Since the uncertainty decreases as $\tau_0 N^{3/2}$ grows, there is significant gain in taking as many points as possible, yet this is seldom done. Similarly, the optimum time difference estimate is best taken as the mean of as large a set of independent individual measurements as is practicable. The confidence of this estimate is the standard deviation of the mean. Such measurements often include systematic effects which make the classical estimates invalid. Characterizing these systematics is important, but is seldom done. Another common analysis error is made in processing time or phase differences between cesium beam clocks, rubidium gas-cell clocks, or any of the passive atomic standards where the intermediate-term stability is characterized by white-noise FM. White-noise FM is the same as random walk PM. Analysts will often calculate a linear regression on the time difference measurements to estimate the frequency difference. This is far from optimum. It is equivalent to throwing away a significant percentage of the data as compared to an optimum estimator. The linear regression on the time differences is optimum for white-noise PM. This noise is found in the very short term for quartz oscillators and active hydrogen masers. Hence, in general, this linear regression on the phase or time difference measurements may give much less than optimum results.

MEASURES FOR STANDARDS AND MEASUREMENT SYSTEMS

Figure 1 illustrates a generic phase locked loop or time (phase) difference measurement system. This figure is a concept diagram. It could apply, for example, to phase locking a local oscillator, comparing data between remote clocks or synchronizing a network. First, consider that the two clocks involved have frequency determining elements. In general, the basic physical quantity of interest in a clock is its frequency, not its phase or

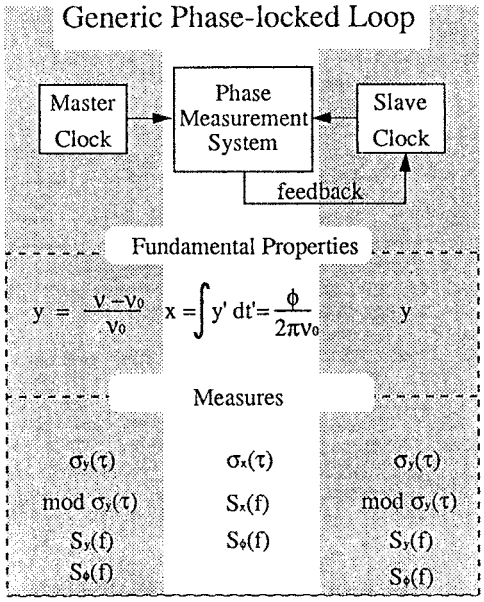


Fig.1. A fundamental concept diagram showing the properties and measures for clocks in contrast to measurement systems. The measurement system can be in the laboratory or it can be a comparison system for measuring clocks remote from each other. These concepts could also apply to a phase-locked servo system or a time dissemination system.

time. If the clocks are ideal the residuals around the mean frequency will have a white-noise FM spectrum (random-walk PM) in some useful regions of Fourier space. In contrast, for a measurement system the basic quantity of importance is the phase or time difference and not the frequency with a few exceptions. Hence, in the case of an ideal measurement system the residuals will have a white-noise PM spectrum.

Our conclusions from this distinction is that $S_y(f)$, $\sigma_y(\tau)$ or $\mathrm{mod}\sigma_y(\tau)$ are useful measures for characterizing most clocks and frequency standards (y is the relative frequency and x is the time residuals gotten by integrating y). Similarly, we conclude that $S_x(f)$, $S_\phi(f)$, $\mathcal{L}(f)$ and $\sigma_x(\tau)$ are useful measures for characterizing typical measurement systems. $S_\phi(f)$ or $\mathcal{L}(f)$ are often used to characterize precise oscillators and have been shown to have useful properties. However, because frequency is basic and phase or time is proportional to the integral of frequency, the spectrum of the phase or the time will behave like the frequency spectrum divided by f^2. Walls and Percival [3] have shown that, within a finite analyzer band-width, significant biases (several decibels) can occur as a result of the steeper slopes typically used in modeling the phase or time spectral density for precision oscillators. Of course, corrections can be made for this, but the potential for error is greater. In general, if a spectrum is more nearly flat, then these biases will be negligible. It might well be wise to limit the use of these latter measures ($S_\phi(f)$ and $\mathcal{L}(f)$) for measurement system characterization. In the time domain the preference for measuring the frequency stability of a measurement system or of a time and frequency comparison or dissemination system is $\mathrm{mod}\sigma_y(\tau)$.

For the frequency domain, IEEE Standard 1139-1988 recommends the measures given in Table 2. [1] From the arguments above we conclude that these form a good set of measures for frequency standards and clocks as well as for measurement systems. In the time domain we conclude that the measures recommended are a good set for frequency standards and clocks. But these are deficient for characterizing measurement systems. Table 3 lists the recommended measures along with a proposed time stability (TVAR) measure which satisfies this deficiency. Table 4 lists the spectral density relationships for this measure and the range of convergence. For a finite data set an estimate of this measure (TVAR) is

$$\sigma_x^2(\tau) \approx \frac{1}{6n^2(N-3n+1)} \sum_{k=1}^{N-3n+1} \left(\sum_{i=k}^{n+k-1} (x_{i+2n} - 2x_{i+n} + x_i) \right)^2, \qquad (2)$$

where $\tau = n\tau_0$.

Equation (2) and Table 3, show that the subscript i denotes data taken from the original set, whereas the subscript k denotes an average over n values of the x_is. The averaging of the x_is has the effect of digitally narrowing the bandwidth in the software to $f_s = f_h/n$, where f_h is the hardware bandwidth of the measurement system. [1][4][5][6]

Figures 2, 3 and 4 are the time-to-frequency domain mappings for $\sigma_y^2(\tau)$, $\mathrm{mod}\sigma_y^2(\tau)$ and for $\sigma_x^2(\tau)$, respectively. Each of the three figures shows the usual range of applicability for precision oscillators and for measurement systems. In Figure 4 we immediately see an advantage of using $\sigma_x(\tau)$.

Table 2. IEEE Standard 1139 (1988)

Frequency Domain Spectral Densities

$S_y(f)$, $S_\phi(f)$, $S_{\dot\phi}(f)$, or $S_x(f)$

Relationships:

$$S_y(f) = \frac{f^2}{\nu_0^2} S_\phi(f)$$

$$S_{\dot\phi}(f) = (2\pi f)^2 S_\phi(f)$$

$$S_x(f) = \frac{1}{(2\pi\nu_0)^2} S_\phi(f)$$

$$S_y(f) = (2\pi f)^2 S_x(f).$$

Table 3. IEEE Standard 1139 (1988)

Time Domain Variances

AVAR:

$$\sigma_y^2(\tau) = \frac{1}{2}\langle(\bar{y}_{i+1} - \bar{y}_i)^2\rangle$$

$$= \frac{1}{2\tau^2}\langle(x_{i+2} - 2x_{i+1} + x_{ik})^2\rangle$$

MVAR:

$$\mathrm{mod}.\sigma_y^2(\tau) = \frac{1}{2\tau^2}\langle(\bar{x}_{k+2} - 2\bar{x}_{k+1} + \bar{x}_k)^2\rangle$$

Proposed

TVAR:

$$\sigma_x^2(\tau) = \frac{\tau^2}{3}\,\mathrm{mod}.\sigma_y^2(\tau)$$

$$= \frac{1}{6}\langle(\bar{x}_{k+2} - 2\bar{x}_{k+1} + \bar{x}_k)^2\rangle$$

Table 4. Spectral Density and Time Domain Relationships

$$S_x(f) \sim f^{\beta},$$

$$\sigma^2_x(\tau) \sim \tau^{\eta},$$

$$\beta = -\eta - 1,$$

$$-5 < \beta \leq 1$$

where $S_x(f)$ is the spectral density of the time difference measurements, x, and β denotes the kind of power-law spectrum.

$$S_y(f) \sim f^{\alpha}$$

$$\alpha = \beta + 2,$$

$$-3 < \alpha \leq +3$$

where y is the normalized frequency.

Since the usual types of measurement noise are centered around $\beta = 0$, this gives a near-zero dependence on τ (a desirable trait for a good measure). Other useful characteristics of this measure are:

- it is equal to the classical standard deviation of the time difference measurements for $\tau = \tau_0$, for white-noise PM;
- it equals the standard deviation of the mean of the time difference measurements for $\tau = N\tau_0$ (the data length), for white-noise PM;
- it is convergent and well behaved for the random processes commonly encountered in time and frequency metrology;
- the τ dependence indicates the power-law spectral density model appropriate for the data;
- the amplitude of $\sigma_x(\tau)$ at a particular value of τ, along with the assumption of one of the five power-law spectral-density models (β = -4, -3, -2, -1, 0), provides enough information to estimate the corresponding level in the frequency domain for any of the recommended IEEE-standard spectral-density measures.

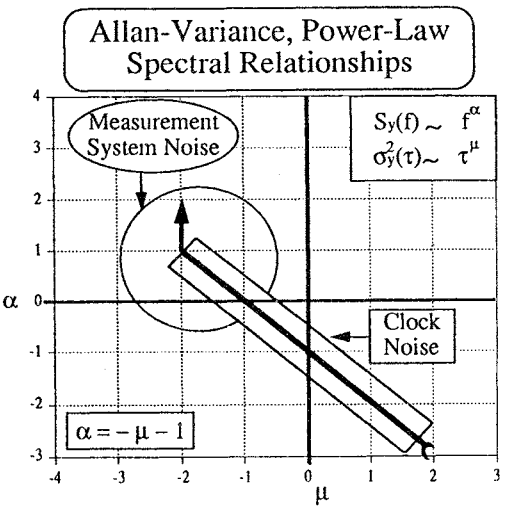


Fig. 2. The broad line represents the corresponding value of α for each value of μ. The rectangle is to graphically illustrate where the broad line is applicable for clock noise. Similarly, the circle is to graphically illustrate where the broad line is applicable for measurement system noise.

Other important considerations for a measurement, comparison, or dissemination system are time accuracy, frequency accuracy and frequency stability. For a measurement system, frequency accuracy will typically be a function of the averaging time as well as the processing method. As shown before, this can be very important. I will give another illustration later.

Whereas the discussion above has focused on time and frequency stability measures, time and frequency accuracy are also very important to metrology. Frequency accuracy as it relates to the fundamental definition of the second seems to be well in hand. However, there is a deficiency in the literature in relation to the definition of and use of time accuracy. Limited work has been done. [7] Absolute frequency has meaning in physics, but absolute time does not. Time accuracy has meaning only as a measurement is traceable to some defined or agreed-upon time standard or reference. Time accuracy can be thought of, for example, as the transport of a perfect portable clock from the agreed upon

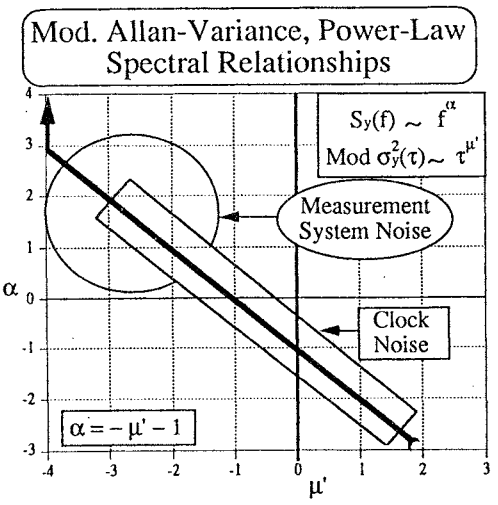


Fig. 3. The broad line represents the corresponding value of α for each value of μ'. The rectangle is to graphically illustrate where the broad line is applicable for clock noise. Similarly, the circle is to graphically illustrate where the broad line is applicable for measurement system noise.

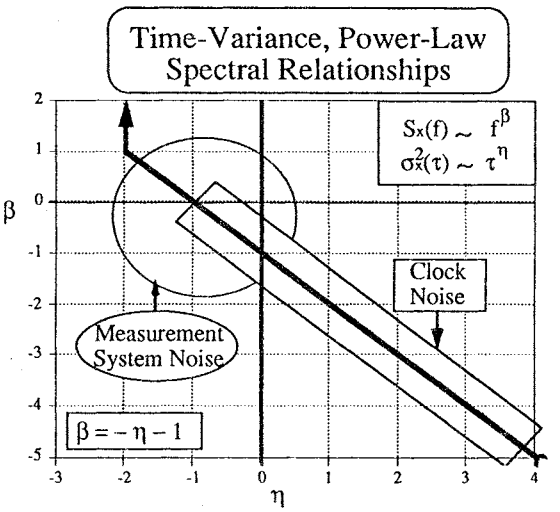


Fig. 4. The broad line represents the corresponding value of β for each value of η. The rectangle is to graphically illustrate where the broad line is applicable for clock noise. Similarly, the circle is to graphically illustrate where the broad line is applicable for measurement system noise.

reference to a desired point including all relativistic effects. The accuracy of transferring time from an agreed upon standard to another point is often hard to evaluate or estimate at high levels of precision. Unknown systematic delays and delay variations often become very important in this regard.

CHARACTERIZATION OF SYSTEMATICS EFFECTS

Properly characterizing the stochastic processes allows optimal estimation of the environmental perturbations due to temperature, humidity or other factors. If, for example, we see white-noise FM (α = 0) in a cesium standard, the optimum estimator of frequency drift is a linear regression on the frequency and not a quadratic on the time residuals or on the phase residuals. Often we see random-walk FM (α = -2) as the long-term behavior of cesium or rubidium standards. If we wish to estimate frequency drift in the presence of this noise, the optimum estimator is a second-difference operator, rather than a simple linear regression. [1]

If a modulation side-band is found in the data, due, for example, to a diurnal effect, then such a side-band can be clearly seen when analyzed in the frequency domain. It can also be observed in the time domain. Figure 5 shows the effect of modulation on a $\sigma_x(\tau)$ plot. The effect of modulation decays as $1/\tau$, whereas the white-noise PM background decays only as $\tau^{-1/2}$.

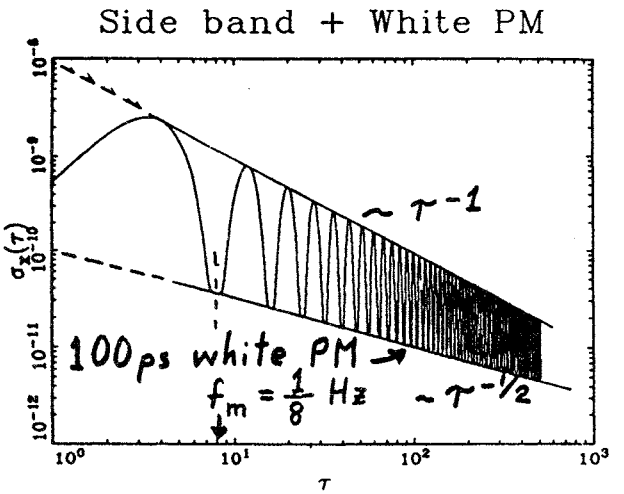


Fig. 5. An illustration of the effect of a modulation side band on top of a white-noise PM background. The solid straight lines are the theoretical maximum and minimum limits for $\sigma_x(\tau)$.

COMPARISON OF $\sigma_y(\tau)$ AND $\sigma_x(\tau)$

The IEEE and CCIR definition for $\sigma_y^2(\tau)$ is

$$\sigma_y^2(\tau) = \tfrac{1}{2}\langle(\bar{y}_{j+1} - \bar{y}_j)^2\rangle, \quad (3)$$

where

$$\bar{y}_j = \frac{x_{j+1} - x_j}{\tau}. \quad (4)$$

This estimate for y is an optimum estimate if the spectrum of the residuals is white-noise FM. This noise is the theoretically expected noise for a passive atomic frequency standard. The first-difference operator allows this measure to remain convergent for the low-frequency divergent power-law spectra met in modeling precision oscillators. Other advantages of $\sigma_y^2(\tau)$ are that it is equal to the classical variance for white FM. It is theoretically derivable from any of the commonly occurring spectral density models in precise oscillators. It is intuitive in that it is a measure of the change in frequency over some interval τ as may be germane, for example, in determining the optimum attack time for a servo design or for choosing the appropriate oscillator for a radar returned signal. In addition to

these theoretical and practical advantages, we have from the arguments above an additional intuitive feeling for why $\sigma_y(\tau)$ is a good measure for clocks and oscillators. It is an optimum estimation of frequency change for white-noise FM.

The principal disadvantage of $\sigma_y(\tau)$ is the ambiguity for power-law processes with $\alpha \geq +1$ where we cannot easily distinguish between white-noise PM and flicker-noise PM. This ambiguity was the main motivation for the development of $\mathrm{mod}\sigma_y(\tau)$. As shown above, $\sigma_y(\tau)$ differences adjacent optimum frequency estimates (for white-noise FM) each measured over an interval τ. While this is a good measure for frequency stability, it has limitations as a measure where time or phase is the basic quantity, such as for measurement systems or for the synchronization of networks.

Since $\sigma_x(\tau)$ is derived directly from $\mathrm{mod}\sigma_y(\tau)$, it has all the associated advantages. For example, $\sigma_x(\tau)$ can be used to differentiate between white and flicker-noise PM. The second-difference expression for $\sigma_x(\tau)$ is

$$\sigma_x(\tau) = \left[\frac{1}{6}\langle(\bar{x}_{k+2}-2\bar{x}_{k+1}+\bar{x}_k)^2\rangle\right]^{1/2}. \qquad (5)$$

The equivalent second-difference equation for $\sigma_y(\tau)$ is similarly

$$\sigma_y(\tau) = \left[\frac{1}{2\tau^2}\langle(x_{j+2}-2x_{j+1}+x_j)^2\rangle\right]^{1/2}. \qquad (6)$$

$\sigma_x(\tau)$ is proportional to the rms second difference of time (phase) averages rather than the rms second difference of the time (phase) measurement points as in $\sigma_y(\tau)$. If we have white-noise PM, then the optimum estimate of the time from a set of time readings is the simple mean. Since $\sigma_x(\tau)$ is proportional to the rms change in the mean value of the time difference measurements averaged over an interval τ with respect to a similar average taken just before and just after τ, we see that it is an optimum estimator of changes in time (phase) over the interval τ. Hence, this measure is most useful where time or phase differences are the basic quantity of interest. Examples of application include network synchronization, time and phase difference measurements, phase-locked servo systems, remote time transfer and comparison systems, and time distribution systems.

A look at the transfer function for $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ for different values of τ provides a good insight into their usefulness. We can show that, for $\tau = n\tau_0$ and $n = 2^i$ ($i = 0, 1, 2, \ldots$), the set of resultant transfer functions form a nearly orthogonal set as viewed from the frequency domain. Figures 6 and 7 are the composite transfer functions for $i = 0$ to 8. For $\sigma_y^2(\tau)$ the composite transfer function is nearly square and flat within about 5% over more than two decades. One of the features of $\mathrm{mod}\sigma_y(\tau)$ is that it changes the bandwidth within the software; hence the composite transfer function is a little less flat than that for $\sigma_y(\tau)$. It is also more nearly square at the high frequency end.

The time-domain, frequency-domain relationships are known for $\sigma_y(\tau)$ and for $\mathrm{mod}\sigma_y(\tau)$ [1]. Since $\sigma_x(\tau) = \tau\,\mathrm{mod}\sigma_y(\tau)/\sqrt{3}$, these known relationships can be used for $\sigma_x(\tau)$ as well.

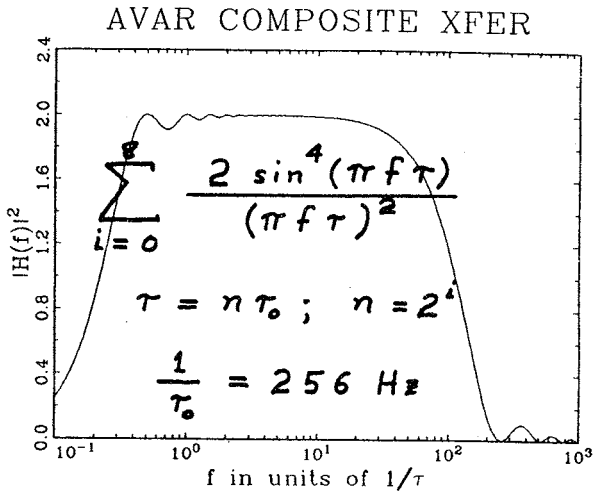


Fig. 6. A plot of the sum of nine individual transfer functions for $\sigma_y^2(\tau)$. The number of transfer functions included in the sum, of course, is arbitrary. The approximate square window, as viewed from the frequency domain, can be made as wide as one wishes -- limited by the data length. The value of n must be less than or equal to N/2.

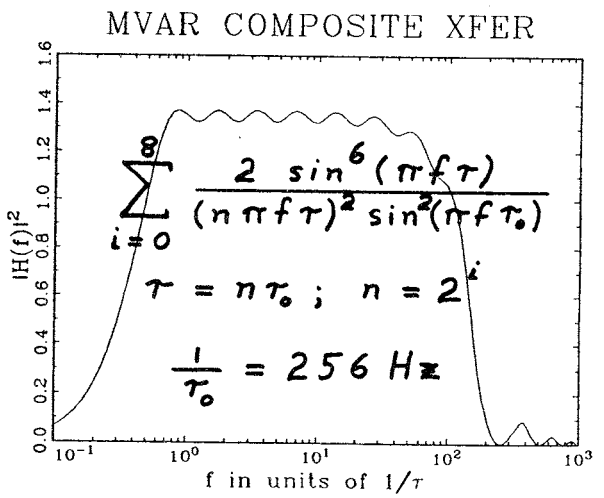


Fig. 7. A plot of the sum of nine individual transfer functions for $\mathrm{mod}\sigma_y^2(\tau)$. The number of transfer functions included in the sum, of course, is arbitrary. The approximate square window, as viewed from the frequency domain, can be made as wide as one wishes -- limited by the data length. The value of n must be less than or equal to N/3. Notice that Figure 7 has a slightly sharper high-frequency edge than Figure 6.

SOFTWARE CONSIDERATIONS FOR $\mathrm{mod}\sigma_y(\tau)$ AND $\sigma_x(\tau)$

The computation time for these two measures is essentially the same. They differ only by a factor. However, because of the double sum in equation (2), a direct software implementation of this form of the equation will take substantial CPU time for large data sets. 10^6 data points may take several hours to process on a main-frame machine, and the CPU time grows as the number of points squared. There are two procedures that can reduce the computation time enormously.

The first procedure can be conceptualized by looking at Figure 8. This Figure illustrates a case where n = 4, and the windows show which time points (spaced τ_0 apart) are used to make up each window's time average. The averages from the three windows provide the entries for one second-difference. An rms of these second differences is computed across all possible sets for a particular n. Equation (2) implies that we recompute the sums over the x_is each time we move the second difference one point forward. However, instead of repeating the sum, we can drop a point at the back end of each window in Figure 8, add the point dropped to the previous window (windows 1 and 2), and add a new point to the third window to the right. This properly implements the sum much more efficiently. This procedure reduces the CPU time to be proportional to the number of points rather than to the number of points squared as above.

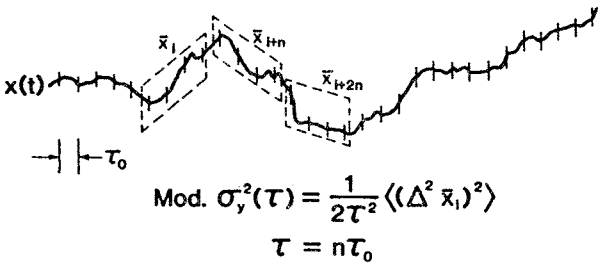


Fig. 8. A plot illustrating a set of x(t) values measured τ_0 apart. The windows block out a single set of these measurements as they are combined to construct a single second difference of the average value taken from each window's set of x(t) readings. For a given value of n, the three windows are moved from the beginning of the data to the end to obtain an estimate of the infinite-time average of these sqared second differences.

The second procedure is probably useful only for very large data sets - on the order of 10^5 or more points. It takes advantage of the confidence of the estimate obtainable from such large sets. It is not necessary to take all possible averages and second-difference combinations to get a very good confidence of the estimate from a large data set. For example, the original data points could be averaged in nonoverlapping blocks of 128 (n = 128) to form a new data set. The number of data points in this new set would be N/128, where N is the original number in the set. This process can, of course, be repeated as many times as needed and for any value of n. As argued before, the set $n = 2^i$ ($i = 0, 1, 2 \ldots \leq \log(N/3)/\log 2$) is a nearly orthogonal set as viewed in the frequency domain. The data can then be analyzed in blocks. If this is done the rms values can be obtained across all blocks for the whole set along with an internal estimate of the confidence interval as calculated from the standard deviation of the mean of the variances at a particular value of n. This internal estimate of the confidence also provides a check on the hetero-schedasticity of the modeling; that is, are the measures and the model well behaved with time. The sacrifice in using this procedure is small for white-noise PM. Fortunately, this is the ideal noise met in time or phase measurement systems. This second procedure also gives a significant reduction in computation time. For large data sets the penalty in loss of confidence of the estimate is typically very small.

SOME PRACTICAL APPLICATIONS

Figure 9 is a $\mathrm{mod}\sigma_y(\tau)$ plot for comparing a passive hydrogen maser at NBS (now NIST) against the NRC (National Research Council) primary cesium standard, Cs 5. Two features are worth noting. First, there is excellent agreement between the theoretical curves and the experimental data. And second, with optimum combining algorithms and analysis procedures, we can see a white-noise PM modulation of $\sigma_x(\tau$ = 1 day) = 0.8 ns and a relative stability for the clocks of $\mathrm{mod}\sigma_y(\tau$ = 8 days) = 2 X 10^{-15}. This experiment was repeated three times with similar results. As an aside, it is apparent that optimally weighted time transfer data using the GPS common-view method serves this comparison of geographically separated clocks quite well.

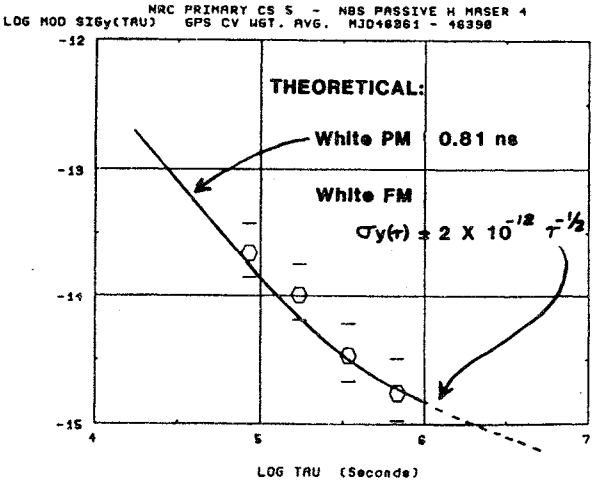


Fig. 9. A $\mathrm{mod}\sigma_y(\tau)$ plot of the frequency stability between a passive-hydrogen maser located in Boulder, Colorado and the Cs 5 Primary Frequency Standard of NRC located in Ottawa, Canada. The distance between these timing centers is about 3 Mm. The measurement system was an optimully weighted set of GPS satellites used in the common-view mode.

Figure 10 is a combined $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ diagram showing performances for the usual kinds of measurement, dissemination, and comparison systems. Figure 11 is a $\sigma_x(\tau)$ diagram for a subset of these measurement and comparison systems along with the performance of typical atomic clocks. The noise seen in the two-way satellite time transfer technique is outstanding for τ equal one second to a few minutes. Unfortunately this white-noise PM behavior does not persist for larger τ values due to systematic effects in the transmitting and receiving equipment. This

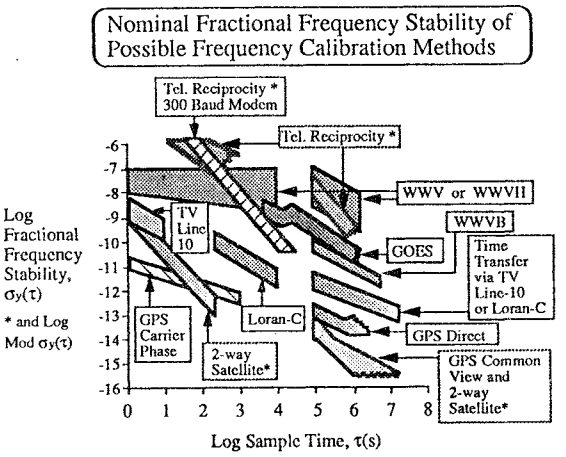


Fig. 10. A plot showing the frequency transfer capability of various comparison and dissemination systems. The accuracy with which frequency can be transfered is obviously a funcion of τ. Where white-noise PM was an appropriate model, $\mathrm{mod}\sigma_y(\tau)$ was used as the measure as indicated by the asterisk, *.

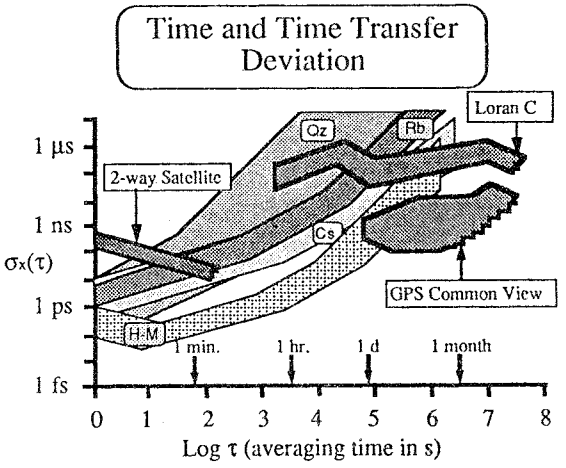


Fig. 11. A $\sigma_x(\tau)$ plot of a subset of comparison or dissemination systems shown in Figure 10. Also plotted are the time stabilities of commonly used precise clocks. It becomes obvious that the measurement noise is often the limiting factor.

problem is currently being addressed, and with the aid of proper measurement techniques a significant resolution is expected. Loran-C shows a double hump due to the combination of a diurnal term and an annual term. These annual variations plagued international time comparisons and the generation of TAI for many years. When we thought the averaging time was sufficiently long to see the behavior of two widely separated clocks, doubts arose about whether the variations were caused by annual or seasonal variations in Loran-C or similar variations in the clocks. The top part of the GPS common-view curve is from an around-the-world comparison performed over several years. We see again an annual term. This is apparently due to annual variations in the errors associated with the ionospheric model broadcast from the satellites. The bottom of the curve reflects results of the NBS/NRC (NBS is now NIST) comparison shown in Figure 9. It is easy to see why GPS has been widely accepted for international time comparison.

As we look forward to the development of advanced clocks and frequency standards with accuracies and stabilities beyond one part in 10^{15}, we face significant challenges in performing comparisons of widely separated clocks and frequency standards (see Figure 11). Table 5 is a list of some of current time comparison techniques as well as some proposed. Important improvements are needed for measurement and comparison systems as we prepare for future clocks and frequency standards.

Table 5. Time & Frequency Transfer Research

Type	$\sigma_x(\tau = 1\ \mathrm{d})$		$\mathrm{mod}\sigma_y(\tau = 20\ \mathrm{d})$
GPS	~50	ns	10^{-13}
GPS C-V	~1-6	ns	10^{-15}
2-Way Sat.	~1-10	ns	10^{-15}
Trilateration	~10	ns	10^{-15}
LASSO	~0.1	ns	10^{-16}
STIFT	~0.02	ns	10^{-17}

The first three types are currently operating systems. The last three are proposed. Trilateration is a specific service being built up by NIST. It will operate in the receive only mode at the KU band. It will use a transponder on board a geostationary satellite. The satellites position will be actively determined using triangulation from three known timing centers. Both the LASSO and STIFT proposed time and frequency transfer techniques have been published.

CONCLUSIONS

The progress in time and frequency metrology over the last three decades is impressive. An excellent bibliography covering the measurements used is the reference list in a paper by Walls and Rutman. [8] Advances in frequency and time sources have been similarly impressive. As we review the past and look to the future, we see a continuum of change. There will be a need to better characterize clocks and oscillators, for outlooks on performance advances are good.

The current standards for characterizing clocks, oscillators, measurement systems, and dissemination systems are useful and adequate except in two areas. These deficiencies are (1) the characterization and modeling of environmentally induced perturbations, which often cause long-term instabilities in clocks and oscillators, and (2) the characterization of the time-domain behavior of measurement, dissemination, and comparison systems. A time-domain time stability measure, $\sigma_x(\tau)$, has been studied, tested and is proposed as a measure which will help resolve these deficiencies. Dealing with both these deficiencies is a necessary step to improve time and frequency metrology and to keep up with the new clocks and oscillators anticipated in the future.

As we develop new international time transfer systems, such as the two-way satellite time-transfer system, it will be extremely helpful to have a common language, as well as common and needed measures to characterize performance. Additional tools may also be needed as we move to higher levels of accuracy and stability.

ACKNOWLEDGEMENTS

I thank the EFTF committee for inviting me to participate in the Forum. Several people have been very helpful in discussions as well as in preparation of this paper. Specifically, I am indebted to Tom Weissert, Marc Weiss, Donald Sullivan, James A. Barnes, Roger E. Beehler and Fred L. Walls.

References:

[1] D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, Eds., "Characterization of clocks and oscillators," NIST Tech Note 1337, March 1990.

[2] A. DeMarchi, G.D. Rovera, R.E. Drullinger, and D.A. Howe, A Beam Reversal Experiment on NBS-6 Primary Cs Standard Including Rabi Pulling Evaluation," Proceedings of the 39th Annual Symposium on Frequency Control, 3-7, 1985.

[3] F. L. Walls, D. B. Percival and W. R. Irelan, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," Proceedings of the 43rd Annual Symposium on Frequency Control, 31 May - 2 June, 1989, pp. 336-341.

[4] D. W. Allan and J. A. Barnes, "A modified Allan variance with increased oscillator characterization ability," Proc. of 35 Annu. Frequency Control Symp., 1981, pp. 470-475.

[5] P. Lesage and T. Ayi, "Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate, IEEE Transactions on Instrumentation and Measurement, 1984, IM-33, pp. 332-337; see also D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, Eds., "Characterization of clocks and oscillators," NIST Tech Note 1337, March 1990. (same as reference 1 above)

[6] L. G. Bernier, "Theoretical analysis of the modified Allan variance," Proc. 41st Ann. Frequency Control Symp., pp. 116-121, 1987.

[7] D.W. Allan, "Synchronization of Clocks," Proc. of the 21st Annual Precise Time and Time Interval Planning Meeting, Redondo Beach, CA, Nov. 30 - Dec. 1, 1989, pp. 1-16.

[8] F. L. Walls and J. Rutman, "Characterization of Frequency Stability in Precision Frequency Sources," to be published in Proc. of the IEEE.

SESSION 1A

ATOMIC CLOCKS I

Chairman : C. AUDOIN

Frequency Standard at 3.39 μm based on Linear Optical Ramsey Resonance in a Methane Molecular Beam.

G. Kramer
PTB, Braunschweig, Germany

The extension of coherent frequency synthesis throughout the infrared part of the spectrum has blurred the distinction between (microwave) frequency and (optical) wavelength standards. A stabilised laser combined with a frequency measurement chain constitutes an optical clock, very much like the cesium clock.

The frequency reference in our optical clock (at PTB) is a molecular transition at 88 THz in a methane beam. Very narrow lines with a fractional width of 6 parts in 10^{11} are being observed using a Linear Optical Ramsey Resonance scheme. While the classical Ramsey excitation requires a narrow, collimated beam, in this scheme a mechanical grating is inserted between the interaction regions, so as to obstruct certain trajectories for the molecules. Resonances free of Doppler broadening result without the use of non-linear effects such as saturation or two-photon absorption.

At a level below 1 part in 10^{12}, the accuracy performance of this optical frequency standard still fails to reach that of the best cesium standards by an order of magnitude. As there is a large potential for further improvement, optical schemes should be kept in mind as candidates for the frequency standard of the future.

Session 1A Plenary Number 3

Frequency Shifts in Cesium Clocks due to Majorana Transitions: Towards a theoretical understanding

A. Bauch, R. Schröder

Physikalisch-Technische Bundesanstalt
3300 Braunschweig, Germany

Abstract

A theory is presented which describes a certain class of shifts of the output frequency of a cesium atomic clock. The successive action of Majorana transitions ($\Delta F=0$) and radio-frequency transitions ($\Delta F=1$, $\Delta m=0, \pm 1$) is identified as the reason of frequency shifts which were reported since long as being caused by Majorana transitions. The treatment is closely adapted to previous experiments made with the PTB's experimental device CSX. First results are given and compared with experimental findings.

1. Introduction

The function of a cesium clock can be described using the following simplifying scheme. A state-selected atomic beam leaves the polarizer magnet. A change in the population caused by the irradiation with a probing field is monitored using an analyzer magnet and an ionization detector. The transitions between the quantum states involved are induced by an applied rf magnetic field. Following Ramsey's idea, the field is applied in two phase coherent bursts of length τ separated by a time interval T /1/. There is an analogy between this kind of excitation and the quantum mechanical double slit experiment: Even when the experimental conditions allow a transition between the two involved states it is yet indiscernible in which irradiation section the transition takes place. Two ways are open for an atom to reach the final state starting from a given initial state. This gives rise to interference products of the two transition amplitudes in the expression for the transition probability with a characteristic frequency of $1/(2T)$ as the resulting linewidth.

For the clock operation, the (4,0)-(3,0) transition, one of the σ-transitions obeying $\Delta F=1$, $\Delta m=0$, is used and an oscillator is stabilized to the line centre of this transition. Because of the selection rules, the direction of rf fields is chosen primarily parallel to the quantization axis which is defined by a static magnetic field.

It is often found that this scheme is not sufficient to describe the performance completely. If the atomic beam passes inhomogeneous portions of the static magnetic field, transitions among the Zeeman sublevels are induced, named after Ettore Majorana, who was the first to give a theoretical treatment on this subject /2/.

The atoms are also subjected to rf magnetic fields perpendicular to the quantization field when the atoms traverse through the field lobes of the standing rf field sustained in the resonator. The unavoidable extension of the atomic beam or even an intentionally introduced misalignment in a two beam commercial tube /3/ are the cause. So-called π-transitions with the selection rule, $\Delta F=1$, $\Delta m=\pm 1$ are induced.

In this work a treatment will be presented which encloses all three kinds of transitions. It is closely adapted to the experimental situation in the CSX, the PTB's experimental frequency standard, where some studies have been made in the past /4/. In the CSX, the initial atomic beam consists of atoms in the states F=4, m=-3, ... +4 /5/. The atoms subsequently pass the region of inhomogeneous fields, a drift region and finally the Ramsey cavity. For simplicity it is assumed in this paper that atoms in the state F=3 are detected. This is not the experimental situation, as in the CSX a flop-out detection scheme is used /6/. However, the essential of the treatment is more easily pictured and made in the way chosen here. It is as follows:
An individual atom [4,m,v], characterized by its quantum numbers F=4,m and its velocity v behind the polarizer magnet is followed up through the beam tube. After having passed the region of inhomogeneous static fields, it is in a superposition of (4,m)-eigenstates. In the resonator the (4,0)-(3,0) transition is induced with a large probability but also other $(4,m)\Rightarrow(3,m)$ and $(4,m)\Rightarrow(3,m\pm1)$ transitions are parasitically induced. So there are several indiscernible ways by which the atom [4,m,v] can reach one of the final levels [3,m',v]. In the calculation of the population of the (3,m') levels many interference products arise but only those are considered which contain not more than one term which

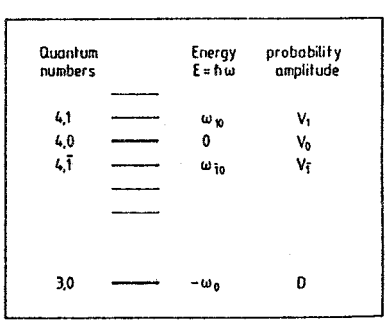


Fig. 1 Definition of the labeling of the quantum states, their energies expressed in frequency units and the probability amplitudes of the individual atom to be in the indicated state.

describes the parasitically induced transitions. In addition, at least one term which describes the (4,0)-(3,0) clock transition must be included as long as frequency shifts of the clock transition are to be explained. This conditions are met only in the probability amplitude of the (3,0) state. So during the passage of the cavity only the levels (4,0), (4,±1) and (3,0) need to be considered whereas in the inhomogeneous static fields the full set of 9 (4,m) levels has to be dealt with. Finally the equation (13) is achieved which links the probability of population of the (3,0) level of the atom of consideration to all the fields present along its way through the beam tube. This result is then multiplied with the relative number of atoms under consideration in the beam as described in /5/.

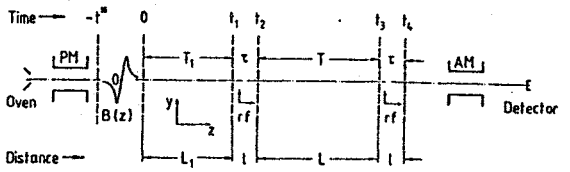


Fig. 2 Schematic of the experimental situation treated in this work, adapted to the PTB's CSX:
PM: Six pole/four pole polarizer magnet
AM: Four pole analyzer magnet
L_1= 250 mm, l=23.26 mm, L=767.7 mm; mean velocities are in the range of 400 m/s.

2. Definition of energy states and transition amplitudes

The quantum states are characterized by their quantum numbers F, m, their energy expressed in frequency units and their (complex) population amplitude in the way depicted in Fig. 1. It is

$$\omega_0 = 2\pi\cdot f_0 = 2\pi \cdot 9192631770\ s^{-1}$$
$$\omega_{10} = 2\pi\cdot(\ 349746\cdot B_0 - 13.349\cdot B_0^2)\ s^{-1} \qquad (1)$$
$$\omega_{\bar{1}0} = 2\pi\cdot(-349746\cdot B_0 - 13.349\cdot B_0^2)\ s^{-1}$$

where B_0 is the magnetic induction in Gauss ($=10^{-4}$T). An atom [F,m,v] which is found in the beam with the probability V_m^0 is now followed up through the subsequently acting fields, illustrated in Fig. 2.

1) Inbetween $-t^*$ and t=0 the atom passes inhomogeneous static magnetic fields. Then at t=0 the atom is in general in a superposition of Zeeman eigenstates and the probability amplitudes are given by

$$V_m\ (t=0) = \sum_{m'} T_{mm'}\cdot V_{m'}^0, \qquad (2)$$

The transfer coefficients $T_{mm'}$ are calculated using a numerical method as described in the appendix A1. The coefficients $\tilde{T}_{mm'}$, to be used later on are defined there, too.

2) Having passed a drift region of length L_1 during the time $T_1 = L_1/v$, the atom enters the first microwave interaction region which is traversed in the time τ, $\tau=l/v$. Here σ- and π- transitions are induced which are treated as independent from each other. As follows from /3/ this is not justified in general, but seems to be here, where much larger effects shall be described.

i) The transition amplitudes for the allowed and desired (4,0)↔(3,0) σ-transition is /1/:

$$Y_0(t) = i\ \sin(b_\sigma\cdot\tau)\exp\{i(\omega t-\phi_0\pm\omega\tau/2)\} \qquad (3)$$

The condition of near resonant excitation is assumed, meaning that the excitation parameter b_σ is much larger than $(\omega-\omega_0)$, where ω is the applied frequency and ω_0 is defined in Fig. 1. The "+" sign is in all cases valid when the atom is finally in the lower energy state whereas the "-" sign applies for the upper state and ϕ_0 is an arbitrary intial phase of the rf field. The corresponding amplitude to describe the non-occurence of a transition, i.e. (4,0)-(4,0) and (3,0)-(3,0), is

$$N_0(t) = \cos(b_\sigma\tau)\ \exp\{\pm i\omega\ \tau/2\}. \qquad (4)$$

ii) The amplitudes for the transitions (4,±1)↔(3,0) are derived in a perturbative solution of the Schrödinger equation which is given in the appendix A2. It follows

$$Y_1(t)=-\Pi_1\ \exp\{i(\omega t-\phi_0-\omega_{10}\tau/2\pm\omega\tau/2)\}$$
$$Y_{\bar{1}}(t)=+\Pi_{\bar{1}}\ \exp\{i(\omega t-\phi_0-\omega_{\bar{1}0}\tau/2\pm\omega\tau/2)\} \qquad (5a)$$

where

$$\Pi_i=2\cdot\Omega_i\cdot\tau^2\cdot\hat{b}_\pi\cdot\cos(\Omega_i\tau/2)/(\Omega_i^2\tau^2-\pi^2),$$
$$\Omega_i=\omega_{i0}+\omega-\omega_0,\ i = +1, -1 \qquad (6)$$

The excitation parameter for the π-transition, $\hat{b}_\pi$, is also derived in the appendix. Correspondingly we define

$$N_1(t)=1\cdot\exp\{-i(-\omega_{10}\tau/2\pm\omega\tau/2\}$$
$$N_{\bar{1}}(t)=1\cdot\exp\{\ i(-\omega_{\bar{1}0}\tau/2\pm\omega\tau/2\} \qquad (5b)$$

In the definitions (5a) and (5b) common phase factors have been neglected as will be justified subsequently.

3) For a period T the quantum states evolve undisturbed before the atom enters the second interaction region. At $t=t_4=T_1+T+2\tau$ the probability amplitude of population of the level (3,0), $D(t_4)$, is calculated. The further evolution in time leads to negligible phase factors as the detector is only sensitive to $|D|^2$.

For to calculate frequency shifts, $|D(t_4)|^2$ averaged over the population of atoms in the beam is calculated. The frequency ω is varied assuming square-wave frequency modulation with the modulation-width W/2 or 3·W/2. W is the approximate linewidth of the clock-transition Ramsey line of 260 Hz /6/. The frequency shift is calculated from $|D(t_4)$, $\omega-\omega_0=-W/2|^2$, $|D(t_4)$, $\omega-\omega_0=+W/2|^2$ and the derivative
$d/d\omega|D(t_4)$, $\omega-\omega_0=W/2|^2$.

3. Setup of analytical equations

According to the verbal description in the previous section and to Fig. 2, the algebraic form of the probability amplitudes is now set up in a time series.

$$
\begin{aligned}
t=0:\quad & V_k(0)=T_{km}\,V_m^0,\ k=-1,0,+1\\
& D(0)=0\\
t=t_1:\quad & V_1(t_1)=V_1(0)\exp\{-i\omega_{10}T_1\}\\
& V_0(t_1)=V_0(0)\\
& V_{\bar1}(t)=V_{\bar1}(0)\exp\{-i\omega_{\bar10}T_1\}\\
& D(t_1)=0\\
t=t_2:\quad & V_k(t_2)=V_k(t_1)\cdot N_k(t_1),\ k=-1,0,+1\\
& D(t_2)=Y_0(t_1)\cdot V_0(t_1)+Y_1(t_1)\cdot V_1(t_1)\\
& \quad +Y_{\bar1}(t_1)V_{\bar1}(t_1)\\
t=t_3:\quad & V_1(t_3)=V_1(t_2)\cdot\exp\{-i\omega_{10}T\}\\
& V_0(t_3)=V_0(t_2)\\
& V_{\bar1}(t_3)=V_{\bar1}(t_2)\cdot\exp\{-i\omega_{\bar10}T\}\\
& D(t_3)=D(t_2)\exp\{i\omega_0T\}\\
t=t_4:\quad & V_k(t_4)=V_k(t_3)\cdot N_k(t_3),\ k=-1,0,+1\\
& D(t_4)=D(t_3)(t_3)\cdot N_0(t_3)+Y_1(t_3)\cdot V_1(t_3)\\
& \quad +Y_0(t_3)\cdot V_0(t_3)+Y_{\bar1}(t_3)\cdot V_{\bar1}(t_3).
\end{aligned}
\tag{7}
$$

As described before, $D(t_4)$ has to be expressed by the initial probability amplitude V_m^0:

$$D(t_4) = V_m^0 \cdot [\mathrm{SUM1}+\mathrm{SUM2}] \tag{8a}$$

where SUM1 and SUM2 describe the transition amplitude to the (3,0) level in the first and second irradiation section, respectively.

It is

$$
\begin{aligned}
\mathrm{SUM1}=&(Y_0(t_1)\cdot T_{0m}+Y_1(t_1)T_{1m}\exp\{-i\omega_{10}T_1\}+\\
&Y_{\bar1}(t_1)T_{\bar1m}\exp\{-i\omega_{\bar10}T_1\})\exp\{i\omega_0T\}\\
\mathrm{SUM2}=&Y_1(t_3)N_1(t_1)\exp\{-i\omega_{10}(T_1+T)\}T_{1m}\\
&+Y_0(t_3)N_0(t_1)T_{0m}\\
&+Y_{\bar1}(t_3)N_{\bar1}(t_1)\exp\{-i\omega_{\bar10}(T_1+T)\}T_{\bar1m}\ .
\end{aligned}
\tag{8b}
$$

Common phase factors are split off by the definition of D':

$$D(t_4)=i^{-(m+1)}\exp\{i(\omega T_1-\phi_0+\omega\tau+m\phi+\omega_0T)\}\cdot D'(t_4) \tag{9}$$

In SUM1 and SUM2 one summand contains a factor T_{0m} and describes the strong transition amplitude $(4,0)\Rightarrow(3,0)$. It gets a second index "0" whereas the others are caused by the π-transitions and get a second index "1". D' can thus be expressed by

$$D' = S_{10} + S_{11} + (S_{20} + S_{21})\ \exp\{i\varepsilon T\} \tag{10}$$

where the S-terms are explicitely given by

$$
\begin{aligned}
S_{10}&=\sin 2b\tau\ \tilde T_{0m}/2, \qquad S_{20} = S_{10}\\
S_{11}&=-\Pi_1\ \tilde T_{1m}\ \cos b_\sigma\tau\ \exp\{-i\omega_{10}(T_1+\tau/2)\}\\
&\quad +\Pi_{\bar1}\ \tilde T_{\bar1m}\ \cos b_\sigma\tau\ \exp\{-i\omega_{\bar10}(T_1+\tau/2)\}\\
S_{21}&=-\Pi_{\bar1}\ \tilde T_{1m}\ \exp\{-i\omega_{10}(T_1+T+\tau)\}\\
&\quad +\Pi_{\bar1}\ \tilde T_{\bar1m}\ \exp\{-i\omega_{\bar10}(T_1+T+\tau)\})
\end{aligned}
\tag{11}
$$

and ε is defined as $\varepsilon = \omega-\omega_0$.

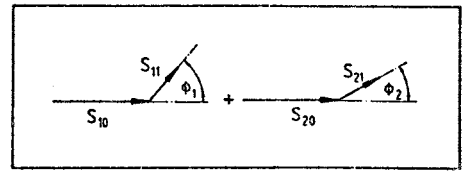


Fig. 3 Graphical representation of $D'(t_4)$ as expressed in (10) in the complex plane.

A graphical representation in the complex plane may help to identify the significance of the terms in equation (10).

If $\varepsilon=0$ and no phase difference is assumed between the two microwave fields, then S_{10} and $S_{20}\exp\{i\varepsilon T\}$ are collinear. S_{11} and S_{21} may have different angles ϕ_1 and ϕ_2 with respect to S_{10} and S_{20}, as illustrated in Fig. 3. If $\phi_2 \neq \phi_1$ then this is equivalent to a phase difference between the two irradiation sections with the consequence that the frequency of maximum transition probability is shifted by /7/.

$$\Delta\omega = (\phi_2-\phi_1)/T. \tag{12}$$

This straightforeward approach to calculate the frequency shifts was given up as the process of averaging over m and v proved to potentially lead to erroneous results in case that ϕ_2 or ϕ_1 become close to $\pm\pi/2$.

Instead $|D|^2$ is evaluated from (8), (10) and (11):

$$|D|^2 = |V_m^0|^2.\ (\Sigma_0 + \Sigma_1 + \Sigma_2) \tag{13a}$$

where

$$\Sigma_0 = \sin^2 2\ b_\sigma\cdot\tau\ \tilde T_{0m}^{\,2}\ (1+\cos\varepsilon T)/2 \tag{13b}$$

is the well known (4,0)-(3,0) transition probability in the near resonant approximation. Note that in case of no Majorana transitions being present, $T_{mm'}=\delta_{mm'}$. Σ_1 is a term which implicitely depends on ε due to the definition of the S-terms above:

$$
\begin{aligned}
\Sigma_1 &= \sin 2b\tau\ \tilde T_{0m}(|S_{11}|\cos\phi_1+|S_{21}|\cos\phi_2)\\
&\quad +|S_{11}|^2+|S_{21}|^2.
\end{aligned}
\tag{13c}
$$

Σ_2 also explicitely depends on ε:

$$
\begin{aligned}
\Sigma_2 &= \sin 2b\tau\ \tilde T_{0m}(|S_{21}|\cos(\varepsilon T+\phi_2)+|S_{11}|\\
&\quad \cos(\varepsilon T-\phi_1))\\
&\quad +2|S_{11}|\cdot|S_{21}|\cos(\varepsilon T+\phi_2-\phi_1).
\end{aligned}
\tag{13d}
$$

The significance of some of the summands in the equations (13c) and (13d) will be illustrated in the subsequent chapters.

4. Some remarks on the experiments previously made

In addition to the general remarks made in the introduction, a few more details will be

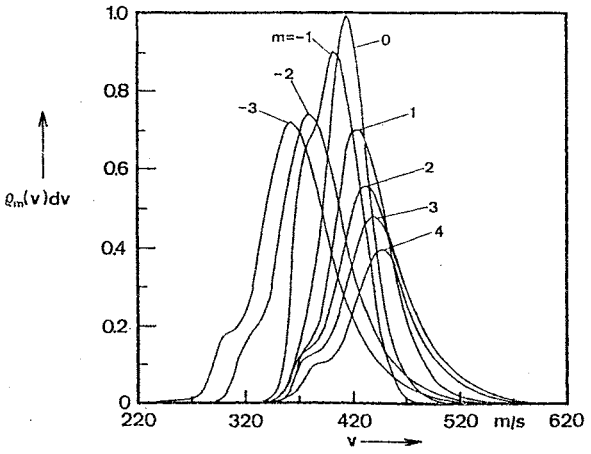


Fig. 4 Velocity distributions $\rho_m(v)dv$ of atoms in the state (4,m) in the CSX atomic beam.

given on the experiments, the results of which had been the motivation for this study. In the ref. /5/, included in this proceedings, the same experiments have been described from a slightly different point of view. The CSX is a conventional cesium beam frequency standard in which sixpole and fourpole magnets are used for the state selection. This kind of state preparation results in velocity distributions $\rho_m(v)dv$ for the beam atoms in the initial states (4,m) as shown in Fig. 4. It is this atomic population which is assumed in all steps of the averaging over m and v of $|D|^2$ (13).

The atomic beam is confined to the symmetry axis of the beam tube which is defined by that of the magnet systems. The cross-section is 3 mm in diameter. The symmetry axis of the coils used to produce the C field and the additional inhomogeneous fields as well as the symmetry line of the two cavity irradiation sections (cf. Appendix A2) are for the next assumed to be aligned with the beam axis.

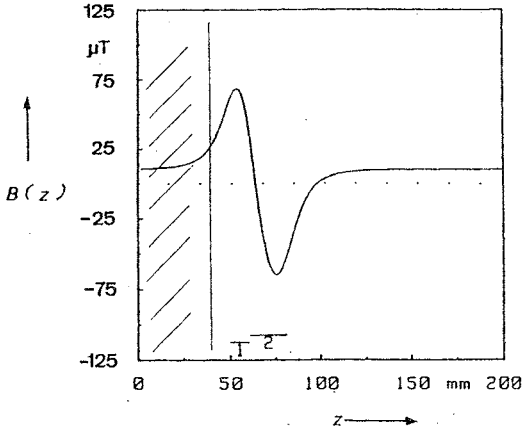


Fig. 5 Calculated magnetic field component $B_z(z)$ as a superposition of the constant C field and the fields produced by the coils 1 and 2. z = 0 corresponds to $t=-t^*$, z = 200 mm to t = 0.

In Fig. 5 the course of the longitudinal field is plotted which is traversed by the atoms between $t=-t^*$ and t=0. It has been found that a proper measure to characterize the passage of an atom through this kind of fields is the number of precessions which a classical magnetic moment with the Larmor frequency alike to that of cesium atoms would undergo inbetween the two zeros of the longitudinal field. In Fig. 6 the transition probability $\sin^2\alpha/2$ and the angle ψ, results of the numerical treatment (see Appendix A1), are plotted for atoms with v=427 m/s. The current in coil 2 (see Fig. 5) is increased whereby the integral field and the number of precessions, N_p, are increased, too. An increase of N_p, plotted on the horizontal axis, by 1 corresponds to the period found in $\sin^2\alpha/2$ and ψ as well.

From Fig. 4 the range of velocities can be inferred which has to be considered. At a fixed setting of the fields, $\sin^2\alpha/2$ is an oscillating function of v, however not strictly periodic /5/.

In the following section, results of computed frequency shifts are presented, obtained under different assumptions of the experimental conditions. In course of this presentation, some more experimental details will be given for to prove the validity of the assumptions made.

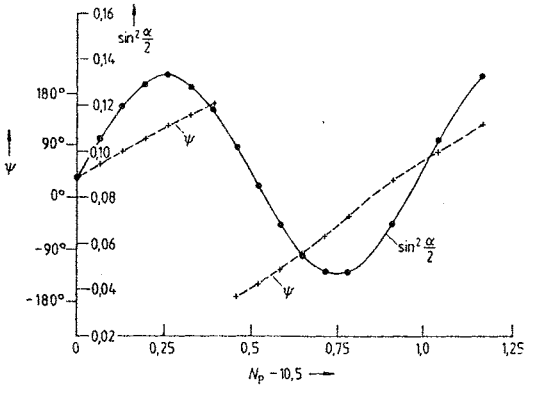


Fig. 6 Calculated transition probability, $\sin^2\alpha/2$, and angle ψ in (A1.1) of a spin 1/2 atom with the g-factor of cesium 1/4, velocity v = 427 m/s as a function of N_p, the number of precessions of the spin between the two zeros in $B_z(z)$ of the field shown in Fig. 5.

5. Computational results

5.1 Absence of π-transitions

If the amplitude of the transversal rf fields is set fixed to zero, then any shift of the clock transition frequency is absent: S_{11} and S_{21} (11) vanish. In dependence on the inhomogeneous fields, the velocity distribution of those atoms which finally are found in the (3,0) level is composed of contributions of all the initial distributions shown in Fig. 4 as a result of the term T_{0m} in S_{10} and S_{20} (11). Eventually, the mean velocity will

vary, but hardly more than by 5 % in relative units. The velocity dependent frequency shifts, present in a real device, will be changed accordingly, but hardly more than by 10^{-13} in case of the CSX.

5.2 Absence of Majorana transitions
An absence of Majorana transitions can be achieved by a proper shaping of the magnetic fields. Then $T_{m'm} = \delta_{m'm}$ and (13a) reduces to a sum of the contribution of atoms starting in the state (4,0) (term $\sum_0$ in (13a)) and (4,±1)($|S_{11}|^2$ and $|S_{21}|^2$ in (13c)). The resulting line shape is that of the unperturbed Ramsey line overlaid to the wings of the two neighbouring π-transitions. The frequency shift has the same signature as the Rabi-pulling, except that the B_0-dependent frequency shifts has twice the period of the corresponding Rabi-pulling /3,9/. It is proportional to $\langle|y|^2\rangle$ where $\langle|y|\rangle$ is the mean distance of the atomic trajectory from the symmetry line of the cavity. From the CSX beam geometry, a value of $\langle|y|\rangle = 0.64$ mm would follow. The calculated magnitude of the π-transition signal is then too large by a factor of ≈ 4 compared to the experimental findings in the CSX. Therefore, $\hat{b}_\pi$ is reduced by a factor of 2. The resulting frequency pulling due to the presence of π-transitions is illustrated in Fig. 7.

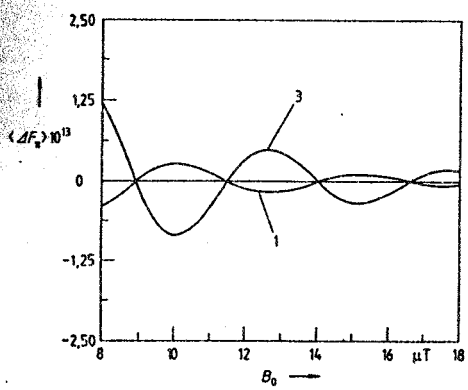


Fig. 7 Frequency shifts as a function of the C-field strength B_0 due to pulling by the wings of the neighbouring ΔF=1, Δm=±1 π-transitions in the CSX assuming square wave frequency modulation with the modulation width W/2 (1) and 3·W/2 (3) to detect the line centre.

5.3 Presence of all disturbances
The transversal rf fields seen by an atom which passes at $y=y_1$ or $y=-y_1$ through the irradiation section of the microwave cavity are out of phase by π. Therefore a factor of $\exp\{i\pi\} = -1$ has to be introduced in Y_1 and Y_1^* (5a) if $y = -y_1$ and finally also the angles ϕ_1 and ϕ_2 in Fig. 3 change their sign. In consequence, the resulting frequency shift is equal in magnitude but different in sign. The results presented now have thus been obtained under the assumption of a certain "misalignment". The same value <y> is assumed for both irradiation sections as if the centre of gravity of the atomic beam was away from the symmetry line by 0.32 mm. During the experimental work with the CSX frequency shifts have been found although the cavity

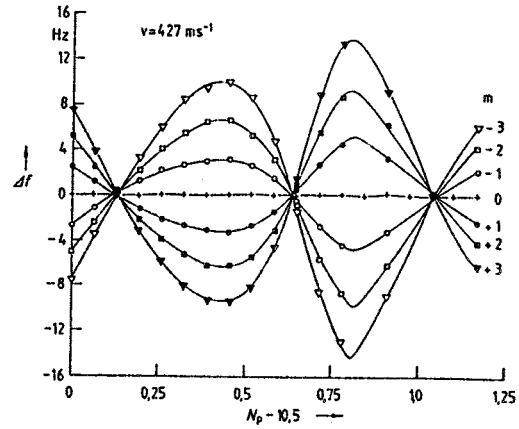


Fig. 8 Shift of the clock transition frequency for a monokinetic atomic beam of v=427 m/s and an initial population of the level (4,m) as a function of the traversed inhomogeneous fields which are characterized by N_p (cf. Fig. 6).

should have been aligned, <y>=0, with the atomic beam. The shifts, however, were noticably sensitive to a rotation of the cavity with respect to the beam axis.

To begin with, in Fig. 8 the calculated frequency shift is plotted for a monokinetic beam of v=427 m/s and initial states of the atoms as indicated. The variation of N_p has been described in the previous section. Again, the full period of the frequency shifts is closely corresponding to an increase of N_p by one. The frequency shift is almost proportional to m. A very small deviation from the proportionality is caused by the quadratic term in the Breit-Rabi formula thus $|\omega_{10}|$ being slightly different from $|\omega_{\bar{1}0}|$ (1).

In the next step it is illustrated how the frequency shift for a given m (here m=-1) and a given field setting depends on the

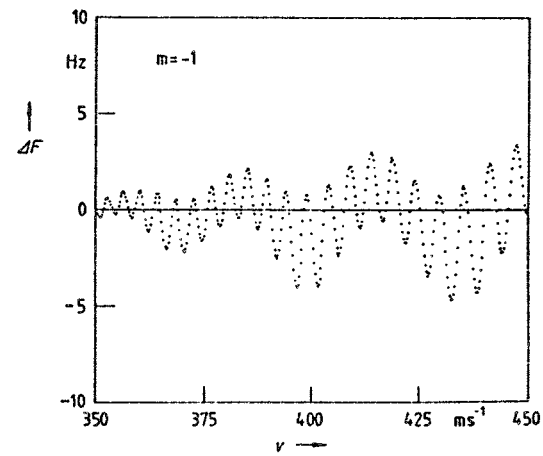


Fig. 9 Shift of the clock transition frequency for a given setting of the magnetic fields and a beam of atoms in the state (4,-1) as a function of the atomic velocity v.

atomic velocity v. In Fig. 9 a look is made on the velocity region around 400 m/s. The dots indicate the resolution in the verlocity space. The v-dependence is threefold: The coefficients $\tilde{T}_{0m}$, $\tilde{T}_{1m}$ and $\tilde{T}_{\bar{1}m}$ are velocity dependent, as is the quantity $\sin^2\alpha/2$. The narrow period to be seen is due to the term $\cos(\omega_{10}(T_1+T+\tau/2))$ in S_{21} whereas a wider period is due to the term $\cos(\omega_{10}(T_1+\tau/2))$ in S_{11}. An average over the velocities found in the atomic beam (cf. Fig. 4) will reduce the frequency shifts considerably.

The result of this average over the population of the atomic beam is shown in Fig. 10. The calculation predicts a frequency shift $\langle\Delta F\rangle_{m,v}$ of several 10^{-12} under the following additional assumptions: B_0=10.4 μT, $\omega_{10} = 2\cdot\pi\cdot 36373$ Hz, b_σ according to optimum excitation condition, square-wave frequency modulation of the central Ramsey fringe. The N_p is calculated for the average atomic velocity of 400 m/s, although now all velocities do contribute. The steps merely correspond now to the change of the field as the current in coil 2 was increased (see Fig. 7). The period in $\langle\Delta F\rangle$ has no easily explainable relation to N_p. This is in contradiction to the previous experimental findings. ΔF_{CSX} had been measured when the fields applied were shaped in the same way as it was tried to model them in the calculations.

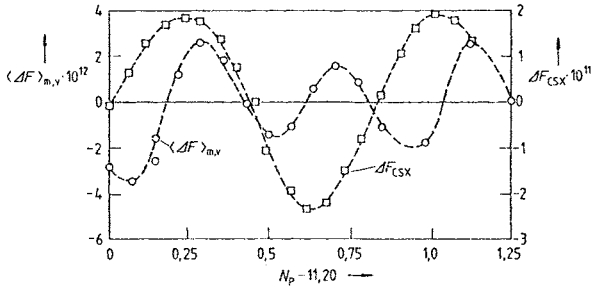


Fig. 10 Shift of the clock transition frequency as a function of the traversed inhomogeneous fields: calculated average over the initial population in the atomic beam from Fig.4, experimental data: ΔF_{CSX} = F(CSX)-F(CS1).

6. Discussion

The work described here had been motivated by the alarming magnitude of frequency shifts detected in previous experiments with the PTB's CSX. Therefore it was tried to model the experimental conditions rather closely. For the time being, no general rule can be given on how to state the contribution of this kind of potential frequency shifts in the atomic clocks's uncertainty budgets. Nevertheless, the key to the treatment is to realize that it is the occurence of Majorana transitions and ΔF=1, Δm=±1 π-transition in the beam tube, where the coherences created are long-living compared with all flight-times. The seminal idea to the intuitive understanding is that for each atom leaving the first state selector, several ways are open to reach the final quantum state. As the ways are indiscernible, the probability amplitudes rather than the squares have to be added. Of course, many of the interference terms are found to be of negligible contribution. An average over atomic velocities, populations and trajectories has to be made, including an average over the values of the magnetic field B_0 which is not perfectly homogeneous across the beam. In the CSX, the field is generated by a long solenoid so that a very small variation of B_0 can be suspected in the radial direction.

Until now, only a special shape of the inhomogeneous static magnetic field has been trated. But some relevant general features have been identified:

i) Atoms leaving the polarizer in the state (4,0) do not contribute to frequency shifts of relevance.
ii) In case of an identical population of the Zeeman states in the initial beam, identical with respect to total number, velocity and trajectories, no frequency shift occurs.
iii) The same is true if the trajectories of the atoms are distributed symmetrically around the symmetry line of the microwave cavity, irrespective of m and v.

The experiments previously made are not yet described satisfactorily by the theoretical treatment done until now. Two necessary steps to improve the calculations have been identified, others may follow:

1) The CSX detector is sensitive to atoms in the (4,m) quantum states. An equivalent relation for the $|V_m(t_4)|^2$ has to be set up, replacing the equation (13).

2) The analyzer magnet does not focus all atoms in the state (4,0), for example, onto the detector, if their velocities differ considerable from the designed optimum velocity of 410 m/s. Approximate transmission functions are currently evaluated using beam trajectory calculations.

7. Bibliography

/1/ N.F. Ramsey, Rev.Mod.Phys. 62, 1990, p. 541-552.

/2/ E. Majorana, Nuovo Cimento 9, 1932, p. 43-50.

/3/ L. Cutler et al. J. of Applied Physics 69, March 1991

/4/ A. Bauch, PhD-thesis, published as PTB-Bericht, PTB-Opt-20, 1986.

/5/ A. Bauch, this proceedings

/6/ T. Heindorff et al., PTB-Mitteilungen 94, 1984, p. 318-326.

/7/ N.F. Ramsey, H.B. Silsbee, Phys.Rev. 84, 1951, p. 506-507.

/8/ J. Vanier et al., Metrologia 20, 1984, p. 101-105.

/9/ A. de Marchi et al., Metrologia 20, 1984, p. 37-47.

/10/ F. Bloch, I.I. Rabi, Rev.Mod.Phys. 17, 1945, p. 237-244.

Appendix A1

The equation (A1.1) yields the basic relation to deduce the transition amplitudes $T_{mm'}$ from the experimental situation. The angles α, ϕ and ψ are independent of the angular momentum F /10/:

$$T_{mm'}=[(F+m)!(F+m')!(F-m)!(F-m')!]^{1/2}\cdot$$
$$i^{2F-m-m'}\exp\{i(m'\phi+m\psi)\}\ \sin^{2F}(\alpha/2). \qquad (A1.1)$$

$$\sum_k (-1)^k \frac{\cot an^{m+m'+2k}(\alpha/2)}{(m+m'+k)!(F-m'-k)!(F-m-k)!k!}$$

In case of F = 1/2, the relations are

$$T_{-+} = i\ \exp\{i(-\psi+\phi)/2\}\ \sin\ \alpha/2$$
$$T_{++} = i\ \exp\{i\ (\psi+\phi)/2\}\ \cos\ \alpha/2. \qquad (A1.2)$$

A numerical treatment of the passage of an atom with velocity v, F=1/2 and g-factor g=1/4 through the applied inhomogeneous fields is made. It is described in some detail elsewhere /4,5/. Starting with an atom in the +state, i.e. C_+=1 at t=-t*, the complex amplitudes $C_+(t=0)$ and $C_-(t=0)$ are obtained (cf. Fig. 2). Together with (A1.2) this allows the determination of α, ϕ, ψ. Inserted back into (A1.1), now with F=4, a set of transition amplitudes is obtained which couples the (4,m') level to all (4,m) levels, depending on the atomic velocity and the fields traversed between t=-t* and t=0. Throughout this paper it is usually dealt with

$$\tilde{T}_{mm'} := T_{mm'}\cdot \exp\{-im'\phi\}\ (i)^{m'+m} \qquad (A1.3)$$

instead of $T_{mm'}$.

Appendix A2

The transition amplitudes Π_1 and $\Pi_{\bar{1}}$ are derived in three steps:
a) The interaction hamiltonian couples the levels (4,0) (4,±1) and (3,0) according to

$$\langle 3,0|\hat{H}_1(t)|3,0\rangle = g\ \mu_0\ B_z(t)/2$$
$$\langle 3,0|\hat{H}_1(t)|4,\pm1\rangle = \pm\sqrt{15}\cdot g\cdot\mu_0\cdot B_y(t)/8\ . \qquad (A2.1)$$

where the state vectors in the low field approximation have been taken from /8/.

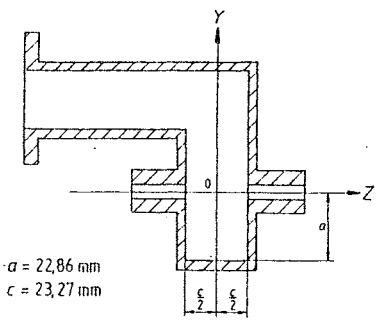


Fig. A1 Sketch of the end section of the Ramsey microwave cavity used in the CSX.

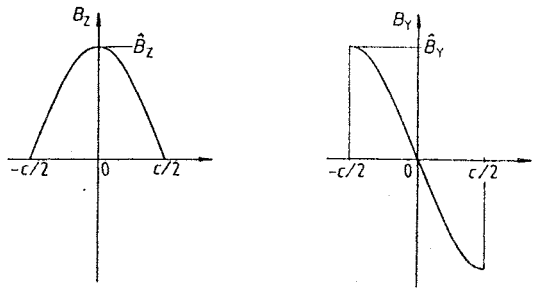


Fig. A2 Shape of the longitudinal and transversal rf magnetic field components in the irradiation section of the cavity structure from Fig. A1.

b) In Fig. A1 a sketch of the end section of the Ramsey cavity used in the CSX is shown. In Fig. A2 the time dependence of the rf magnetic field amplitude is illustrated for an atom which passes the irradiation section along a trajectory parallel to the symmetry line y=0. The fields B_z and B_y are given by

$$B_z = \hat{B}_z\ \cos(\pi\cdot y/a)\cdot\cos(\pi z/c)$$
$$B_y = \hat{B}_z\ \sin(\pi\cdot y/a)\cdot\sin(\pi z/c)\cdot(a/c) \qquad (A2.2)$$

The maximum value of the excitation parameter, $\hat{b}_\sigma$, is determined from the optimum excitation condition of the clock transition. With the help of (A2.1) and (A2.2), $\hat{b}_\pi$ is calculated.

c) The transition amplitude is then found as the Fourier transform of the time dependent excitation parameter, $b_\pi := b_\pi(t)$ as it is described in /9/. The result is

$$\Pi_1 = 2\cdot\Omega_1\ \hat{b}_\pi\cdot\tau^2\ \cos(\Omega_1\tau/2)/(\Omega_1^2\tau^2-\pi^2) \qquad (A2.3)$$

where $\Omega_1 = \omega_{10}+\omega-\omega_0$, i = 1, -1.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 1A Plenary Number 4

RECENT PROGRESSES IN THE Mg BEAM FREQUENCY STANDARD

A. Godone - C. Novero

Istituto Elettrotecnico Nazionale "G. Ferraris" - Torino, Italy

Abstract

We report in this paper the present status of the experiments aiming to increase the stability and to improve the accuracy of the Mg frequency standard prototype. In particular we shall describe the optical pumping experiments which allow an increase by an order of magnitude of the short-term frequency stability and the solid state frequency multiplication chain which can remove the main technical problems limiting the standard accuracy well below its theoretical capabilities.

Introduction

The stability of the Mg beam frequency standard is mainly limited by the S/N ratio in the detection of the clock transition: a significant background-noise contribution comes from an incomplete population inversion just before the interrogation region. The feasibility of optical pumping techniques to overcome the problem has been already demonstrated |1|; other different approaches will be discussed in this work.
As regards the accuracy, it is at present limited at the level of 10^{-11} by the use of a BWO, phase-locked to a high spectral purity quartz, for the generation of the microwave radiation |2|; moreover the magnetic field dispersed by the strong BWO magnet puts a limitation to the definition of the C-field in the interaction region, thereby affecting directly the accuracy. Work has begun recently to build up a solid state chain for the generation of the 600 GHz signal, based on a 100 GHz Gunn oscillator and solid state multipliers.
We shall report here the present status of the experiments aiming to reduce the above limitations: optical pumping, beam diagnostic and solid state fequency synthesis up to submillimetric wavelengths.

Optical pumping

Several possibilities exist to fully deplete the $^3P_1$ state in a Mg metastable beam before the excitation of the clock transition $^3P_0 - {}^3P_1$: the optical transitions at 333.2 nm, 383.2 nm, 277.7 nm and 517.4 nm (see Fig. 1) are well suited for optical pumping of the $^3P_1$ state. All of them have a spontaneous decay time constant lower than 10 ns, which guarantees a large enough number of cycles for the pumping process during the atom transit-time through a low intensity laser field. The wavelengths 333.2 nm and 383.2 nm lie in the range of the second harmonic of existing semiconductor lasers and will represent the most interesting solution, from the technical point of view, if efficient second-harmonic generation (SHG) is

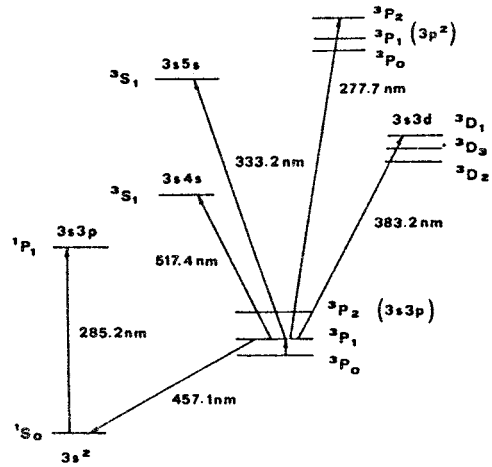


Fig. 1 - Energy level diagram of Mg.

achieved at the above wavelengths. The 383.2 nm line, moreover, may be excited by a frequency doubled Ti: sapphire laser (up to 200 mW). The 277.7 nm and 517.4 nm wavelengths may be generated by a frequency-doubled Rhodamine 110 dye-laser and by a Coumarin 6 dye-laser respectively.
In fig. 2 the set-up used for the first optical pumping experiments is reported; in this case the $^3S_1(3s4s) - {}^3P_1(3s3p)$ line at 517.4 nm has been used (see also |1|). The Ramsey signal corresponding to the $^3P_1 - {}^3P_0$ $\Delta m_j = 0$ microwave-induced transition is reported in fig. 3 with (upper trace) and without (lower trace) optical pumping; an increase by a factor 4.8 is observed for the peak value of the central Ramsey fringe. A reduction of the background noise level by a factor 2 has also been measured, limited by the ^{26}Mg and ^{25}Mg fluorescence emissions. An improvement by an order of magnitude of the S/N ratio in the detection of the clock transition is then obtained (S/N = 2,500; BW = 1 Hz), which implies a short-term frequency stability of $\sigma_y(\tau) = 8\times10^{-13}\,\tau^{-1/2}$ for the frequency standard. The laser intensity required to achieve an optical pumping efficiency of 90% is 5 mW/cm^2.
The results obtained are interesting also from another point of view, in fact it may be demonstrated |3| that the ratio between the peak values of the central Ramsey fringe with and without optical pumping is related in a definite way to the $^3P_1$ decay time constant ; the value obtained from these experiments turns out to be: $\tau = 5.1 \pm 0.7$ ms.
Optical pumping of the $^3P_1$ state has been obtained also with the line at 277.7 nm; the near UV radiation was generated by a R110 dye-laser frequency doubled through an intracavity KDP crystal, providing 1.5 mW

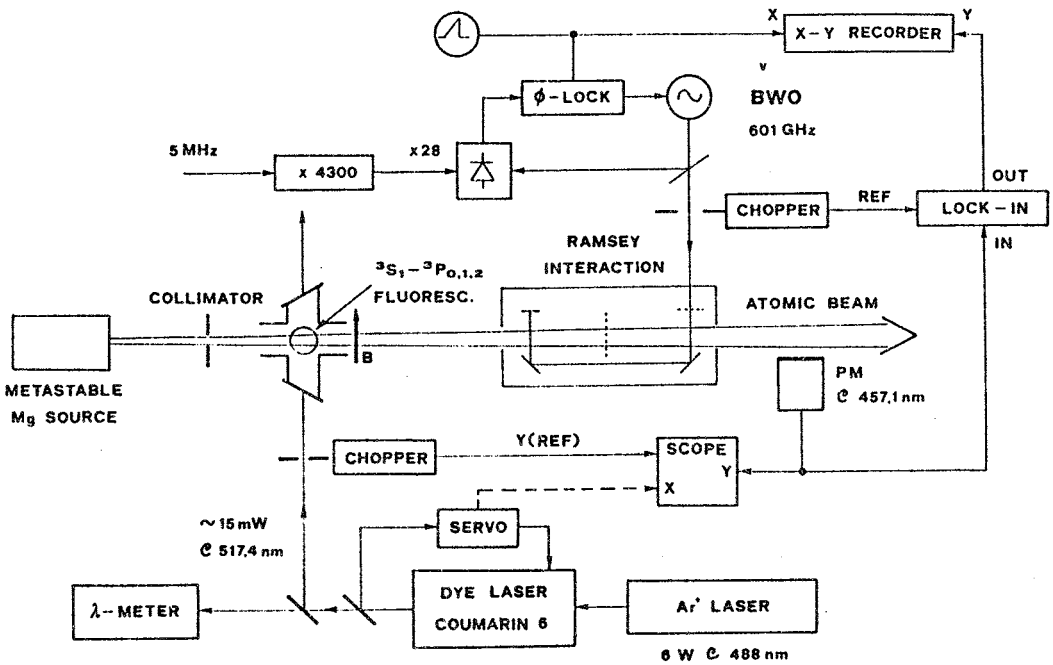


Fig. 2 - Experimental set-up for the optical pumping experiments.

at the required wavelength: an optical pumping efficiency of 50% has been reached in this case.

Atomic beam laser-diagnostic

In fig. 4(a) a record of the Mg $^1S_0 - ^1P_1$ resonance transition at 285 nm is reported (lower trace) together with the R115(20-1) I_2 linear absorption line (upper trace); the atoms effused by the oven are not excited to the metastable states in this case (discharge current $I_s = 0$). The resonance is observed through the fluorescence light emitted in the interaction region between the laser radiation and the atomic beam, 0.5 m downstream from the oven exit. The frequency separation between the ^{24}Mg and the I_2 line is 1 ± 0.10 GHz, in good agreement with the value reported in |4|, and the absolute wave-number turns out $\lambda^{-1} = 35051.263 \pm 0.008$ cm^{-1} |5|.

The broad pedestal observed in the record is the Doppler profile of the Mg gas present in the interaction region; its frequency width (FWHM) is 3.2 GHz, corresponding to a temperature lower than that of the oven (520°C or 4 GHz).

The record of Fig. 4(b) is obtained in the same condition of fig. 4(a) but with a discharge current I_s = 380 mA for the excitation of the atoms to the metastable triplet. The peak reduction observed in this case is directly related to the fraction of atoms excited to the ^3P states allowing to evaluate the efficiency η of production of metastable atoms. The fluorescence peak intensity V_F of the resonance line and the corresponding η values are reported in fig. 5 versus the discharge current; the efficiency curve is in good agreement with that reported in |6|. It is worth noting that the maximum efficiency is near the theoretical value and that the gas of Mg in the ground state does not affect the metastable beam.

Solid-state chain up to 600 GHz

In order to achieve the accuracy level of 1×10^{-13} estimated for the Mg frequency standard |7| it is necessary to remove the technical problems related to the use of a BWO for the generation of the interrogating signal at 600 GHz, mainly the intense stray magnetic fields and the cycle-slipping of the BWO phase-lock |8|.

To this end we are setting up an all-solid-state frequency multiplication chain from 5 MHz up to 601.3 GHz, whose block diagram is shown in fig. 6. Recent technological advances have allowed the development of highly efficient varactor frequency multipliers in the millimetric and submillimetric regions |9|. An InP Gunn diode delivering 75 mW at 100 GHz drives a varactor balanced doubler whose output power is 20 mW; the available power at 200 GHz is high enough to drive a varactor tripler which delivers about 0.3 mW at

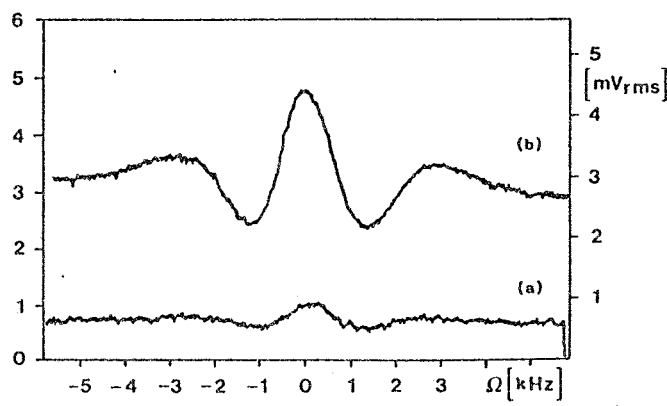
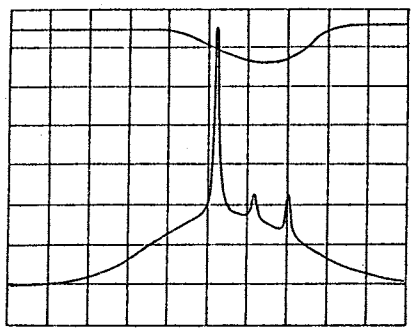
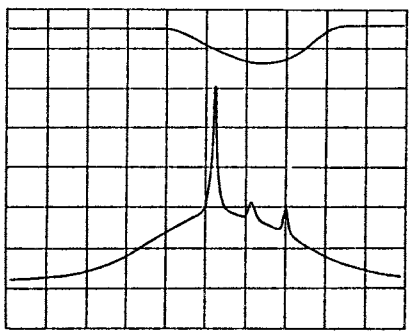


Fig. 3 - $^3P_1 - ^3P_0$ $m_j = 0$ Ramsey signal of ^{24}Mg: a) without optical pumping; b) with optical pumping. The vert. axis is normalized to the peak-value of curve a).



4(a)



4(b)

Fig. 4 - $^1S_0$ - $^1P_1$ resonance transition (lower traces) of Mg at 285 nm; R115(20-1) absorption of I_2 (upper traces). a) discharge current I_S = 0, b) I_S = 380 mA.

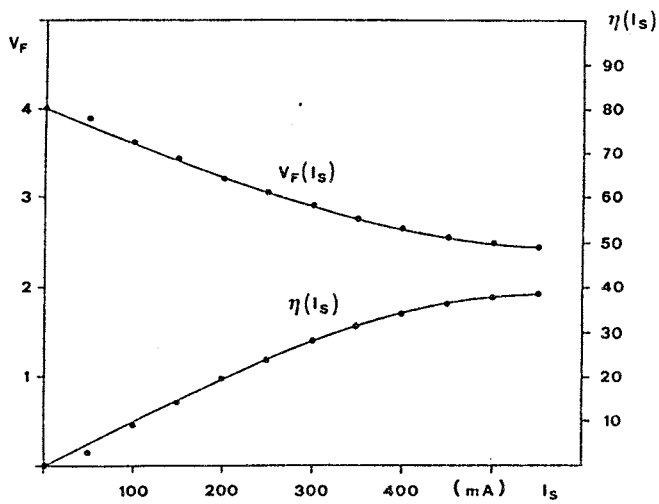


Fig. 5 - Fluorescence peak of the resonance line and efficiency of production of metastable atoms as a function of discharge current I_s.

601.3 GHz. The optimum power required, with the Ramsey cavity used to induce the clock transition, is about 130 μW |10|, less than the output power of the solid state multiplier chain.
Work is now in progress in order to phase-lock the 100 GHz Gunn oscillator to the X-band reference signal and then to remeasure the Mg frequency with respect to the Cs standard.

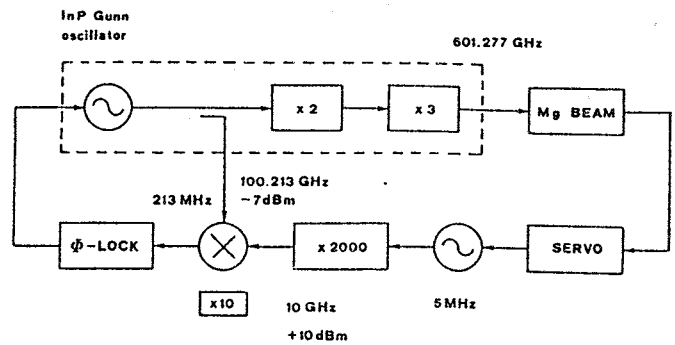


Fig. 6 - Block-diagram of the solid-state multiplication chain (5 MHz - 601 GHz).

Conclusions

We have reported in this paper the recent advances towards the realization of a frequency standard based on the fine structure transition $^3P_1$ - $^3P_0$ $\Delta m_j = 0$ of ^{24}Mg. An improvement of an order of magnitude has been demonstrated for the short-term frequency stability via optical pumping on the $^3S_1$ - $^3P_1$ green transition. Laser diagnostic of the atomic beam has been used to acquire more details about the metastable beam. Finally a solid state multiplier chain from 5 MHz to 601.3 GHz has been described which is supposed to eliminate some important obstacles to the improvement of the Mg standard accuracy up to the 10^{-13} region.

References

|1| A. Godone, C. Novero and E. Bava: Optical pumping of the metastable states of Mg - IEEE Trans. Instr. Meas., April 1991.

|2| A. Godone, E. Bava and C. Novero: Phase-lock of submillimetric backward-wave-oscillators - IEEE Trans. Instr. Meas., IM-38, pp. 794-798, 1989.

|3| A. Godone and C. Novero: Lifetime measurement of the Mg I intercombination line - Submitted for publication.

|4| N. Beverini et al.: Wavelength, isotopic shift, and transition rate of the Mg I resonance line - Optics Communications, 77, pp. 299-302, 1990.

|5| S. Gerstenkorn and P. Lue: J. Physique, 46, 867, 1985.

|6| G. Giusfredi, A. Godone, E. Bava and C. Novero: Metastable atoms in a Mg beam: excitation dynamics and velocity distribution - J. Appl. Phys. 63, pp. 1279-1285, 1988.

|7| A. Godone, E. Bava and C. Novero: Mg beam frequency standard - Frequency Standards and Metrology (Ed. A. De Marchi) Springer Verlag, Berlin, Heidelberg, pp. 78-83, 1989.

|8| A. Godone, E. Bava and C. Novero: High resolution frequency measurement of the $^3P_1$ - $^3P_0$ ^{24}Mg transition - Metrologia 24, pp. 133-138, 1987.

|9| Millitech Corp. Mod. MUS-601-6.

|10| E. Bava, A. Godone and G. Rietto: Rabi and Ramsey interrogations of metastable beams - Appl. Phys. B 41, pp. 187-196, 1986.

Session 1A Plenary Number 5

Determining the Effects of Microwave Power Variations on the Output Frequency of a High-Performance Cesium Frequency Standard

S. K. Karuza, W. A. Johnson, J. P. Hurrell, F. J. Voit, and A. Leong

Communications Electronics Department, The Aerospace Corporation
P. O. Box 92957, Los Angeles, CA 90009 USA

Abstract

Previous investigations of cesium (Cs) frequency standards demonstrated that Rabi pulling and cavity pulling were the most significant effects that translated microwave power variations into output frequency variations. It was also shown that there exist optimum values of the C-field that make the output frequency insensitive to variations in microwave power. Using a completely automated measurement system, we performed a similar study on a dual-beam, high-performance Cs frequency standard. The microwave power was precisely calibrated and the effects of microwave cavity tuning on the output frequency were measured.

1. Introduction

Previous investigators presented experimental evidence [1,2], based on one manufacturer's cesium (Cs) frequency standards, confirming that Rabi pulling and cavity pulling are the major transducing effects that turn microwave power variations into output frequency variations. It was shown that there were C-field values for which the output frequency of the Cs standard is insensitive to changes in microwave power. In our earlier studies [3,4] we performed these C-field experiments on Cs clocks made by other manufacturers. We found C-field curves that were somewhat different from those found by the earlier investigators.

In this study we made C-field measurements on a Cs standard that was from the same manufacturer as the clocks measured by those earlier investigators. We should point out that this particular standard is an older model and may not be representative of the manufacturer's present design. We should also make clear that it is the only Cs standard from this manufacturer that we have measured at this time.

The C-field frequency dependence described in this paper are for the sixth Cs standard measured in our laboratory. Previously we measured two each from two manufacturers and one from a third. The results presented herein are for a Cs standard from a fourth manufacturer. As in the case of the other manufacturers' standards, this standard also had to be modified to provide access to the microwave power and the C-field circuitry.

2. Measurement System

Figure 1 is a block diagram of the complete measurement system. Measurements were automated to allow for the careful sequencing of data gathering without having to adjust or change any microwave connections manually. Both of the parameters that are varied, namely the C-field current and the microwave power, are computer controlled; the current is set by a precision constant-current generator and the microwave power is changed by a calibrated PIN diode attenuator. Good microwave matching at the 9.192-GHz interfaces into and out of the standard has been attained through the use of tuners and isolators at each interface. An amplifier has been added to compensate for the losses through the various components. The entire system is controlled by a Hewlett-Packard (HP) series 300 computer, which also acquires and processes the data.

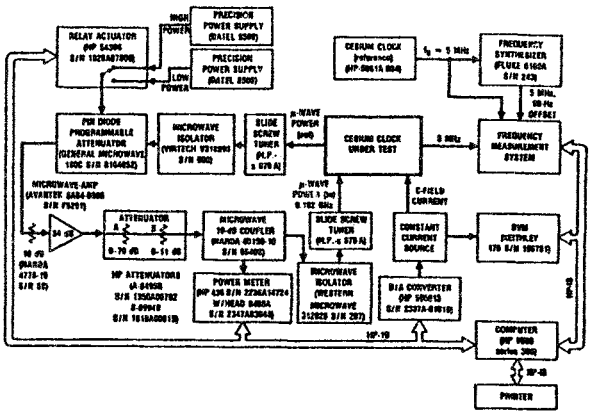


Figure 1. Block diagram of the C-field measurement system for a Cs frequency standard

Figure 2 is a block diagram of the frequency measurement system. The frequency reference for both the Fluke synthesizer and the HP counter is an HP model 5061A-004 Cs frequency standard. Figure 3 shows a plot of the noise floor of the frequency measurement system and a plot of the frequency stability $\sigma_y(\tau)$ of the Cs standard used in our measurements with the HP frequency standard as the reference.

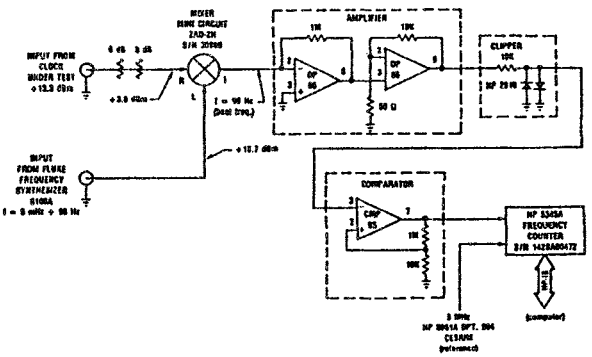


Figure 2. Circuit diagram of the single-mixer frequency-measurement system used to determine the fractional frequency changes at different C-field settings

Before the data are taken, the microwave cavity of the Cs tube must be tuned. Figure 4 is a block diagram of the system that measures the return loss of the microwave cavity in the Cs tube. The microwave tuning is adjusted to obtain a maximum return loss at the resonant frequency f_0 of 9.192631770 GHz, as shown in Figure 5. The input microwave power is then adjusted to maximize the output current from the beam tube, as shown in Figure 6. The resulting microwave power is called the optimum power P_o.

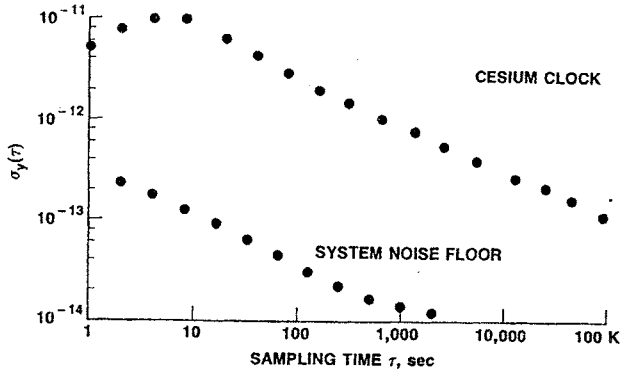


Figure 3. The measured square root of the Allan variance of the Cs frequency standard used in our measurements and the noise floor of the frequency measurement system

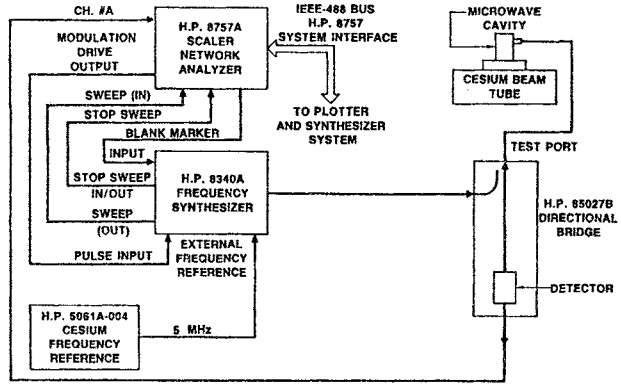


Figure 4. Block diagram of the Cs beam tube microwave-cavity measurement system

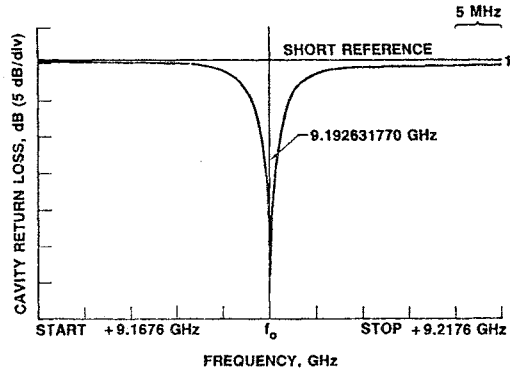


Figure 5. Plot of the microwave cavity return loss of the measured Cs frequency standard tuned to a Cs resonance frequency (f_0)

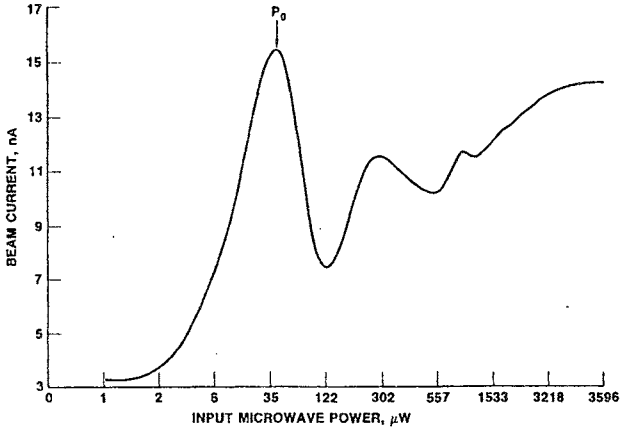


Figure 6. Plot of the Cs beam current vs. input microwave power of the Cs frequency standard tuned to f_0

A typical data-taking sequence consists of the following steps:

1. Set the C-field current at some low value (typically from 6 to 8 mA) and the microwave power at some value (e.g. at the optimum value P_0).
2. Measure the beat frequency over some long averaging time T (typically 3000 sec).
3. Change the microwave power level (e.g. to P_0 + 1 dB).
4. Measure the beat frequency over T again.
5. Increase the C-field current by some programmed amount (typically 0.5 mA).
6. Measure the beat frequency over T again.
7. Change the microwave power back to the initial value.
8. Repeat steps 2 through 7 until the final C-field current (typically 20 to 25 mA) is reached.

3. Measurement Results

In order to plot the resulting C-field data as a function of the Zeeman frequency, measurements were made on the upper and lower microwave transitions to obtain the relationship between the Zeeman frequency f_Z and the applied C-field current I_C. A linear fit to the data yielded the relation

$$f_Z \text{ (kHZ)} = 0.115 + 3.300 \times I_C \text{ (mA)}$$

with a standard deviation of 29 Hz.

For each C-field current a data point is calculated as the difference in beat frequency between the frequency at the higher power and that at the lower power, both of which are normalized to the nominal output frequency (5 MHz). In other words,

$$\text{ordinate} = (\bar{f}_H - \bar{f}_L)/5 \times 10^6$$

where $\bar{f}_H$ is the average beat frequency for the higher microwave power and $\bar{f}_H$ is the average output frequency for the lower microwave power. The reference frequency is set higher than the unknown frequency. Consequently, an increase in beat frequency corresponds to a reduction in the frequency of the standard being investigated.

Figure 7 is a plot of the average of the data runs on the standard after the tube's microwave cavity was tuned to f_0. Data were taken for power changes of both +1 dB and +3 dB above P_0. For the +1-dB data, each and every frequency measurement was averaged for a time that ranged from 17,300 to 45,300 seconds, depending on the data point. The associated standard deviations were between 2.7 and 1.7 x 10^{-13}. For the +3-dB data, each frequency measurement was averaged for a time that ranged from 7,000 to 9,000 seconds, again depending on the data point. The standard deviations in this case are between 4.2 and 3.7 x 10^{-13}.

The +1-dB data shown in Figure 7 exhibit a zero crossing at about 25 kHz and a maximum $\Delta f/f_0$ of about 1.7 x 10^{-12} at 31 kHz. For the +3-dB data there is a zero crossing near 26 kHz and possibly one at about 45 kHz; the maximum $\Delta f/f_0$ of 5.7 x 10^{-12} also occurs at 31 kHz. The frequency offset tends to decrease as the Zeeman frequency (and, hence, the resonance-line separation) increases. These results are similar to those found in our earlier studies [3,4] on Cs standards made by other manufacturers.

The effect of cavity pulling on the C-field frequency dependence was also measured. Using the test setup depicted schematically in Figure 4 to measure return loss, we varied the slide-screw tuner until the tuner and tube were tuned to 5 MHz above f_0. The shape of the return-loss curve was essentially identical to that of Figure 5, except that the maximum return loss was at f_0 + 5 MHz. The C-field data were then taken for a power change of +3 dB. Then the cavity was tuned to 5 MHz below f_0 and the measurements were repeated. The results are shown in Figure 8. The averaging times T ranged between 6,000 and 9,000 seconds, and the standard deviations were between 4.6 and 3.7 x 10^{-13}.

The collapse of the frequency offsets is interesting and is entirely consistent with the predictions for a monovelocity beam. The 5-MHz cavity offset caused a 6-dB reduction in cavity microwave power. This should reduce the magnitude of the measured Rabi pulling by a factor of 18 and the cavity pulling by a factor of 7. We raised the microwave power level by 6 dB and repeated the measurement for a power change of +3 dB, while maintaining the cavity offset at 5 MHz. We recovered the original Rabi pulling seen in Figure 7 and acquired a cavity pulling of 1 x 10^{-12}. This corresponds to a coefficient of about 1.5 x 10^{-13} per dB per MHz under the original conditions of P_0 and zero cavity offset.

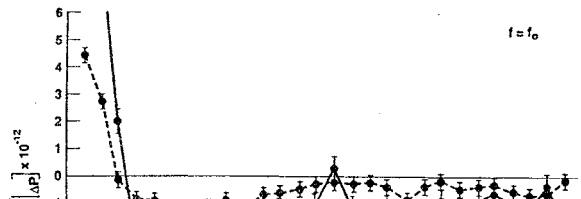


Figure 7. Average of final data regarding the difference of the average frequencies as a function of C-field for a microwave power change of +1 and +3 dB above the optimum power level P_0

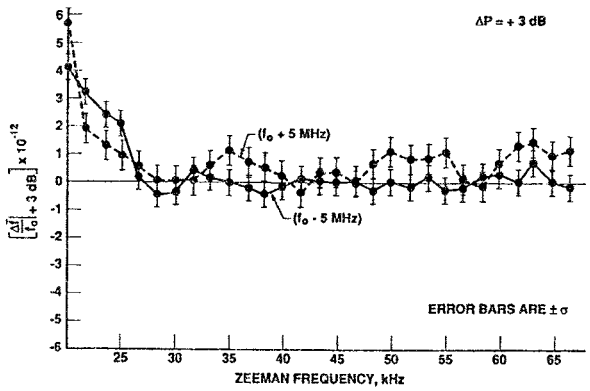


Figure 8. Plot of the difference of the average frequencies as a function of the C-field of the Cs frequency standard for two microwave cavity tuned frequencies (f_0 and ±5 MHz) for a microwave power change of +3 dB above the optimum power level P_0

4. Inaccuracies in the Measurement System

Our measurement system introduces three sources of error or uncertainty: (1) frequency measurement noise, (2) C-field current setting errors, and (3) microwave-power measurement errors. The first contribution has been shown (see Figure 3) to be almost two orders of magnitude below the standard deviation of the data. The uncertainty in the C-field setting is probably on the order of parts in 10^4 in our laboratory environment over the three months during which data were taken; this stability is largely set by the stability of a precision film resistor. The third source of uncertainty, the measurement of the microwave power, is the most difficult of the three to assess. For a microwave source and measurement system that is very similar to that of Figure 1, it was found [5] that the standard deviation of microwave power was between 0.04 and 0.06 dB for measurement intervals between 0.14 and 2.5 days. In addition, the drifts in the power measurement system (consisting of t power meter, the power measurement head, and the calibrator) we examined separately. In two long runs of 10 and 16 days each, the drifts were found to have power-measurement standard deviations 0.0012 and 0.0025 dB, respectively.

Conclusions

This paper has presented the results of a measurement of the se sitivity of output frequency to variations in microwave power, as a fur tion of C-field, on a particular Cs frequency standard. Although th standard may not necessarily reflect the manufacturer's most curre technology, it is nevertheless a high-performance unit. The resu showed that Rabi pulling in this standard was similar to the offsets measured in our laboratory on other Cs standards from different ma ufacturers. In particular, the Rabi pulling becomes extremely small the vicinity of 45 kHz and remains less than 1×10^{-12} per dB for high Zeeman frequencies. In addition, we estimate the cavity pulling at op mum power to be 1.5×10^{-13} per dB per MHz offset.

Acknowledgments

The authors are grateful to Ronald L. Beard, Dr. Joe White, a Frederich Danzy of the Naval Research Laboratory for supplying the frequency standard used in our studies. The authors also thank M chael T. Meyer of The Aerospace Corporation for editing and prepari this paper.

References

[1] A. De Marchi, "New Insights into Causes and Cures of Frequency Instabilities (Drift and Long-Term Noise) in Cesiu Beam Frequency Standards," *Proc. 41st Frequency Control Sym posium*, Philadelphia, Pa. (1987), pp. 54-58.

[2] A. De Marchi, "Rabi Pulling and Long-Term Stability in Cesium Beam Frequency Standards," *IEEE Trans. Ultrasonics Ferroelectronics, and Frequency Control* **UFFC-34** [6], 598-601 (November 1987).

[3] S. K. Karuza, W. A. Johnson, and F. J. Voit, "Determinin the Effects of Microwave Power and C-field Settings on the Fr quency of a Cesium Atomic Frequency Standard," 3rd European Frequency and Time Forum (21-23 March 1989), Be sançon, France.

[4] S. K. Karuza, W. A. Johnson, J. P. Hurrell, and F. J. Voit, "Determining Optimum C-field Settings that Minimize Outpu Frequency Variations in Cesium Atomic Frequency Standards," 21st Annual Precise Time and Time Interval Applications and Planning Meeting (30 November 1989), Redondo Beach, Calif.

[5] W. A. Johnson, S. K. Karuza, and F. J. Voit, "Long-Term Microwave Power Drift of a Cesium Frequency Standard and Its Effect on Output Frequency," 22nd Annual Precise Time and Time Interval Applications and Planning Meeting (5 December 1990), Vienna, Va.

Figure Captions

Figure 1. Block diagram of the C-field measurement system for a Cs frequency standard

Figure 2. Circuit diagram of the single-mixer frequency-measurement system used to determine the fractional frequency changes at different C-field settings

Figure 3. The measured square root of the Allan variance of the Cs frequency standard used in our measurements and the noise floor of the frequency measurement system

Figure 4. Block diagram of the Cs beam tube microwave-cavity measurement system

Figure 5. Plot of the microwave cavity return loss of the measured Cs frequency standard tuned to a Cs resonance frequency (f_0)

Figure 6. Plot of the Cs beam current vs. input microwave power of the Cs frequency standard tuned to f_0

Figure 7. Average of final data regarding the difference of the average frequencies as a function of C-field for a microwave power change of + 1 and + 3 dB above the optimum power level P_o

Figure 8. Plot of the difference of the average frequencies as a function of the C-field of the Cs frequency standard for two microwave cavity tuned frequencies (f_0 and ± 5 MHz) for a microwave power change of + 3 dB above the optimum power level P_o

SESSION 2A

RESONATORS I

Chairman : H. DUCHAUSSOY

Session 2A Number 6

A NEW APPROACH TO MODELING QUARTZ RESONATOR AND OSCILLATOR ACCELERATION SENSITIVITY

J. A. Kosinski, A. Ballato, and T. J. Lukaszek, US ARMY LABCOM
Electronics Technology & Devices Laboratory, Ft. Monmouth, N.J.

S. Mallikarjun, Department of Chemistry and Physics
Monmouth College, West Long Branch, N.J.

Abstract

A new approach to modeling the acceleration sensitivity of quartz crystal resonators and oscillators has been developed in the hope of providing a clearer understanding of the acceleration induced frequency perturbation. The current approach to modeling dwells primarily on the perturbation in either eigenfrequency (in most theoretical treatments) or resonance frequency (in most experimental approaches), as stated in the familiar vector description

$$f(\mathbf{a}) = f(\mathbf{a}=0)*(1+\Gamma\cdot\mathbf{a})$$

with the net frequency perturbation being the scalar product of the acceleration field and the device acceleration sensitivity.

Results presented by Valdois, et al. (1974), and Ballato, et al. (1990), have made clear the inadequacy of such treatments in describing the acceleration effect under certain operating conditions and in applications such as voltage controlled oscillators. This has led to the development of a modeling scheme wherein the effects of acceleration on the quartz resonator and oscillator may be examined separately. That is to say, the observed acceleration induced frequency perturbation in a particular oscillator configuration is well represented by first properly modeling the acceleration effects on the resonator, and then examining the interaction of the perturbed resonator with the oscillator loop.

Validation of the modeling procedure has been completed. Simulations of the effects of acceleration on typical resonators will be compared to the measured effects obtained in a unique measuring system. Simulations of the effects of acceleration on a typical VCO will also be compared to the effects on the VCO measured under various operating conditions.

Introduction

The vector description of the acceleration sensitivity of quartz resonators and oscillators has been widely used during the past decade. The familiar vector equation

$$f(\mathbf{a}) = f(\mathbf{a}=0)*(1+\Gamma\cdot\mathbf{a}) \qquad (1)$$

has been used interchangeably to describe the changes in eigenfrequency of the quartz resonator and operating frequency of the crystal oscillator. That these two quantities are in fact different was first suggested by the data of Valdois, et al. [1], and recently shown conclusively by the measurements of voltage controlled oscillators (VCO's) by Ballato, et al. [2]. These measurements have continued, with primary emphasis on measuring independently the resonator and oscillator effects, and developing a model to link the two.

Two separate measurement systems have been employed in this work. Figure 1 shows a block diagram of the shake table system used for VCO tests. The capabilities of this system have been described elsewhere [3]. Figure 2 shows a block diagram of the centrifugal accelerator system used for resonator tests. This system can produce acceleration levels up to 600g. The linkage of the network analyzer with the centrifuge allows for improved data acquisition and analysis as compared to previous implementations [4]. Although improvements in the commutator and interconnections have been made, commutator related systematic errors present a challenge in measuring the acceleration effects. The accurate determination of the acceleration effects along any resonator axis requires two tests taking advantage of the change in sign associated with the acceleration induced

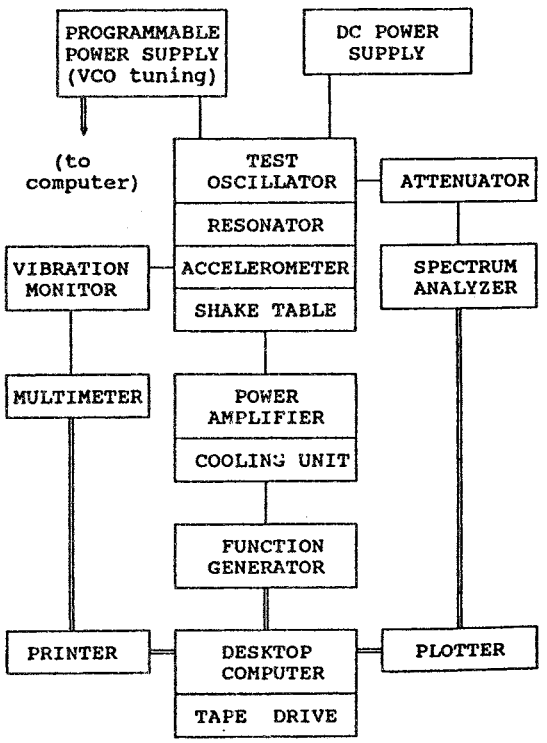


Figure 1. Shake table test system.

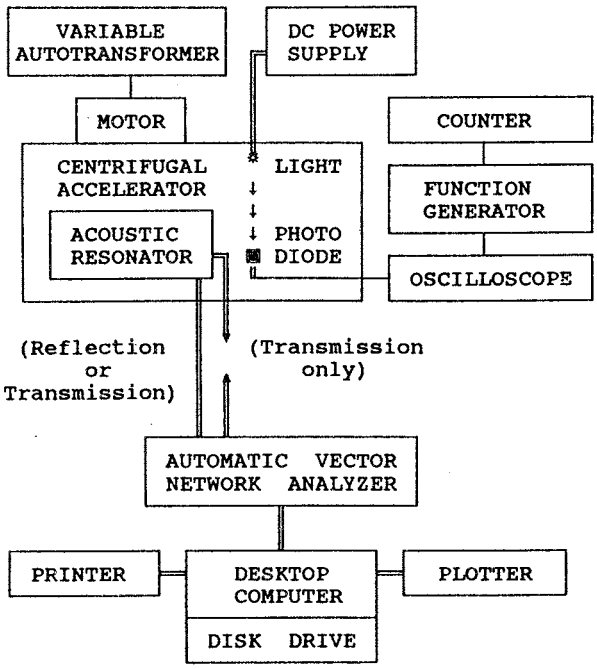


Figure 2. Centrifugal test system.

perturbation when the angle between the applied acceleration and the resonator sensitivity vector is changed by π radians.

Modeling Resonator Effects

The crystal resonator is typically employed as a feedback element in an oscillator loop wherein the steep phase-slope of the resonator transimpedance near the resonance frequency is used to stabilize the oscillator operation. It is important to note that (1) the resonator is almost always used "off resonance" in conjunction with a series load impedance, and (2) the quantity which is actually of importance is not the perturbation in resonance frequency but rather the perturbation in resonator transimpedance at the actual operating point. These considerations argue in favor of an equivalent circuit approach to modeling resonator effects.

The Butterworth-Van Dyke single-mode lumped element approximate equivalent electrical circuit is well known. The equivalent circuit parameters may be determined from the crystal material properties using well known relations obtained using the linear piezoelectric theory for the case of the infinite flat plate:

$$C_o = \frac{\epsilon A}{2 h} \tag{2}$$

$$R_1 = \frac{n^2 \pi^2 h}{4 A e^2} \eta_s = \frac{n^2 \pi^2 h \bar{c} r_1}{4 A e^2} \tag{3}$$

$$L_1 = \frac{h^3 \rho}{A e^2} \tag{4}$$

$$C_1 = \frac{4 A e^2}{n^2 \pi^2 h \bar{c}} \tag{5}$$

The point of note is that R_1 and C_1 depend on $\bar{c}$, while C_o and L_1 do not. Since the acceleration induced effects are considered to arise out of purely elastic nonlinearities, one is led to consider the

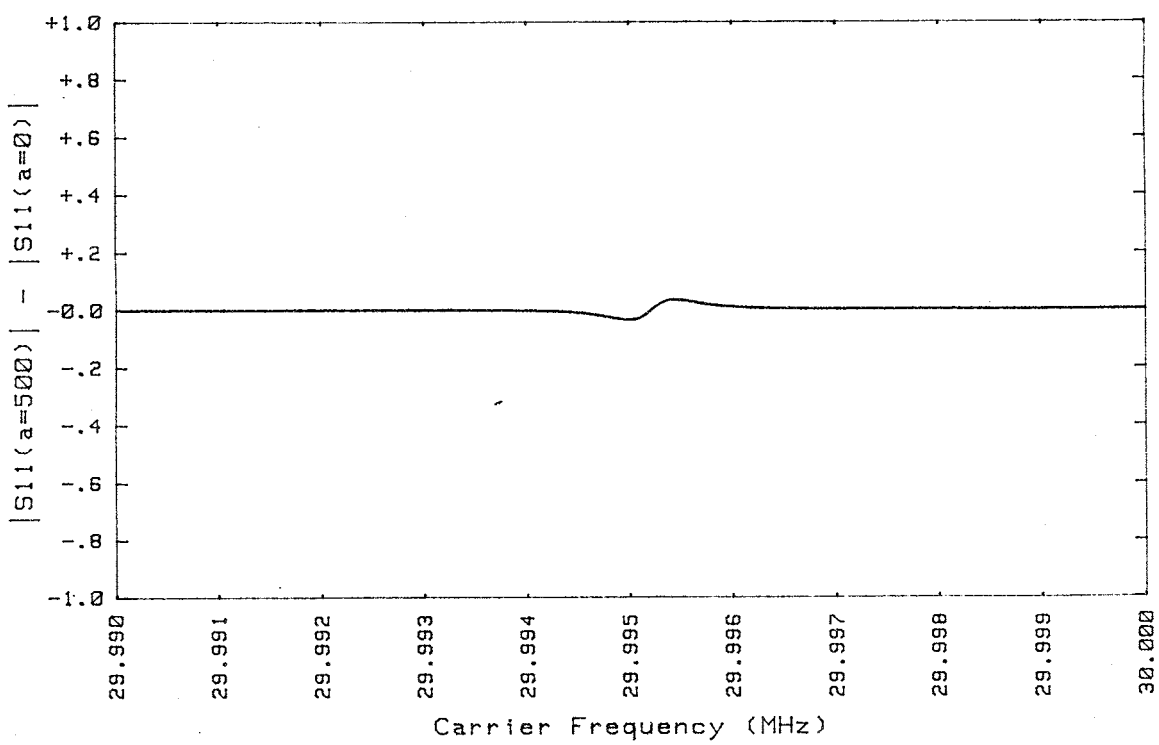


Figure 3. Perturbation in the magnitude of S_{11} for a 30MHz resonator with $|\Gamma|=6\times10^{-9}/g$ as $|\mathbf{a}|$ varies between 0 and 500g.

effect on these relations of introducing an acceleration sensitive elastic stiffness. Recognizing that the piezoelectric contribution to the stiffened elastic constant is small in the case of quartz resonators, we model the acceleration sensitive elastic stiffness as

$$\bar{c}(\mathbf{a}) = \bar{c}(\mathbf{a}=0)*(1+\boldsymbol{\gamma}\cdot\mathbf{a}) \qquad (6)$$

wherein we use the lower case Greek gamma to denote the acceleration sensitivity of the elastic stiffness. The relationship between the acceleration sensitivity $\boldsymbol{\gamma}$ of the elastic stiffness and the acceleration sensitivity Γ of the resonance frequency is simply $\boldsymbol{\gamma} \approx 2\,\Gamma$.

Examination of the acceleration induced transimpedance perturbation follows directly from the substitution of equation (6) into equations (3) and (5). The validity of this modeling approach was tested through a series of s-parameter measurements made with the centrifugal accelerator system in both reflection and transmission modes. The form of the perturbation in S_{11} predicted for a one-port reflection measurement is shown in Figure 3.

Modeling Oscillator Effects

Once the perturbation in resonator transimpedance is known, it is a simple task to obtain the corresponding frequency shift and acceleration sensitivity of an ideal oscillator employing such a resonator as

$$\Gamma = -\frac{[\phi(\mathbf{a}) - \phi(\mathbf{a}=0)]}{2\,\mathbf{a}\,Q_L} \qquad (7)$$

where $\phi(\mathbf{a})$ and $\phi(\mathbf{a}=0)$ are the perturbed and unperturbed phase of the resonator transimpedance, and Q_L is the loaded or phase slope Q of the resonator at the operating point. Implicit in equation (7) is that no other components in the oscillator loop are perturbed by the acceleration. For actual hardware implementations, this is not necessarily true. The acceleration sensitivity for the ideal case with no loop component contributions of an ideal oscillator based on this modeling technique is shown in Figure 4, with the corresponding resonator S_{21} response shown in Figure 5.

In the case where other loop components are also sensitive to the external acceleration, the effects on the oscillator loop are well represented by an ideal amplifier in series with an acceleration sensitive load reactance. Based on experimental observations, we treat the load reactance as an ideal component with an acceleration sensitive lead inductance. The effect on the lead inductance may be modeled as a linear perturbation in similar fashion to equation (6), and the perturbation in transimpedance phase of the resonator/series load reactance combination may be employed in equation (7) to obtain the oscillator acceleration sensitivity under various operating conditions. Figure 6 compares the application of this modeling technique to a crystal VCO with the measured data obtained for the unit.

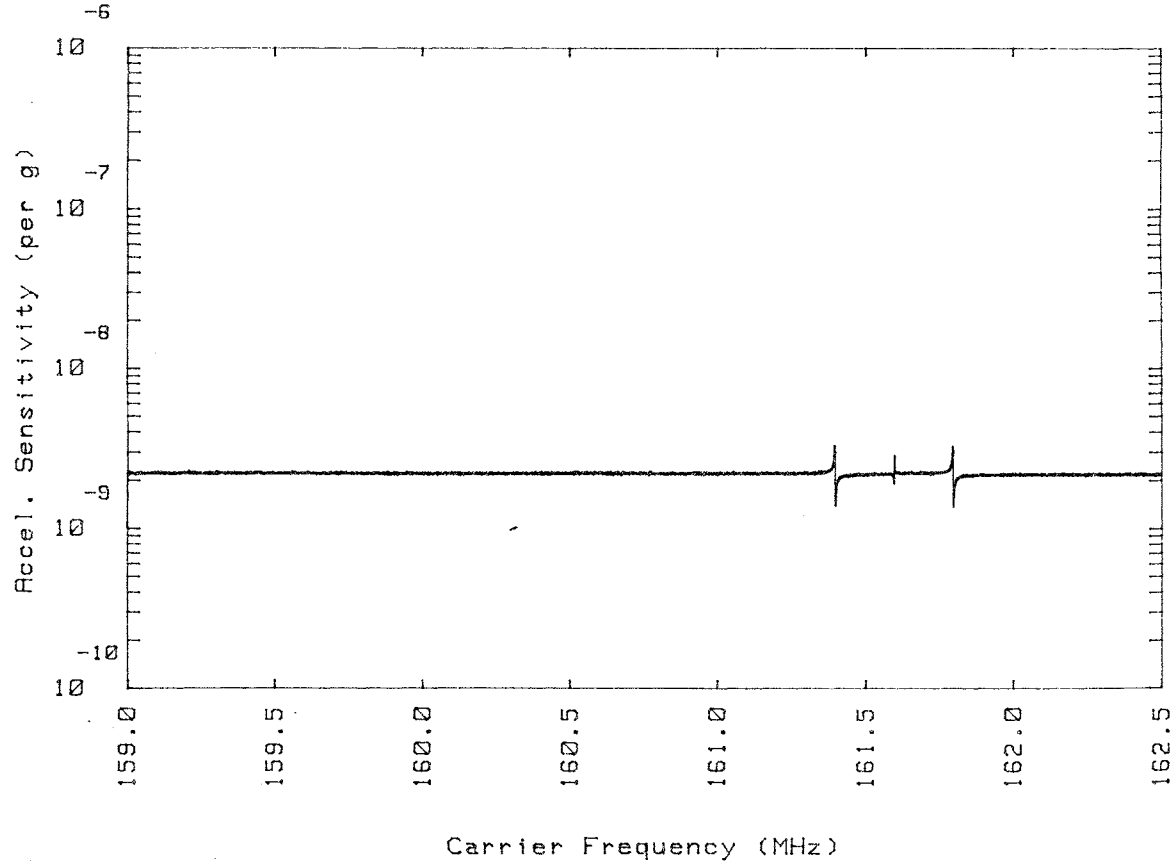


Figure 4. Acceleration sensitivity of an ideal oscillator in which only the resonator transimpedance is acceleration sensitive. For this resonator, $|\Gamma|$ is nominally 2.2×10^{-9}/g.

Discussion

The results obtained using the equivalent circuit approach are in good agreement with the measured data. As a practical consideration, the importance of maximizing the resonator loaded Q in order to minimize the effects arising from other loop components is well seen from Figure 6. The importance of applying good construction techniques is also clearly illustrated. The equivalent circuit modeling approach should prove particularly helpful in analyzing dual-resonator and other acceleration sensitivity compensation schemes.

References

[1] M. Valdois, J. Besson, and J. Gagnepain, "Influence of Environment Conditions on a Quartz Resonator," Proc. 28th Annual Symposium on Frequency Control (ASFC), May 1974, pp. 19-32.

[2] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, and R. McGowan,"Electronic Desensitization of Resonators to Accelerations," Proc. 44th ASFC, May 1990, pp. 444-451.

[3] J. Kosinski, A. Ballato, and T. Lukaszek, "Measurements of Acceleration-Induced Phase Noise in Surface Acoustic Wave Devices," Proc. 44th ASFC, May 1990, pp. 488-492.

[4] M. Nakazawa, T. Lukaszek, and A. Ballato, "Force- and Acceleration- Frequency Effects in Grooved and Ring- Supported Resonators," Proc. 35th ASFC, May 1981, pp. 71-91.

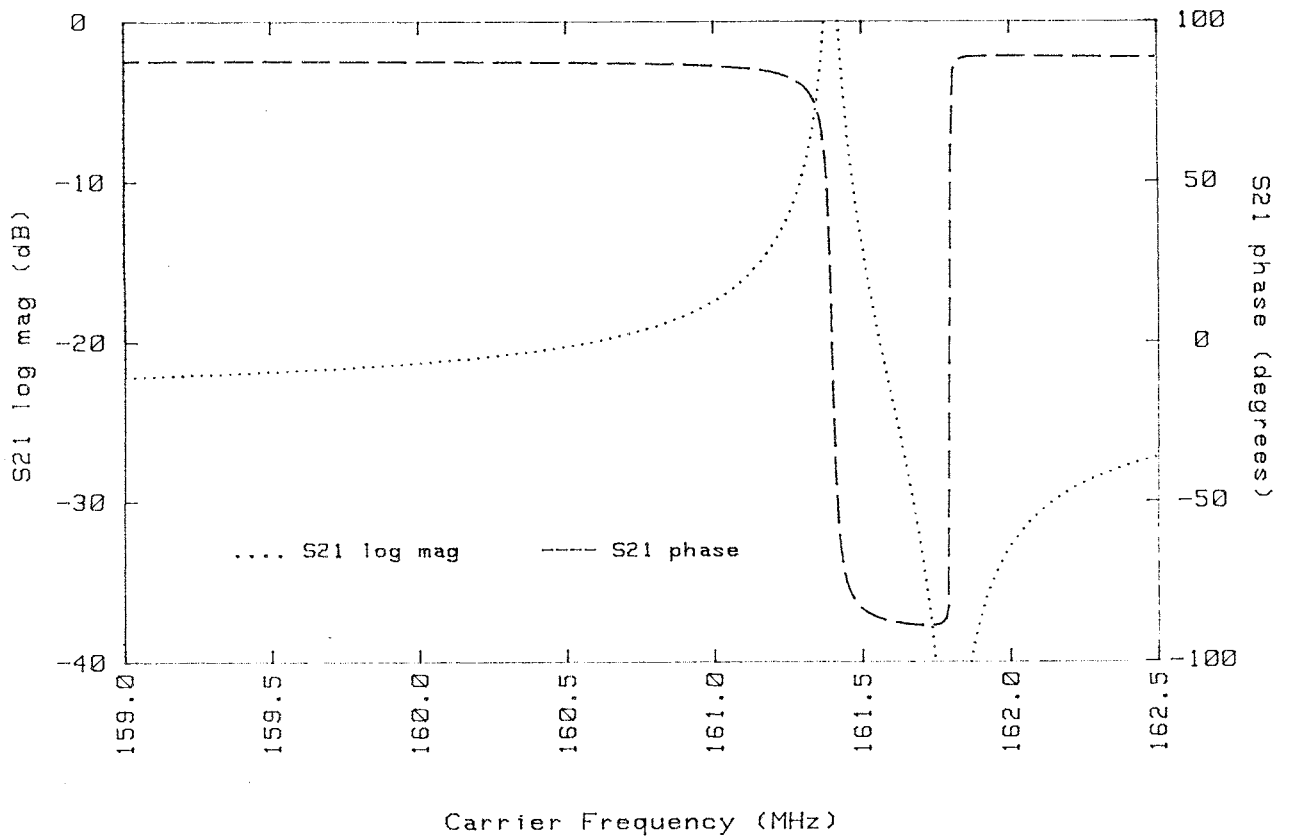


Figure 5. S_{21} response for the resonator of Figure 4.

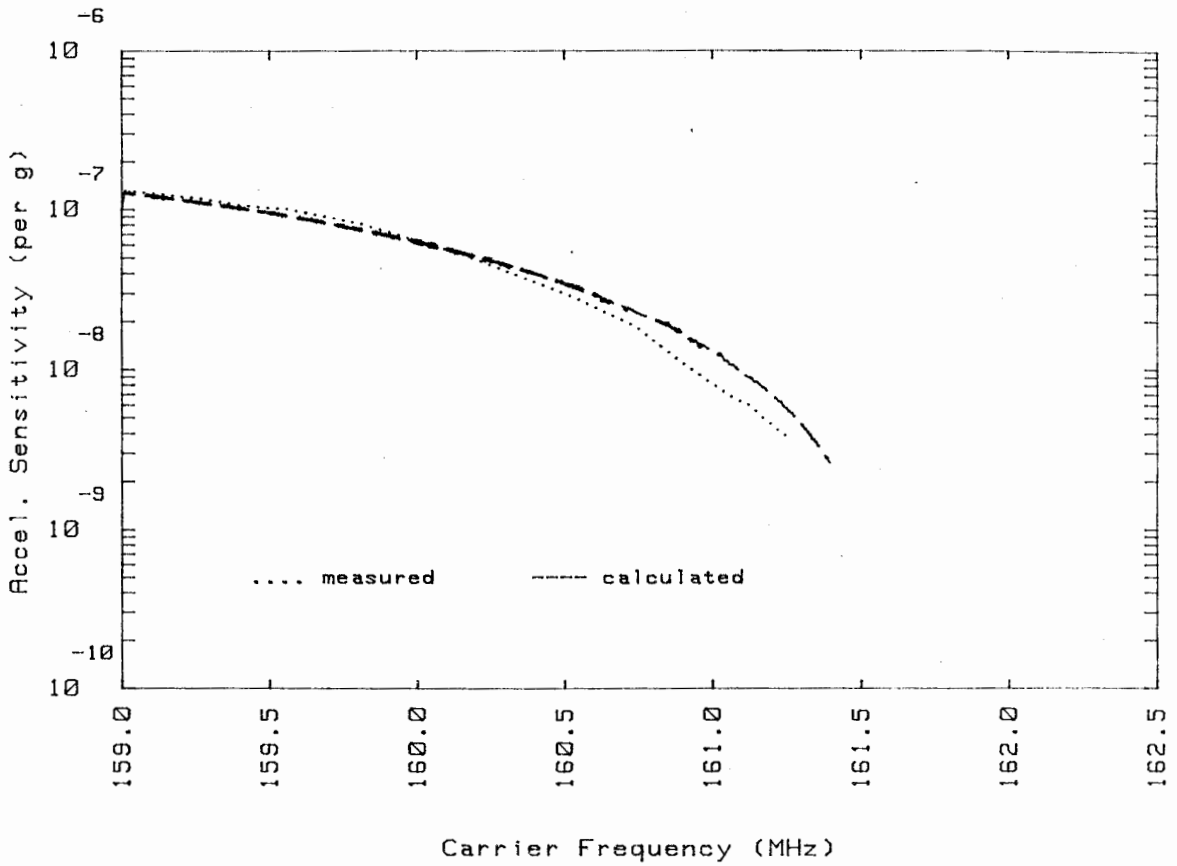


Figure 6. Measured and calculated acceleration sensitivity of an actual VCO employing the resonator of Figure 4.

Session 2A Number 7

Optimization of the Design of Thickness Shear Devices using the AT Cut of Berlinite

J.Détaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot*

Centre National d'Etudes des Télécommunications dept PAB/BAG/MCT/CMM
196, Av. H.Ravéra 92220 Bagneux France
*Université de Montpellier, Laboratoire de Physico-Chimie des Solides
Place E.Bataillon 34000 Montpellier France.

Abstract.

Systematic investigations concerning the design of berlinite resonators were made in order to find the best design for the elements of filters for the intermediate frequency of the future pan-european numerical radiotelephone system. The nature and the severity of the requirements for this filter prescribe to build it with elements having a very reduced sensitivity to the dispersions induced by the elaboration process.

Using the crystals with a low impurity concentration now available, the angular position of the AT cut for the 3rd overtone thickness shear mode was determined for the plane resonators. This was done using resonators cut nearly every 10' in the vicinity of the approximate position of this cut that was estimated from the measurements previously made on third overtone resonators [1]. The laws of variation of the temperature coefficients of the resonance frequency (up to the third order) as a function of the cut angle were extracted by least square analysis. Similar investigations were made for the fundamental mode.

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 on the elements of the equivalent scheme and on the anharmonic spectrum was made . This was done using numerical models based upon the theory of the essentially thickness modes [2]. Several family of designs were found to be quite satisfactory , but when a constraint was put on the impedance level (inductance) and on the sensitivity to the variations of several parameters that are difficult or costly to totally control in an industrial process, one family of design 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 to 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 90MHz. It is generally estimated that more than 10 millions of radiotelephone sets will be sold in the next ten years in Europe.

It was previously shown that the specifications for this application can be perfectly fulfilled by a bulk wave filter with a linearized phase response using four discrete berlinite resonators eventually integrated on the same plate [3] [4]. In order to optimize the characteristics of the filters and also to maximize the yield , we have research an optimization of the design of the piezoelectric components of the filters relatively to several criteria. The most important one being to minimize the sensitivity of the elements of the equivalent scheme towards the dispersions of the physical parameters that result of the technological process. Another important criteria was to obtain resonators with a sufficiently low value of the inductance, since it is an important advantage to keep the iterative impedance of the filter as low as possible in order to reduce the influence of parasitic capacitances and also to facilitate the botheration of the transformers.

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 for the fundamental mode of plane berlinite resonators was determined using the crystals with a low impurity concentration now available. In order to analyse the possibilities offered by berlinite, and to obtain for the piezelectric components of the filters the most interesting designs, a systematic investigation of the influence of the parameters (electrode shape and surface, mass loading, exact orientation etc....) of the plane fundamental resonators using the AT cut of berlinite on the elements of the equivalent scheme and on the anharmonic spectrum was made. This was done using numerical models based upon the theory of the essentially thickness modes [5]. Several family of designs were found to be quite satisfactory , but when a constraint was put on the impedance level (inductance) and /or on the sensitivity to the apparition of anharmonic modes, one family of designs was found to be much better. This family of designs is based upon resonators that take into account the lateral anisotropy of AT berlinite plates [6] (the "in plane" anisotropy of berlinite is higher than that of quartz so that more gain is obtained with this material when it is taken into account). The results of these investigations are given and discussed in this communication.

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 a. The intermediate frequency (I.F.) is not defined by the GSM specifications which only specify the global channel selectivity (Figure 1b).

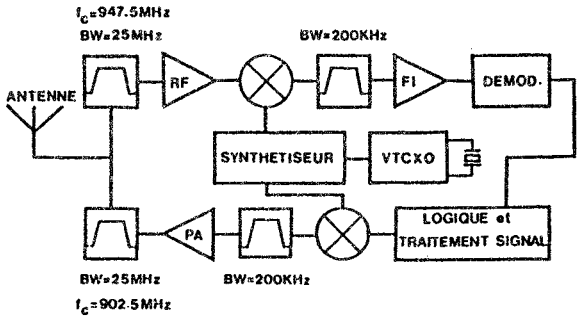


Figure 1a: Typical scheme of the radio sub-system of a GSM radiotelephone.

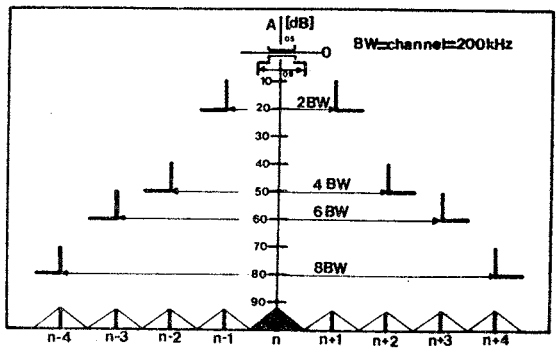
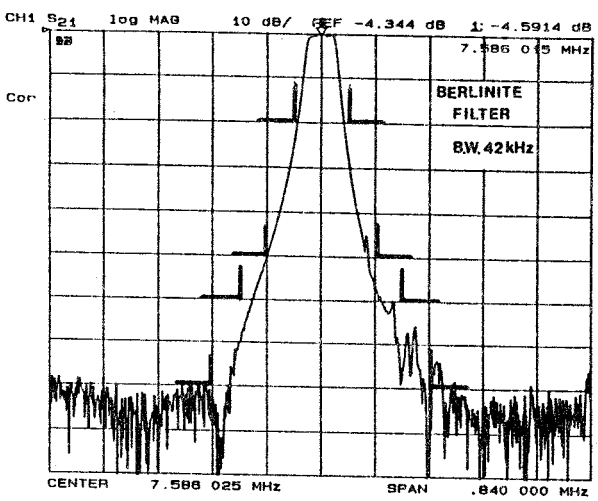
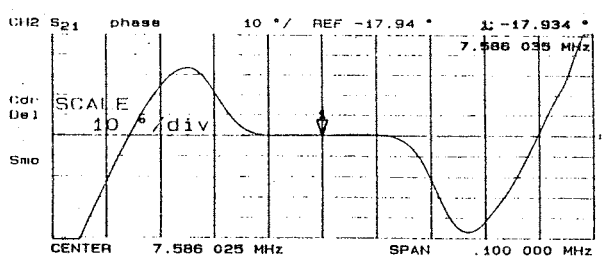


Figure1b: Global specifications of the channel selectivity for the G.S.M. radiotelephone.

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 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 take into account the existing solutions for the I.F. filter. 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 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 2 are those of such a filter that was obtained in 1987 [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].





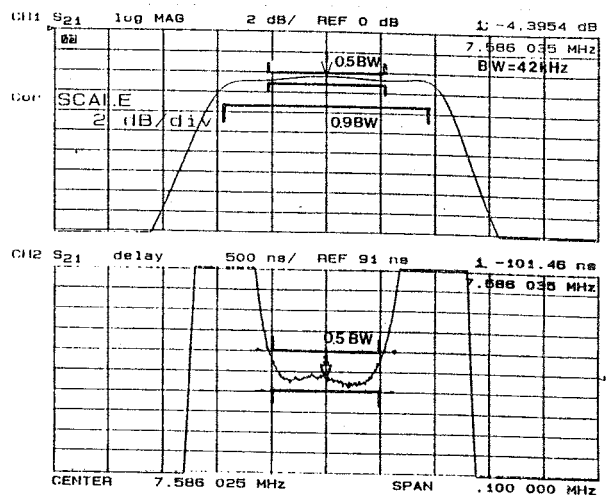


Figure 2: Response of a 4 poles Jaumann berlinite filter with a linearized phase response. [2a global amplitude response, 2b amplitude , phase and delay in the passband.]

Other choices of the distribution of selectivity in the G.S.M. set appears to be made, and it seems, in the present state of receiver circuits, that the minimum selectivity that can be placed in the IF filter is some thing like what is represented in figure 3. Such specifications can be fulfilled either by a low loss surface wave filter or a two poles berlinite bulk wave filter.

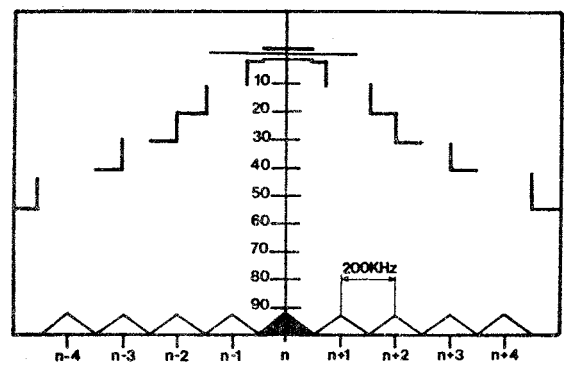


Figure3: An example of specifications for the I.F. filter for a set win which a very large amount of the selectivity is obtained by additional filtering in the base band.

Among the other constraints put on the components for this system are an important degree of miniaturization and a very reduced cost. In order to find filter that have the best characteristics at the lowest cost, we have undertaken an investigation of the designs of the piezoelectric components of the filter to reduce the sensibility of the equivalent scheme to technological dispersions that are costly and difficult to avoid at the very high frequencies required. Three important features , particularly investigated, were to avoid the existence of anharmonic modes in the resonators, to obtain a low value of the equivalent inductance and to keep the temperature shift of the response to a minimum. The first point is important to fulfil the specifications of out of band rejection , it is a known difficulty for V.H.F. filters and several solutions were already proposed [8] [9]. The second point is important to reduce the cost of transformers and to minimize the sensitivity to parasitic capacitances. The third point is of evident interest since the mobile (car) and portable equipments have to operate in a large temperature range.

II Optimization of the crystalline orientation.

Using the results that we have previously published [1] and also new results of thermal behaviour of berlinite 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 obtained by least square analysis, the variations of the 1st, the 2nd and the 3rd order temperature coefficients of the resonance frequency (third overtone) as a function of the cut angle. The obtained 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 inflexions 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 the values of the 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 cut situated between Y-32°47.5' and Y-33°10'.

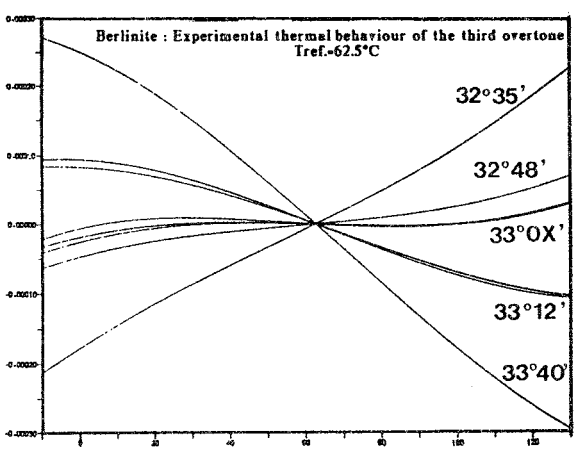


Figure 4: Experimental results of the thermal behaviour of Y rotated berlinite resonators near the AT cut [Third overtone].

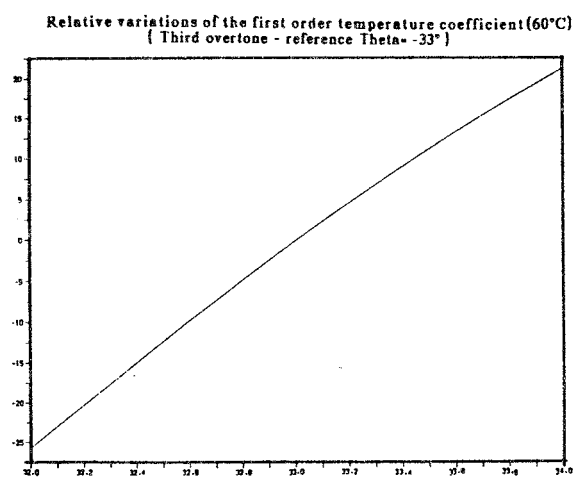


Figure 5: Variation of the first order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone].

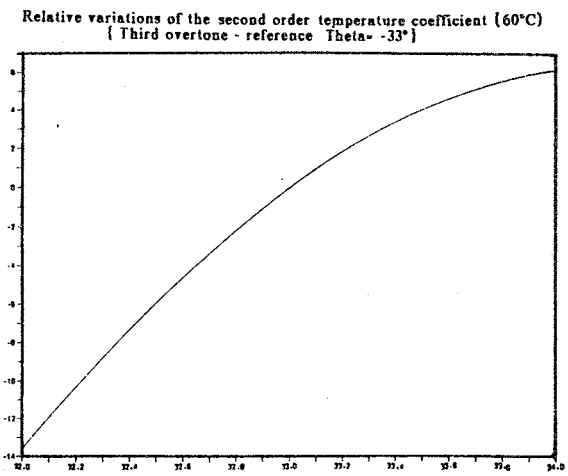


Figure 6: Variation of the second order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone] .

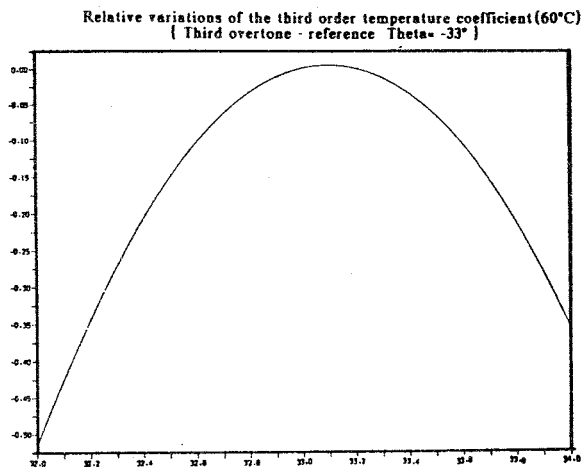


Figure 7: Variation of the third order temperature coefficient of the resonance frequency in the range[Y-32°, Y-34°] [Third overtone].

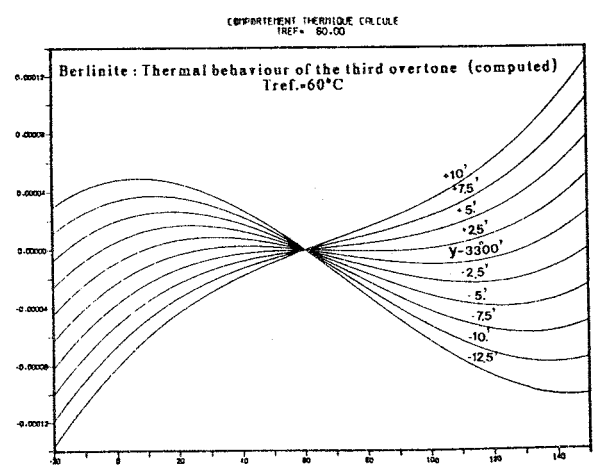


Figure 8: Computed thermal behaviour for the third overtone near AT.

The same measurements were also made for the fundamental mode using resonators made with the high Q factor crystals now available and having nominal cuts spaced nearly every 10 minutes in an interval situated around the cut presenting the horizontal inflexion tangent (AT+0°] for the third overtone mode . On figure 9 is represented the frequency response of one of the resonator used for these measurements. An example of the obtained results is displayed on figure 10 ; it represents the thermal behaviour of resonators with orientation near Y -32°48' presenting the angular dispersion resulting of the all the technological operations needed to obtain a VHF plate (made here starting with thick plates) 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 and present an angular dispersion that was voluntary slightly increased to obtain more informations around this angle.

III Optimization of the design of berlinite resonators for filter applications III.1 Physical and electrical parameters of a resonator, sensitivity.

The physical parameters of a piezoelectric resonator are listed in table I together with the origin of their dispersions. The case of the plate orientation and of the corresponding dispersions is particular since it is practically uncoupled of the other since small angular errors only affect the thermal behaviour and not , in a significant direct manner, the values of the elements of the equivalent scheme at the temperature at which the frequency adjustment is made.. For this parameter 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 appears. 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 action 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 in 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 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 VHF frequency 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.

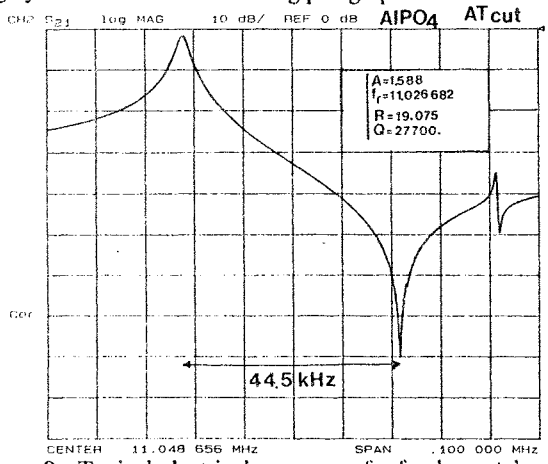


Figure 9: Typical electrical response of a fundamental mode resonator near AT.

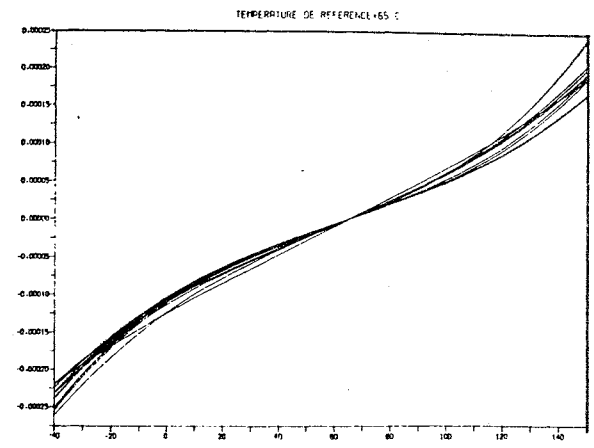


Figure 10: Thermal behaviour of fundamental mode resonators [Y-32°48' +- 2'].

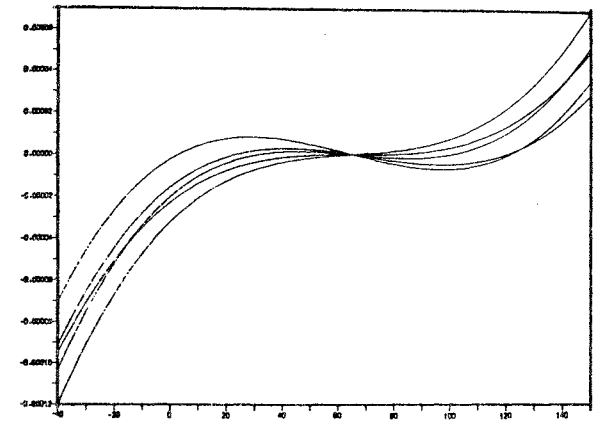


Figure 11: Thermal behaviour near the AT cut for the fundamental mode.

PHYSICAL PARAMETERS OF RESONATORS	
-Plate orientation	-(angular dispersion) -(effect on T.C.)
-Thickness of the plate	-(thickness distribution) -(parallelism)
-Electrode geometry	-(dispersion of dimensions) -(misalinement of electrodes)
-Mass loading R	-(accumulation of all the the preceding errors)
ELECTRICAL PARAMETERS	
-Resonance frequency	(accuracy of D.t.F. plating)
-Inductance	Include all dispersions
-Static Capacitance	(errors on geometry)
-Parasic responses	highly sensitive on most of the physical parameters.

For V.H.F. filters, a low sensitivity of the inductance towards the physical parameters is mandatory.

A very reduced sensitivity towards the apparition of anharmonics is of large importance.

A reasonably low value of L is also required.

Table 1: Parameters of a thickness shear resonator.

The electrical response of a crystal filter is sensitive (more or less depending on their designs) to the dispersion of their constitutive elements and in particular to the dispersion on the electrical parameters of the crystals which are elements having reactances with fast frequency variations. It it thus important to find designs for these resonators that minimise the influence of the unavoidable technological dispersions of their physical parameters on the values of the equivalent scheme. This can be done using a model relating their physical parameters to their electrical properties.

III.2 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 [10] then for plates of an arbitrary orientation [2]. In this case the equation has, in 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, $\bar{u}_2(x_1,x_3,t)$ 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 dispersion relation. 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 the resonances frequencies [11][12].

Eigen mode analysis : The eigen modes of the resonator can be obtained following the simultaneous solution of the equations for the electroded and unelectroded region 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 :

$$f^* = f_{ce1} = \frac{n}{4h}\sqrt{\frac{\hat{c}^{(1)}}{\rho}}$$

for the electroded region.

or:

$$f^* = f_{el} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$$

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 :

$$r^2R'' + rR' + R(r^2A - \upsilon^2) = 0$$

$$T'' + \upsilon^2 T = 0$$

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 = ν 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 = ν 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 [13][14][6], this can be approximated by the expression of these conditions at only a discrete number q of points (figure 4). It the solutions for the two regions are :

$$u_1^I = \left(\sum_{m=0} A_m J_n(r\sqrt{A})\right)\cos mt \sin(n\pi x_2/2h)$$

$$u_1^{II} = \left(\sum_{m=0} B_m K_m(r\sqrt{A'})\right)\cos mt \sin(n\pi x_2/2h) \qquad \{8\}$$

at the point $P(x_1^P x_2^P x_3^P)$ we have :

$$u_1^I(x_1^P x_2^P x_3^P) = u_1^{II}(x_1^P x_2^P x_3^P)$$

$$n_1\frac{\partial u_1^I(P)}{\partial x_1} + n_3\frac{\partial u_1^I(P)}{\partial x_3} = n_1\frac{\partial u^{II}{}_1(P)}{\partial x_1} + n_3\frac{\partial u_1^{II}(P)}{\partial x_3} \qquad \{9\}$$

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 :

$$\hat{u}_{1e}^n = sin(n\pi x_2/2h)\bar{u}_{1e}^n$$

$$\bar{u}_{1e}^{n\mu} = e^{j\omega t}\sum_{i=0}^{i=q-1} A_{2i}^{n\mu} J_{2i}\left(r.\sqrt{A_e^{n\mu}}\right)cos(2it) \qquad \{10\}$$

and for the unelectroded region :

$$\hat{u}_{1l}^{n\mu} = sin(n\pi x_2/2h)\bar{u}_{1l}^{n\mu}$$

$$\bar{u}_{1l}^{n\mu} = e^{j\omega t}\sum_{i=0}^{i=q-1} B_{2i}^{n\mu} K_{2i}(r.\sqrt{-A'^{n\mu}_l})cos(2it) \qquad \{11\}$$

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 :

$$\iiint_{V_t} \hat{u}_1^{n\mu}\hat{u}_1^{k\lambda}dV = \iiint_{Ve} \hat{u}_{1e}^{n\mu}\hat{u}_{1e}^{k\lambda}dV + \iiint_{Vl} \hat{u}_{1l}^{n\mu}\hat{u}_{1l}^{k\lambda}dV$$

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 suplementary term must be added [15]

$$\hat{u}_1 = \sum_n\sum_\mu H^{n\mu}\hat{u}_1^{n\mu}\sin(n\pi x_2/2h) \qquad \{14\}$$

$$\hat{\phi} = \sum_n\sum_\mu H^{n\mu}\hat{\phi}^{n\mu} + \frac{x_2 V_0 e^{j\omega t}}{2h} \qquad \{15\}$$

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 dependance 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 dependance 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:

$$H^{n\mu} = (-1)^{(n-1)/2}\frac{f^2 h e_{26}}{(f^2 - f_{n\mu}^2)n^2\pi^2 C_{66}}\frac{4V_0 e^{j\omega t}\iint_{Se}\bar{u}_1^{n\mu}dS}{2N^{(n)(\mu)}} \qquad \{16\}$$

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 $Y = \frac{-1.}{V_0 e^{j\omega t}}\iint_{Se} D_2 dS$

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 :

$$C_z = \frac{S_e \varepsilon_{22}^S}{2h}(k_{26}^2 + 1)$$

is the capacitance at f = 0 {17}

$$C_{n\mu} = \left[\frac{8k_{26}^2 \varepsilon_{22}}{n^2\pi^2 2h}\right] \frac{h\left(\int\int_{Se} \tilde{u}_1^{n\mu} dS\right)^2}{N^{(n)(\mu)}} \quad \{18\}$$

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 :

$$C_0 = C_Z - \sum_n \sum_\mu C_{n\mu} \quad \{19\}$$

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.

III.3 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 radii ranging from 0.15mm to 0.4mm and a mass loading ranging from 0.5% to 1.5%. The results relative to the resonance frequency 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 different shapes of electrodes.

Let us first recall that the relative sensitivity of a quantity Q (x,y,z,....) to the variations of x is defined by: $S_x(Q) = (x/Q).\partial Q/\partial x$, the absolute sensitivity being $\partial Q/\partial x$.

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 for a much lower value of the electrode radius with the largest value of the mass loading 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 to the resonance frequency of the partially electroded plate); we can observe that this frequency lowering is greatly dependant of the dimensions of the electrodes.

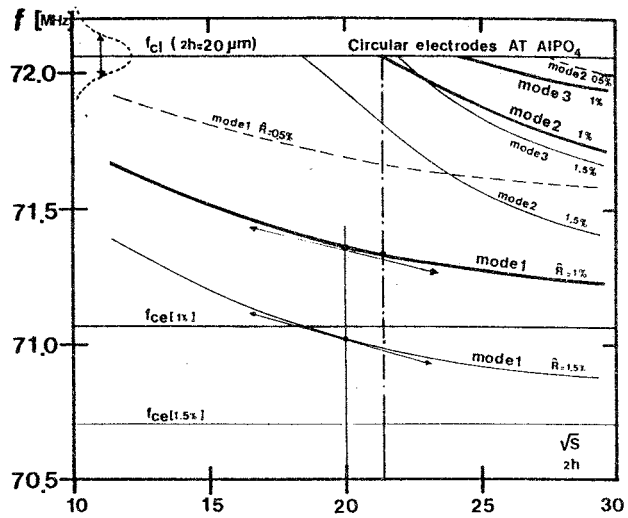


Figure 12: Thickness shear modes of AT berlinite resonator with circular electrodes.(Plate thickness = 20 microns , mass loading 0.5, 1., 1.5 %).

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}). This representation of Ω in function of the same normalized electrode dimension as for figure 12 is independant of the thickness of the plate.

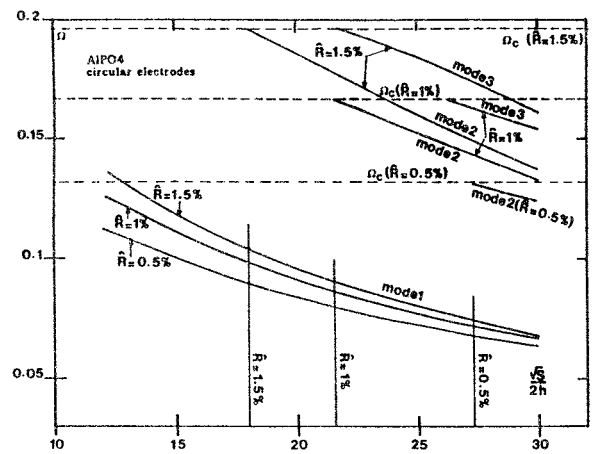


Figure 13: Thickness shear modes of AT berlinite with circular electrodes in reduced frequency coordinates .

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 independant of the mass loading, at least for 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.

The sensitivity of the inductance to the mass loading vanishes for a value of the normalized electrode dimension situated near $\sqrt{S}/2h=25$ depending on the value of the mass loading. This value near 25 can be used to obtain filter resonators free of spurious response if the mass loading is kept below 0.7 to 0.6%. Then a low value of the inductance is obtained.

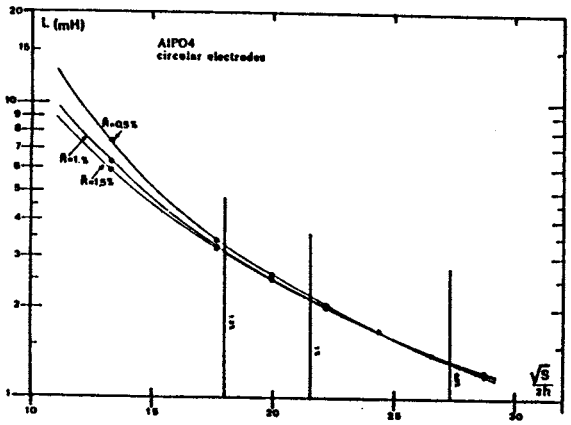


Figure 14: Inductance of 20 ,m AT resonators with circular electrodes.

III.4 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.

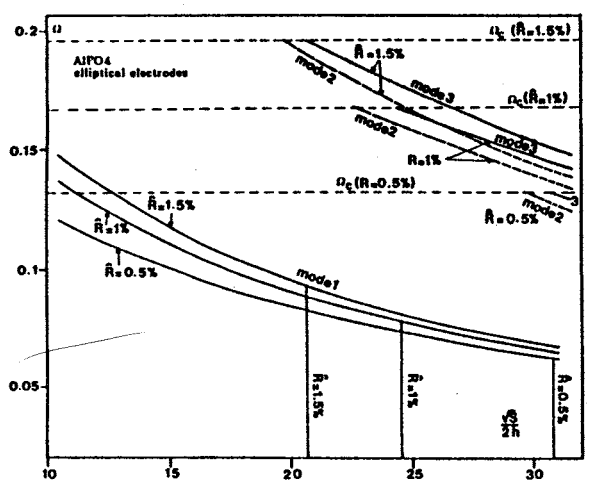
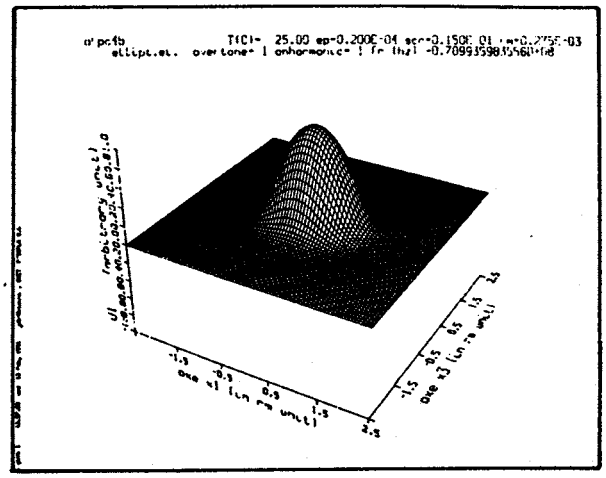
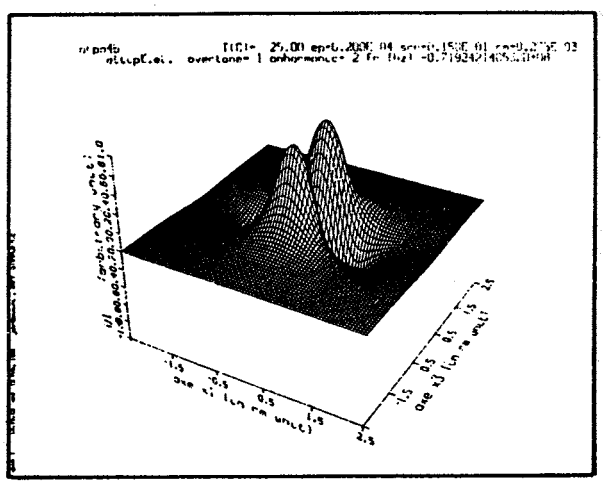


Figure 15: Thickness shear modes of resonators with elliptical electrodes respecting the in plane anisotropy of the AT cut.

On figure 17 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] which correspond to 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.





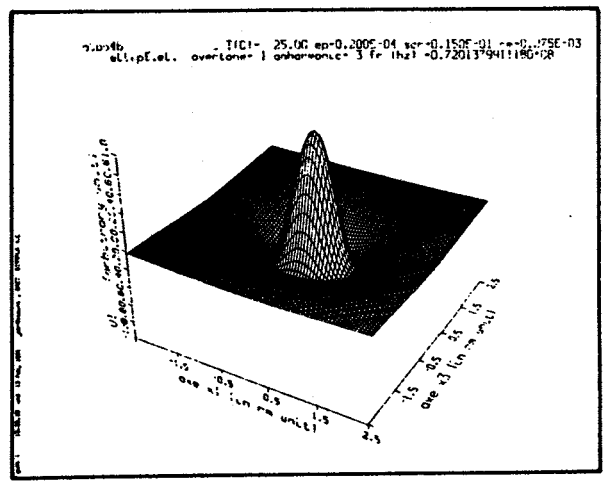


Figure 16: Example of computed modes for AT resonators with elliptical electrodes.

On figure 18 are represented the computed modes of a resonator having a large a/b (2.5) ratio but a smaller mass loading (1.0%), these mode are drawn at the same scale as in figure 16. In this case (figure 17), 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 19), we can observe that the in plane anisotropy of AT berlinite makes that the shape of the mode is different of the shape of the electrodes, this is particularly visible in the case of a circular electrode (a/b=1.).

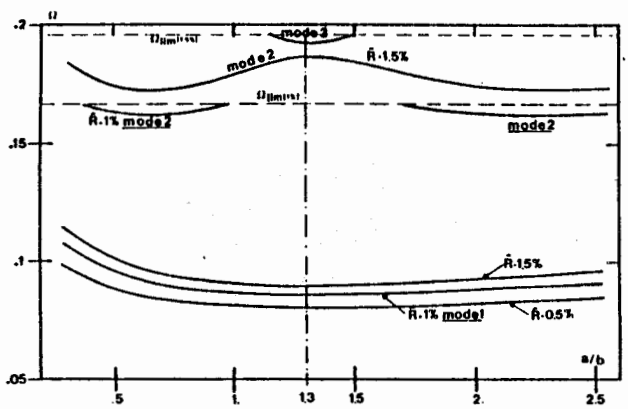


Figure 17: Thickness shear modes of elliptical electrodes with constant surface and varying axis ratio.

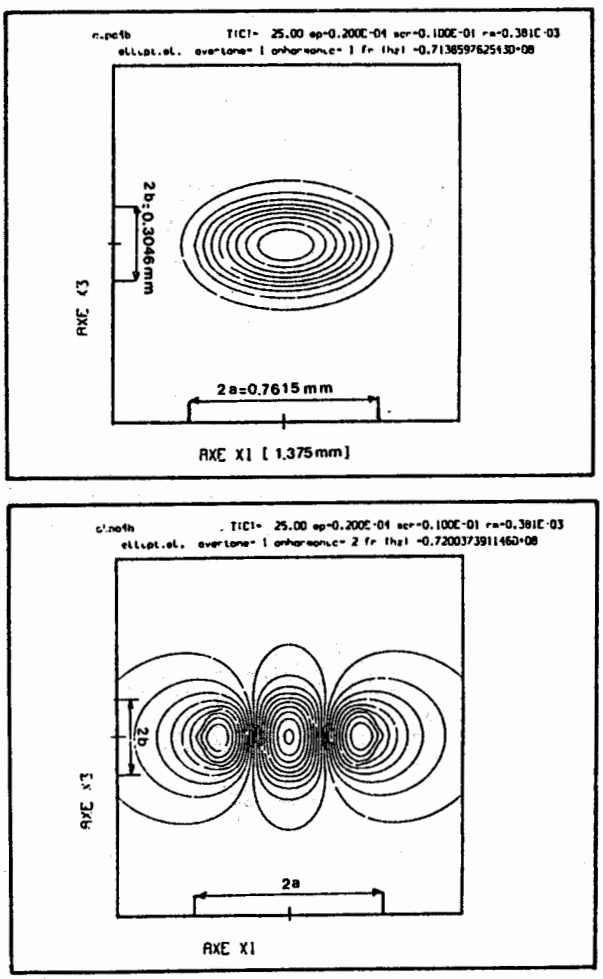


Figure 18: Computed modes of an AT berlinite resonator with elliptical electrodes having a large a/b ratio.

On figure 20 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 $u^2(dmax) < 2.10^{-4} * u^2max$, where u^2max is the maximun of u_1^2. On figure 19 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 optimal choice, for each mass loading , of the electrode dimension and of the distance between resonators. This permit a simultanneous optimization of the plate dimension and of the value of the inductance ofthe integrated resonators. Moreover, the minimum of dmax occurs for values of the electrode dimensions that give favourable sensibility properties.

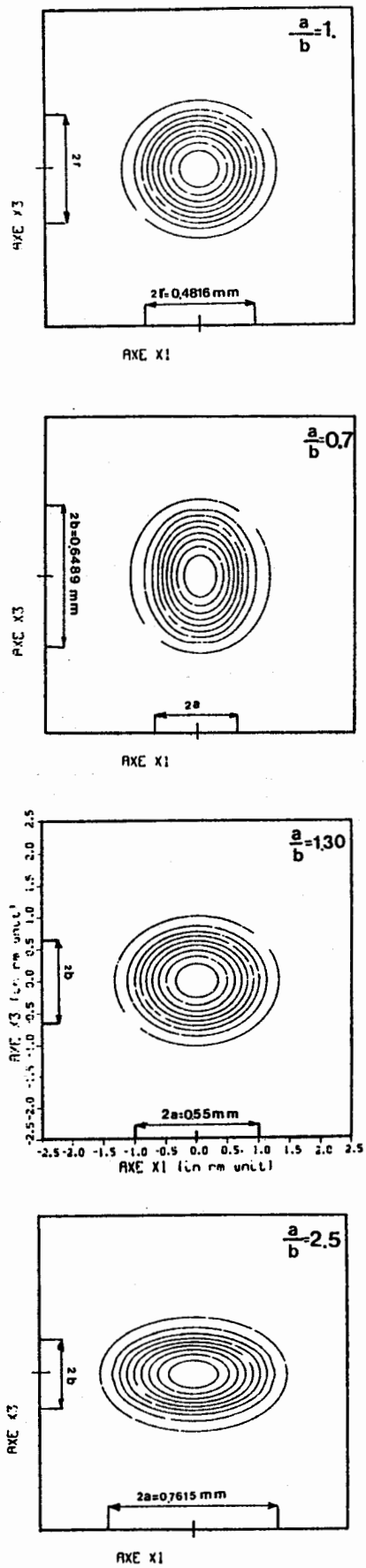


Figure 19: Comparison of the computed fundamental modes of berlinite resonators with elliptical electrodes having a constant surface and different axis ratios.

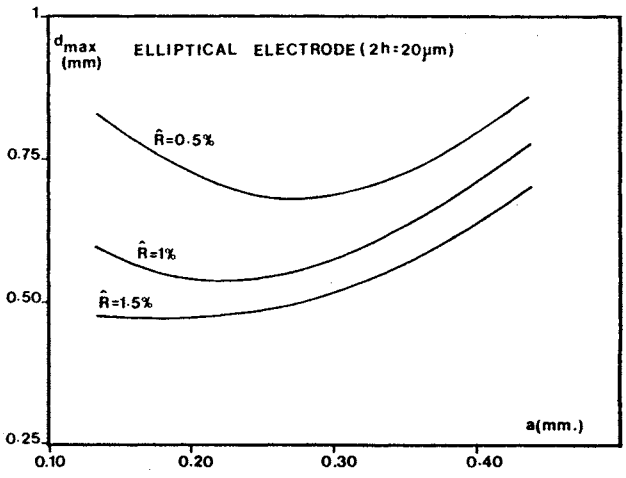


Figure 20: "Dimension" of the vibration mode in elliptical resonators, as a function of the electrode dimension (a) for three mass loading (0.5%, 1.0%, 1.5%) . dmax is here the radius at which the square of the displacement is 1./5.10**3 of the maximum value for the mode (in the center for the fundamental).

III.5 Comparison of resonators having electrodes of different geometry:

In table 2 we compare the computed inductances of resonators having the same thickness, the same area and the same mass loading but different electrode geometry.

In table 2 we can observe that the resonance frequencies of the fundamental mode and the corresponding 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 quantity are observed for the elliptical electrodes that respects the lateral anisotropy of AT berlinite. It can be also noticed that the rectangular and square electrodes give always a slightly higher values than the corresponding elliptical or round electrodes. This is due to the unefficiency 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 the conditions of apparition of the first symmetrical anharmonic mode, the electrode shapes having a ratio a/b corresponding to the in plane anisotropy of AT berlinite as it is described by equation {1} are much better than the others (see figure 19). Among all the electrode shapes having the proper a/b ratio , the elliptical ones are from far the best.

TABLE 2

S = mm²; $\hat{R}$ = 1 %; 2 h = 20 μm

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**

Table 2: Comparison of the inductance of resonators with different electrode geometry.

III.6 Comparison with experimental results

Computed results were compared to experimental ones for AT berlinite resonators made with resonance frequencies in the range of 10MHz and round electrodes, a raisonably good agreement was found for the inductance and the frequency. A comparison was made for VHF resonators with elliptical electrodes . These resonators were made using chemically milled very thin plates of quartz, commercially available. 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 21 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 are in the range 1.52-1.58 mH.

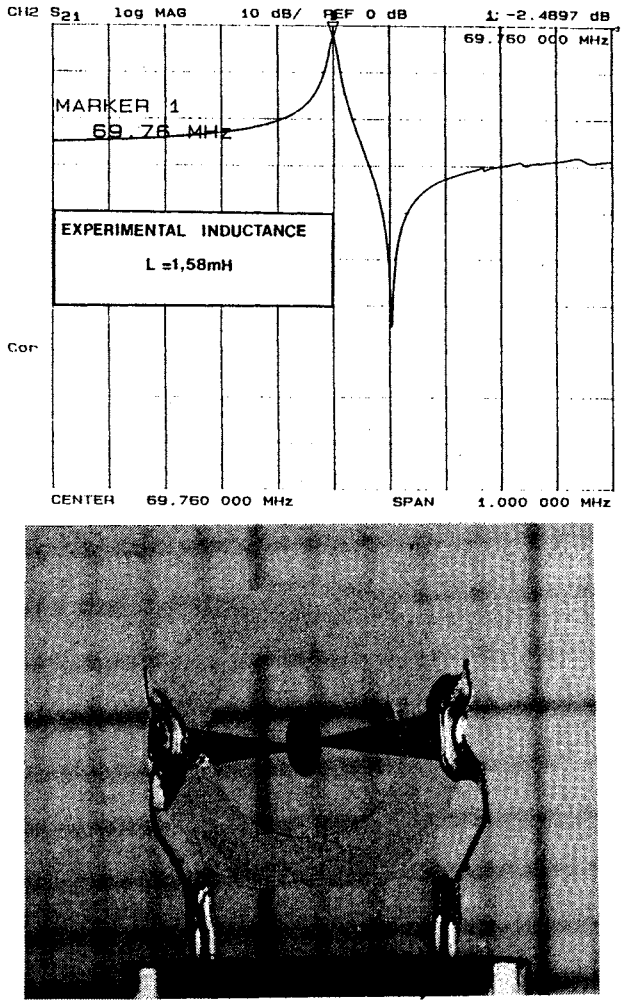


Figure 21: Response curve and photography of a V.H.F. Resonator with an elliptical electrode (0.9*05mm) having a mass loading of 0.5% giving no anharmonics (AT Quartz).

IV Conclusion

An investigation of the properties of VHF resonators using the fundamental mode of the AT cut of berlinite 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 an interesting insensitivity to several technological 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 for a value of the electrode surface that is close to the optimal value. This can permit the optimization of filters integrating several resonators on the same plate.

Acknowledgments: The authors whish to acknowledge MM. X.Buisson and R.Arnaud of SICN for kindly providing the berlinite samples used in this study , They acknowledge the assistance of M. Caumont and Mrs. Daniel for the preparation of the samples used in this study and they thank particularly Mrs M.P.Louis for the preparation 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.P.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 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] T. J.Lukaszek I.E.E.E. trans. Sonics Ultrason. vol. SU 16 n°' p238 (1971)

[9] J.Detaint, H.Carru, J.Schwartzel, C.Joly, B.Capelle, A.Zarka proc. 43th Annual Frequency Control Symposium p.563 (1989)

[10] H.F. Tiersten J. Acoust. Soc. Am. vol. 59 p.879 (1976).

[11] B.Dulmet, H.F.Tiersten Proc. 1986 I.E.E.E. Ultrasonics Symposium p. 355 (1986).

[12] J.Detaint, H.Carru, J.Schwartzel,B.Capelle, A. Zarka proc. 42nd Annual Frequency Control Symposium p. 19 (1988).

[13] K.Nakamura, M. Shimizu proc. 1976 I.E.E.E. Ultrasonics Symposium p.606 (1976).

[14] 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)

[15] R.Holland, E.P. Eernisse Design of Resonant Piezoelectric Devices M.I.T.Press (1969).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 2A Number 8

OBSERVATION OF DEFECTS AND VIBRATION MODES IN LITHIUM TANTALATE

B. Capelle, J. Détaint*, J. Schwartzel*, A. Zarka, Y. Zheng, C. Joly*

Laboratoire de Minéralogie-Cristallographie, Université Paris VI et Paris VII, Associé au CNRS, 75005 Paris France
*Centre National d'Etudes des Télécommunications dept. PAB/BAG/MCT/CMM 92220 Bagneux France

ABSTRACT:

Using the white radiation delivered by the DCI synchrotron at the LURE (Orsay France), the extended defects of lithium tantalate were studied by reflexion and transmission topography. Some low angle subgrains were observed in slices cut perpendicularly to the y and y-45° 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 the first exhaustive 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. To 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 DCI 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 thousand) 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 CNET at the begining 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 extensional 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 (SAW 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 DCI storage ring at the LURE (Orsay France), we have characterized the crystalline defects of experimental and commercial LiTaO3 crystals using either transmission or reflexion topography. The new possibilities given by the intensity and the time structure of the X-ray radiation of the DCI synchrotron, have permitted the first exhaustive 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.

II CHARACTERIZATION OF LiTaO3 CRYSTALS.

To characterize crystals having a 2 inches diameter pulled in the y direction we have cut Y and Y-45°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 .200mm to 0.070mm.

II.1 Experimental settings.

In principle, for crystals too thick to be examined in transmission, a Laue back-setting is used, in which the film is put between the white radiation source and the crystal (Figure 1a). With the synchrotron source, this setting is not interesting to obtain topographs with a high quality because of the large diffusion of the low energy radiations occuring in air. To overcome this limitation, a reflection Laue setting using an oblique incidence was used (Figure 1b). In this setting we have put the film parallel to the sample at a distance (10 to 20 cm) chosen 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 .

II.2 Experimental results.

In the following figures we present typical examples of the observations made. In Figure 2 is represented a reflexion 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

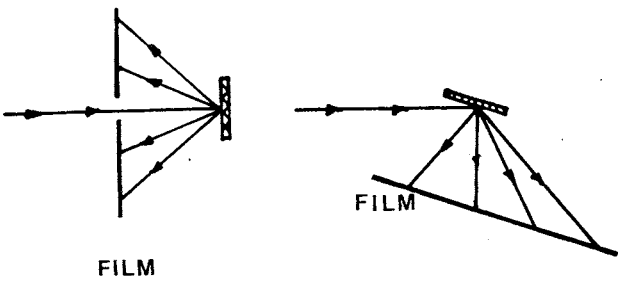


Figure 1: Laue back-setting (1a) and oblique incidence setting (reflection) (1b).

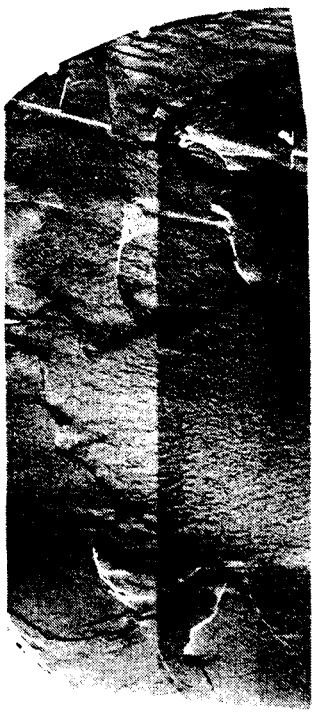


Figure 2: Reflexion topograph of an Y-cut slice.



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

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 cm 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 or 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 centimetre.

In Figure 3 is represented a reflexion topograph covering 3.2cmx1.7cm 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 4 is a transmission topograph of a Y-45°thin (0.100mm) plate, 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.

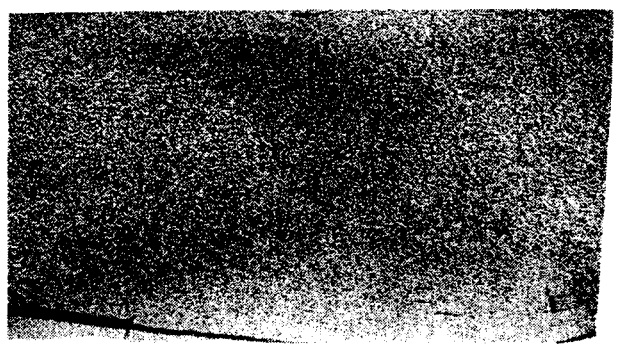


Figure 4: Transmission Laue topograph of a thin (0.100mm) Y-45° plate.



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

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 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.

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 6, a typical response curve of an X-cut resonator (Figure7) 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). 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.

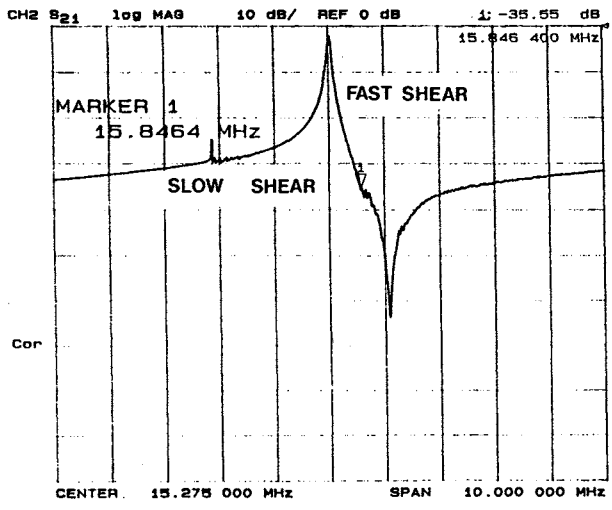


Figure 6: Typical response curve of an X-cut resonator with small electrodes: (a) fundamental of the slow shear. (b) fundamental fast shear.

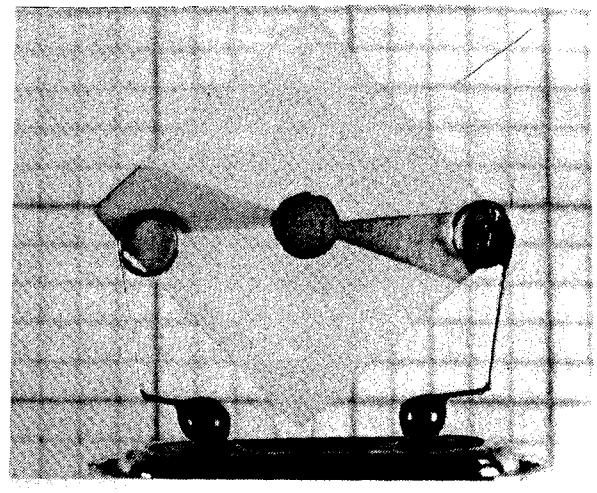


Figure 7: Photography of an X-cut resonator designed to have no anharmonics of the fast shear.

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.

The resonator represented in Figure 8 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 9). 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 6 and 9 has a quite surprising frequency response presenting several anharmonics.

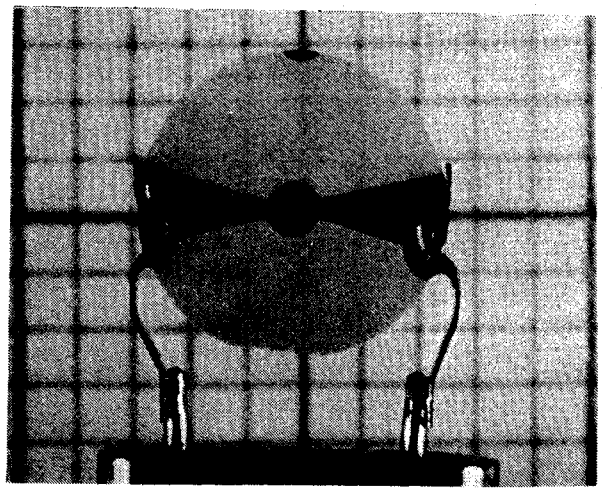


Figure 8: Photography of an X-cut resonator designed to have two symmetrical anharmonics of the fast shear.

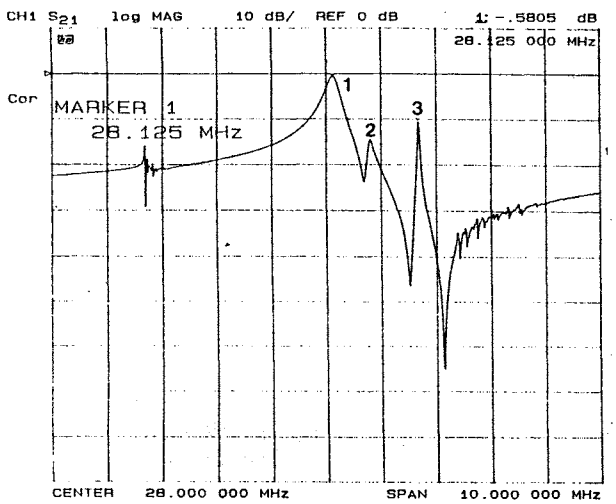


Figure 9: Electrical response of the resonator represented in Figure 8.

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 10 are obtained. This diagram is relative to a resonator similar to that represented in figure 7 , 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 11. 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.

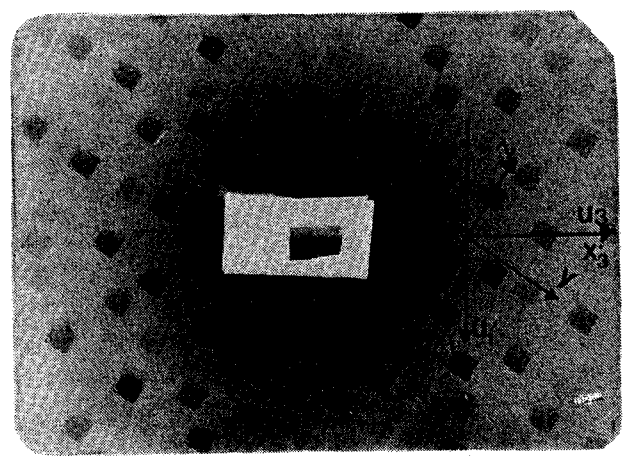


Figure 10: Laue pattern of an X-cut resonator (normal incidence). The projections of u_1 (axe x'_1) and of u_3 (axe x'_3) are drawn on the diagram.

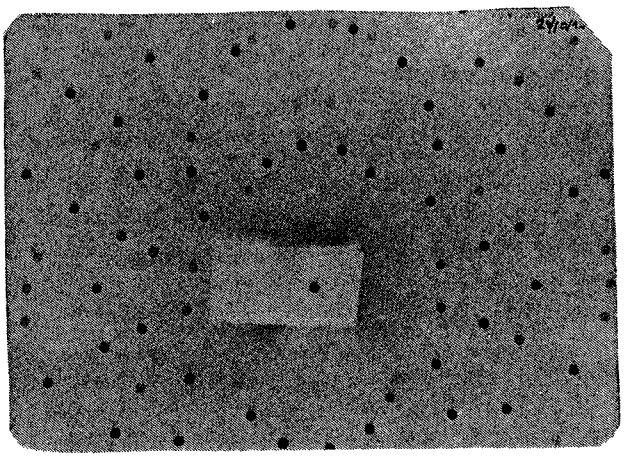


Figure 11: Laue pattern of an X-cut resonator (normal incidence and reduced sample- film distance).

III.3 Fast shear mode of a resonator with small and symmetrical electrodes.

On figure 12 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 13.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 13.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.

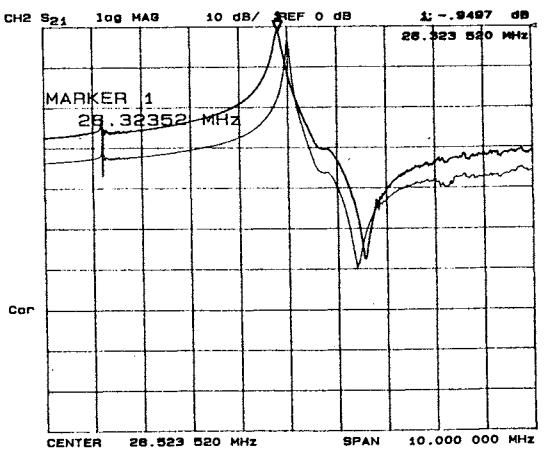
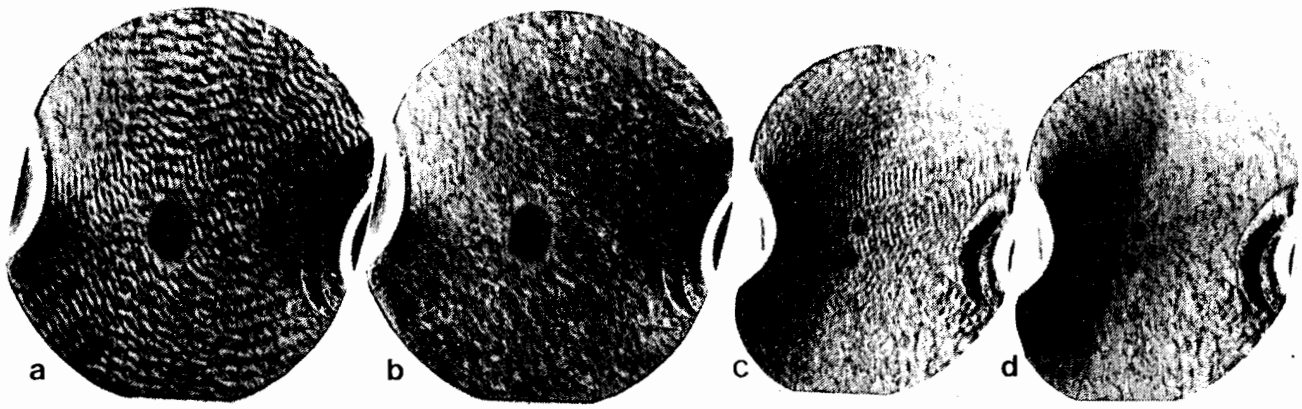


Figure 12: Electrical response of a resonator with small and symmetrical rectangular electrodes .



13a: u_1 time resolved observation (stroboscopic topograph)

13b: u_1 time integrated observation. (Conventional topograph)

13c: u_3 time resolved observation. (Stroboscopic topograph)

13d: u_3 time integrated observation. (Conventional topograph)

Figure 13: Observed fast shear mode of a resonator having symmetrical small electrodes .

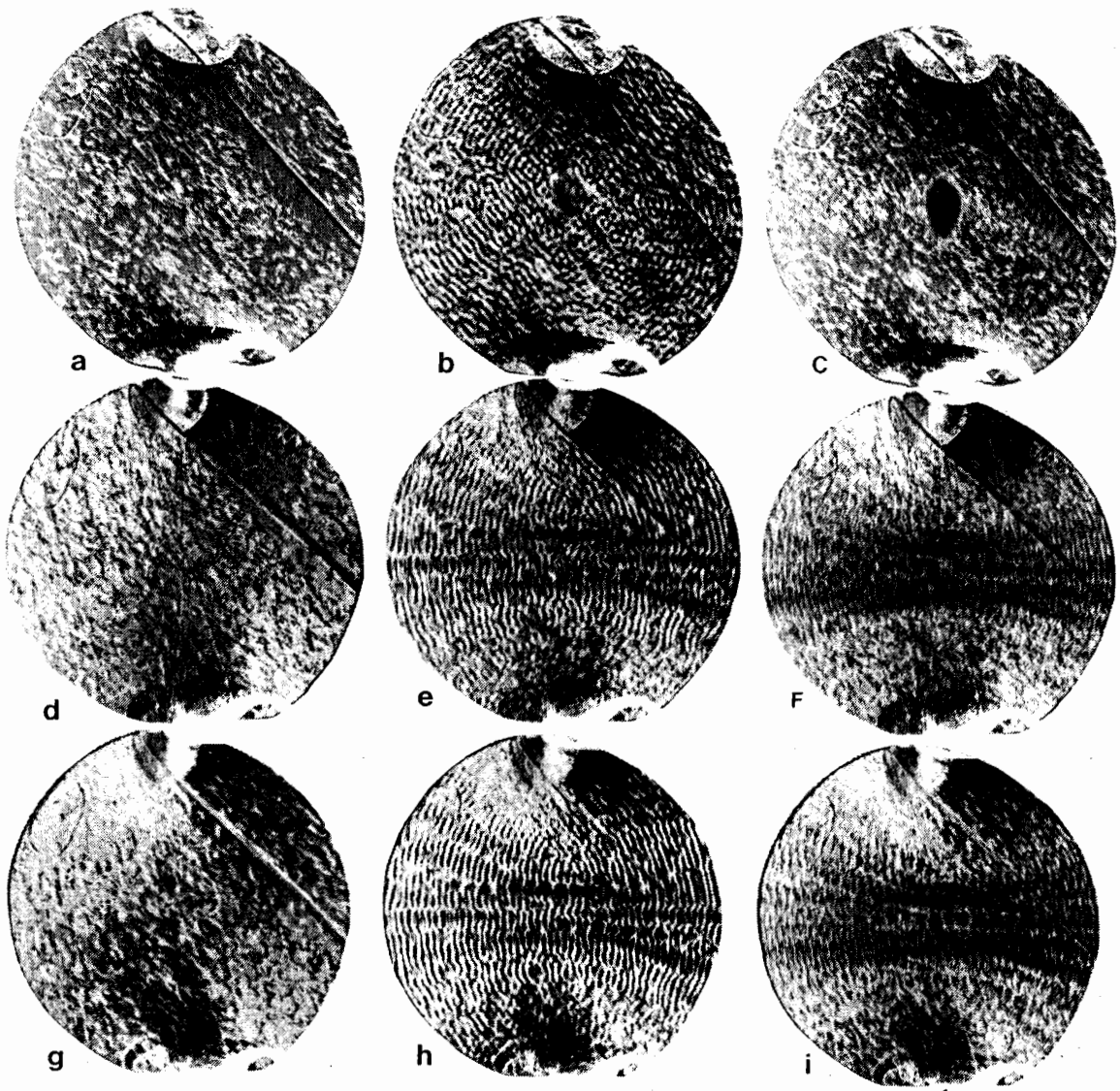


Figure 14: Observation of the fundamental fast shear mode in a resonator with large symmetrical elliptical electrodes.
14a, b, c: images obtained with a diffraction vector nearly in the u_1 direction.
14d, e, f: images obtained with a diffraction vector nearly in the u_3 direction.
14g, h, i: images obtained with a diffraction vector close to the u_3 direction.
(In the columns: without excitation, stroboscopic topographs, conventional topographs).

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 13.c and 13.d 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).

III.4 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 14 for the fundamental fast shear mode. Three diffraction vectors are considered in this figure (one on each rank of the figure) . The first one (14.a, 14.b, 14.c) 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 14.d, 14.e. 14.f and figures 14.g, 14.h, 14.i). The three columns of this figure correspond respectively to topographs obtained without excitation, stroboscopic topographs and conventional topographs. On the topograph taken without vibration (first column), it is possible to observe the characteristic contrasts of the dislocations existing in the plate and one subgrain boundary. In figure 14a and partially in figure 14.g, these contrasts are inverted. This results of the very important value of the μxt 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 14.c) 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 14.f and 14.i). 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 14.b) and u_3 (Figures 14.e and 14.h) progressive components. Although no separate examination of the u_2 component have been made (it requires reflexion topography), the examination of the diffractions spots on the transmission topographs seems indicate that this component is also non negligible.

III.5 Mode of a resonator having symmetrical larges electrodes (anharmonics of the fast shear).

On Figure 15 are represented the u_1 and u_3 components of the three fast shear modes of a resonator similar to that of figure 8 having large electrodes . These components were observed using the conventional topography technique. Figure 15.a, 15.b, 15.c 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 figure in figures 15.b and 15.c the fundamental mode has a weak trapped u_3 component (Figure 15.d) whereas the two anharmonics have large u_3 stationary components extending up to fixation of the plate (Figure 15.e and 15.f).

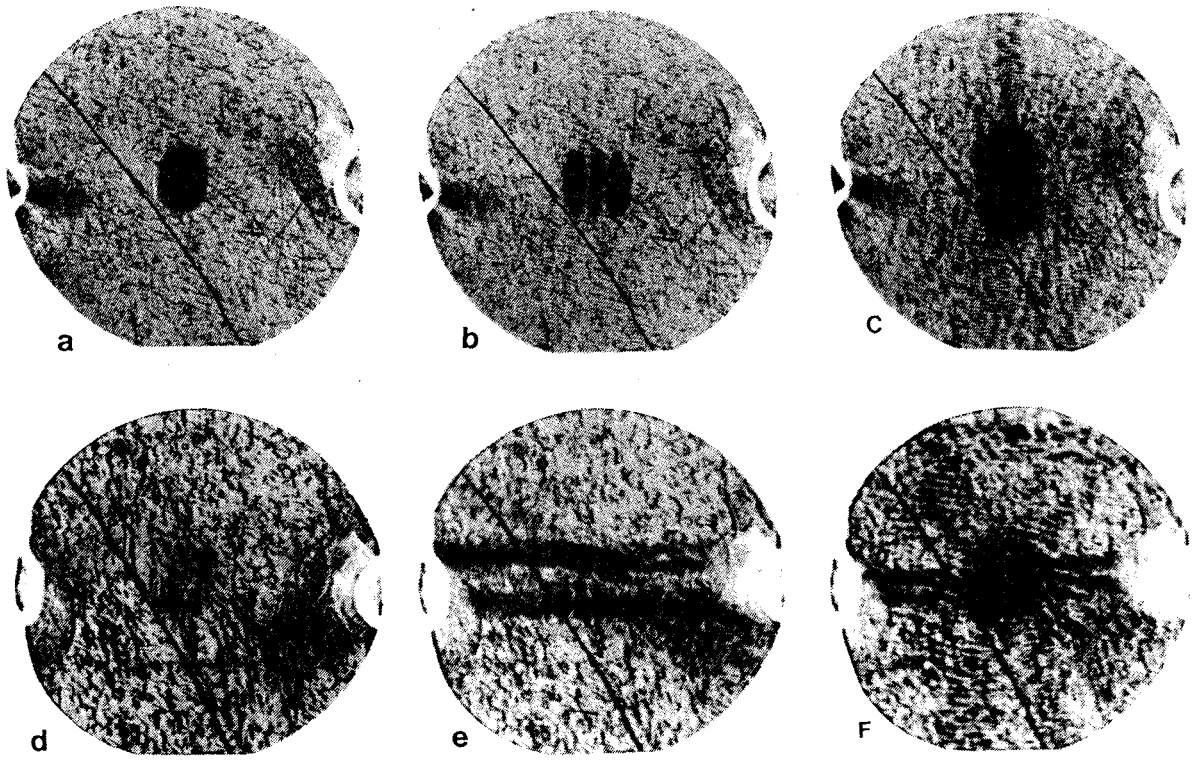


Figure 15: The fundamental and the two trapped anharmonics of the fast shear.
15a, b, c: the u_1 components of the three modes.
15d, e, f: the u_3 components of the three modes.

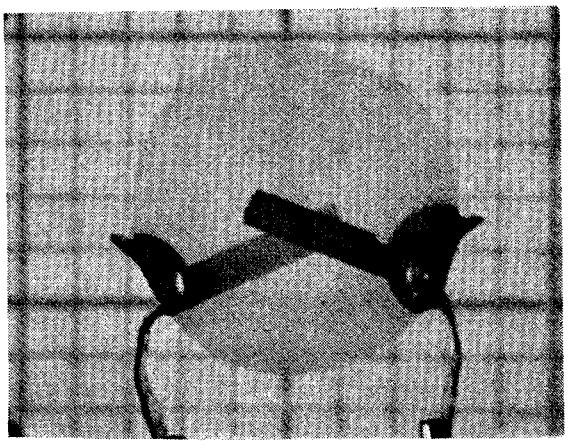


Figure 16: Photography of a resonator having unsymmetrical electrodes not centred in the plate.

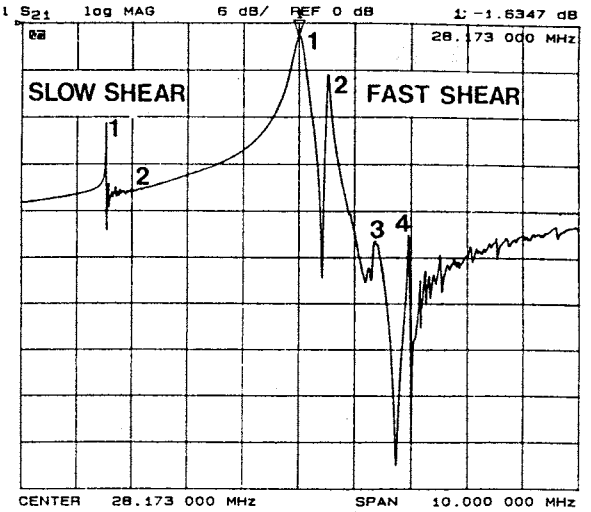


Figure 17: Electrical response of the resonator represented in Figure 16. (fundamental slow shear, fundamental fast shear and theirs anharmonics).

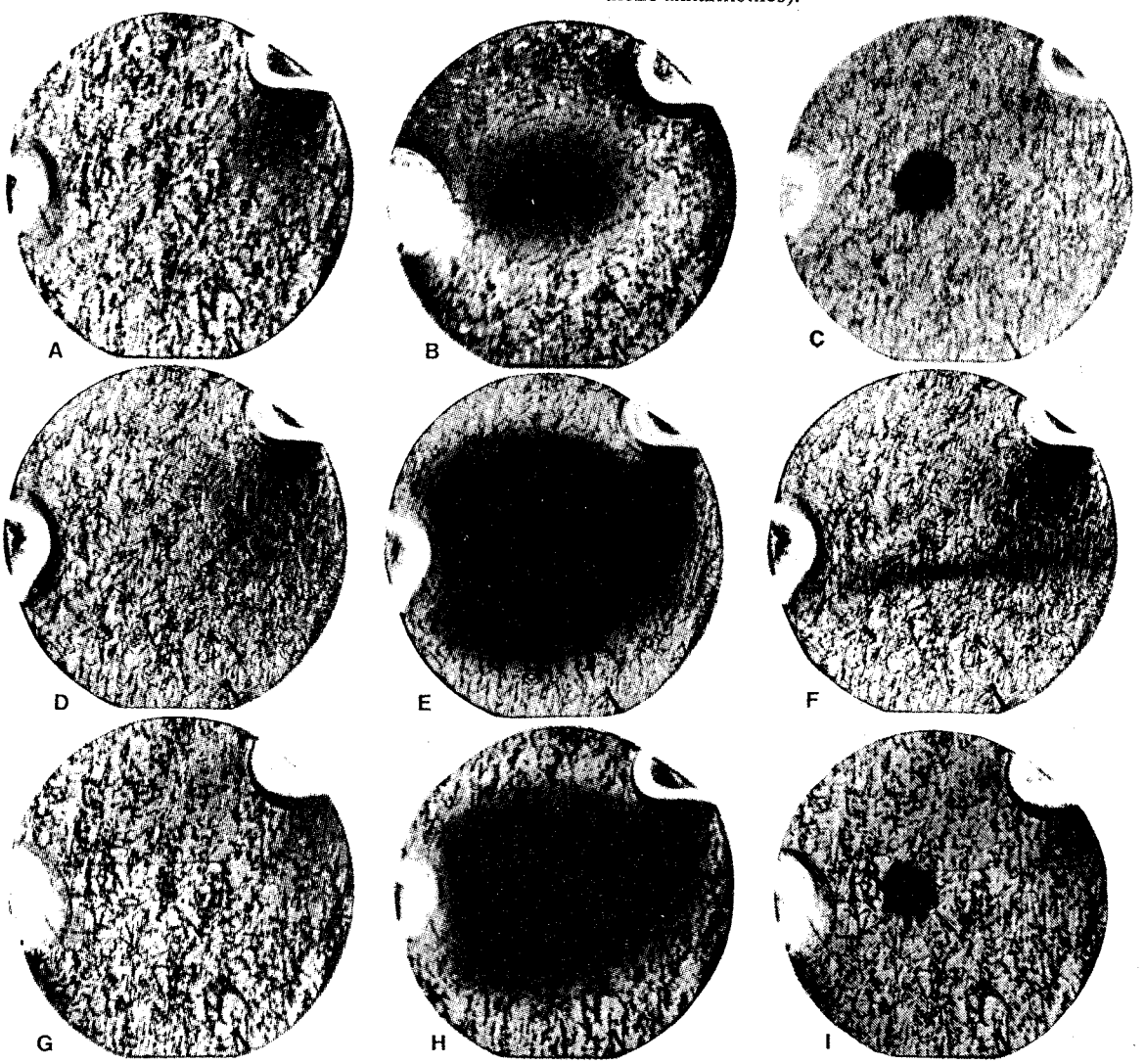


Figure 19: Topographs of the resonator represented in Figure 16, without vibration, fundamental slow shear mode and fundamental fast shear mode.

19a, b, c: images obtained with a diffraction vector nearly in the u_1 direction.
19d, e, f: images obtained with a diffraction vector nearly in the u_3 direction.
19g, h, i: images obtained with a diffraction vector in a composite direction .
(in the columns: without excitation, slow shear, fast shear).

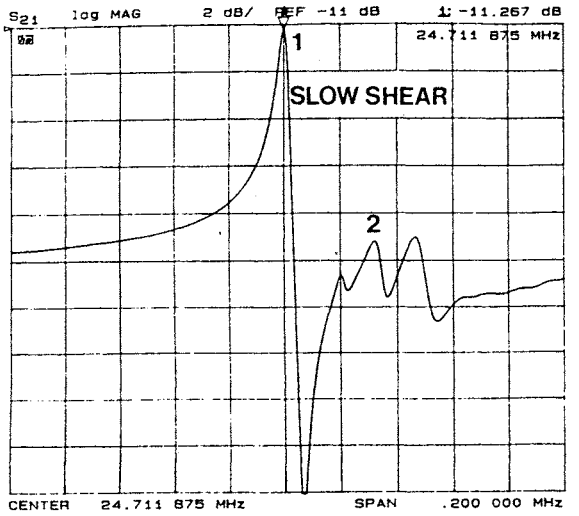


Figure 18: Electrical response near the slow shear.

III.6 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 16). The response curve of this resonator is given on Figure 17. 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 18.

In figure 19 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 19 are respectively represented the u_1 component (19b), the u_3 component (19e) and a mix of the three components (19i) 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

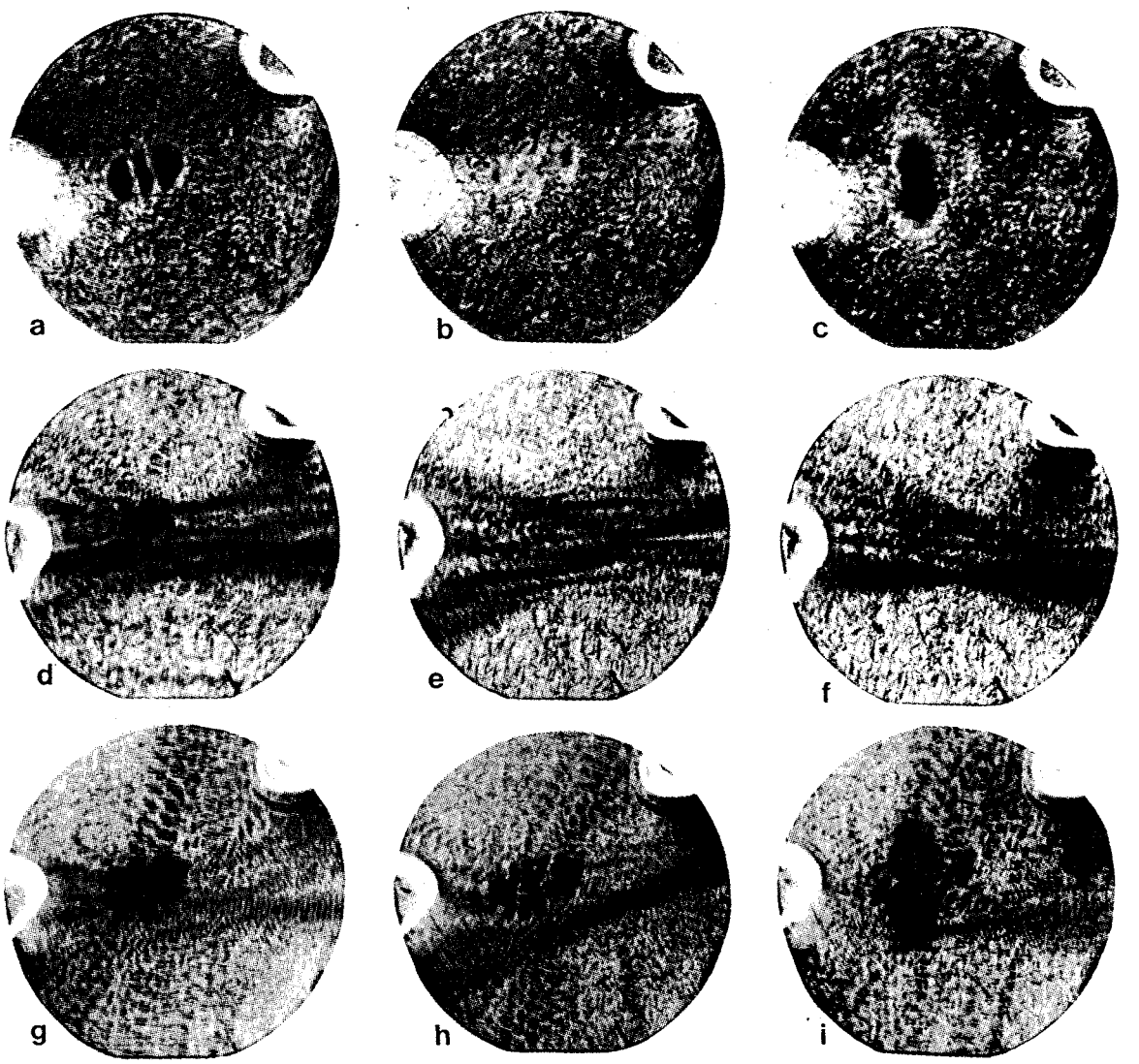


Figure 20: The fast shear anharmonics of the resonator with unsymmetrical electrodes.
20a, b, c: modes imaged with a diffraction vector nearly in the u_1 direction.
20d, e, f: modes imaged with a diffraction vector nearly in the u_3 direction.
20g, h, i: modes imaged with a diffraction vector in a composite direction.

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.

Conventional topographs of the three anharmonics of the fast shear are given in Figure 20. 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 15) . 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.

VI CONCLUSION.

In this study the crystalline perfection of LiTaO3 sample was examined using synchrotron radiations X-ray topography. In thick plates imaged in reflexion, 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 large intensity and the time structure of the X-ray radiation of the DCI synchrotron of the L.U.R.E. (Orsay France) have permitted to make the first exhaustive 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.

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 p132-140 (1976).
[5] P. Auvray CNET/LAB/OCM in reférence [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.416-423 (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.337-348 (1990).
[9] H.F.Tiersten J. Acoust. Soc. Amer. vol. 59 n°4 p.879 (1.976).
[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).
[14] A. Zarka, B.Capelle, J.Détaint, J.Schwartzel J.Appl. Cryst. vol.21 p.967 (1988).

Session 2A Number 9

SENSITIVITY OF BAW DEVICES TO RADIAL IN-PLANE STRESS DISTRIBUTION, COMPARISON BETWEEN ANALYTICAL AND FINITE ELEMENT RESULTS

Sylvain Ballandras

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

Abstract

Usually sensitivity to radial in-plane stress distribution for Bulk Acoustic Wave (BAW) devices is calculated using simple stress models. Recent developments in numerical reckoning techniques yield to 3-dimensions anisotropic models, especially by using finite element methods. The present work is based on analytical calculations for the unperturbed piezoelectric BAW propagation and for static stress distributions, holding anisotropy into account. We study the case of radial in-plane stress distribution using Janiaud's calculation, which is an analytical 2D anisotropic one. Our goal is to estimate differences between recent results obtained by finite element analysis and analytical solutions.

1. Introduction

Sensitivity to radial in-plane stress distributions for Bulk Acoustic Waves (BAW) devices has been studied by many authors, using simple stress models [1] or more complicated ones [2, 3]. E.P. Eernisse showed [4] that a circular plane-sided quartz resonator of particular cut angles (SC-cut) is low sensitive to double radial compression because of its anisotropic properties. Recent developments in numerical reckoning techniques yield to modelized quartz resonators submitted to diametrical forces, taking anisotropy into account. With such new methods like finite element analysis, one can predict frequency shifts of the resonator due to mechanical biases with a good precision.

Present study is based on an analytical calculation for the static stress distribution due to diametrical forces applied on the circular plate. The model has been presented yet [5] but was used with a second order approximation which yield to inaccurate results particularly in the case of doubly rotated cuts. Present calculation is mathematically rigourous and our results can be compared to these obtained by P.C.Y. Lee and M.S.H. Tang [6] who used finite element analysis to estimate frequency shifts due to radial in-plane stress distributions.

2. Perturbation method

When the problem of unperturbed propagation is resolved, equilibrium equations can be written as

$$\rho_o \frac{d^2 u_i^{(s)}}{dt^2} = C_{ikjm} \frac{\partial^2 u_j^{(s)}}{\partial a_k \partial a_m} \qquad (1)$$

with

$$u_i^{(s)} = u_i^{(s)} e^{-j\omega\left(t - \frac{n_\ell - a_\ell}{V^{(s)}}\right)} \qquad (2)$$

where u_i is the mean magnitude of displacements, $V^{(s)}$ is the velocity of the wave and n_ℓ are the cosine director. Upper subscript (s) represent one of the three possible propagation modes.

When the device is prestrained, equation (1) must be modified to take into account nonlinear effects due to the physical properties of the crystal. This yields to the following equation

$$\rho_o \frac{d^2 u_i^{(s)}}{dt^2} = \overline{A_{ikjm}} \frac{\partial^2 u_j^{(s)}}{\partial a_k \partial a_m} \qquad (3)$$

where $\overline{A_{ikjm}}$ are the modified elastic constants and may be written as

$$\overline{A_{ikjm}} = C_{ikjm} + \overline{H_{ikjm}} \qquad (4)$$

with $\overline{H_{ikjm}}$ the perturbation tensor which is function of stresses, $\overline{T_{km}}$ strains $\overline{S_{uv}}$ and gradient displacements $\partial \overline{u_i}/\partial a_p$

$$\overline{H_{ikjm}} = \delta_{ij} \overline{T_{km}} + \frac{\partial \overline{u_i}}{\partial a_p} C_{pkjm} + \frac{\partial \overline{u_j}}{\partial a_q} C_{ikqm} + C_{ikqmuv} \overline{S_{uv}} \qquad (5)$$

Note that C_{ikjmuv} are the 3rd nonlinear elastic constants.

Applying a perturbation method found and developed by H.F. Tiersten [7] for BAW propagating on a biased middle of volume v yields to the relative frequency shift $\Delta F/F_o$ equation

$$\left(\frac{\Delta F}{F_o}\right) = \frac{1}{2\rho_o \omega_o^2} \frac{\iiint_v \frac{\partial u_i^{(s)}}{\partial a_k} \overline{H_{ikjm}} \frac{\partial u_j^{(s)}}{\partial a_m} dv}{\iiint_v u_\ell^{(s)} u_\ell^{(s)} dv} \qquad (6)$$

If static terms are assumed to be homogeneous in the volume v, equation (6) reduces to the following formula

$$\left(\frac{\Delta F}{F_o}\right)^{(s)} = U_{ikjm}^{(s)} \overline{H_{ikjm}} \qquad (7)$$

where U_{ikjm} is a tensor that depends on the unperturbed wave parameters

$$U_{ikjm}^{(s)} = \frac{\ell_i^{(s)} \ell_j^{(s)} n_k n_m}{2\rho_o V^{(s)2}} \qquad (8)$$

with $n_1 = 0$, $n_2 = 1$, $n_3 = 0$ in the plate axes (Fig.1). $\ell_i^{(s)}$ represent the eigenvectors in the same set of axes.

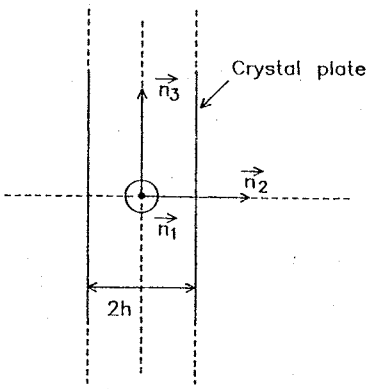


Fig. 1 : Representation of the plate axes

If only the stress distribution is known as in previous diametrical compression models [4], one have to reduce the complexity of equation (5) by using following assumptions :

- there is a fixed point in the center of the plate, where no displacement can take place.
- there is no rigid rotation around this point, and as a consequence of this, $\partial \overline{u_i}/\partial a_j = \partial \overline{u_j}/\partial a_i$.

By applying Hooke's law in (5), the new expression of the frequency shift is obtained as a function of the static stress distribution

$$\left(\frac{\Delta F}{F_o}\right)^{(s)} = {}^{s}a_{\ell n}^{(s)}\ \overline{T_{\ell n}} \qquad (9)$$

where a new tensor appears, ${}^{s}a_{\ell n}{}^{(s)}$ written for the (s) mode, which is only dependent on mechanical properties of the material and of the unperturbed wave parameters

$$ {}^{s}a_{\ell n}^{(s)} = (\delta_{ij}\delta_{k\ell}\delta_{mn} + C_{pkjm}S_{ip\ell n} + C_{ikqm}S_{jq\ell n} + C_{ikqmuv}S_{uv\ell n})\,U_{ikjm}^{(r)} \qquad (10)$$

3. Anisotropic diametrical compression

In this part, Janiaud's model [8] will be used to calculate the static stress distribution on a 2 dimension circular plate. We assumed that the effects of radial compression are homogeneous in regard with depth coordinate a_2.

As in the isotropic case [9], the stress distribution in a point M (Fig. 2) due to the application of a force F on the boundary of a semi-infinite plate is assumed to be as following

$$\overline{\sigma_r} = \frac{A(\theta)}{r} \; ; \; \overline{\sigma_\theta} = \overline{\tau_{r\theta}} = 0 \qquad (11)$$

where $\overline{\sigma_r}$, the radial stress, is the only nonzero term and is a function of the angular position θ of the point M and of its distance r to the point where the force F is applied.

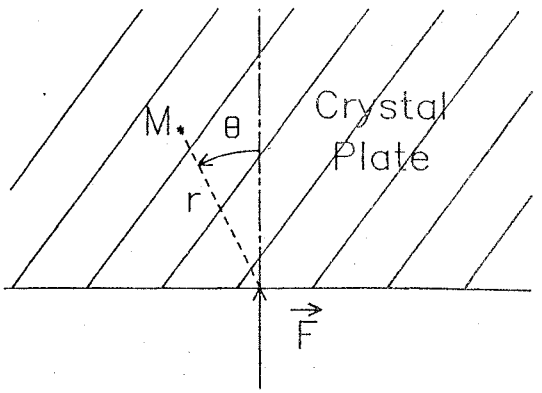


Fig. 2 : Radial force applied on a semi-infinite plate

It has been shown [5], using static equilibrium equations, that the compatibility equation in the case of a 2 dimension anisotropic disk takes the following form

$$\frac{d^2(C_1\,A(\theta))}{d\theta^2} + C_1\,A(\theta) = 0 \qquad (12)$$

where C1 is function of the material stiffness and of the angular coordinate θ. This equation can be easily resolved, and the general solution for the radial stress may be written

$$\overline{\sigma_r} = \frac{\lambda\cos\theta + \mu\sin\theta}{r(1 + \alpha\cos(2\theta) + \beta\cos(4\theta) + \gamma\sin(2\theta) + \delta\sin(4\theta))} \qquad (13)$$

with

$$\alpha = \frac{4(s_{11} - s_{33})}{2s_{13} + s_{55} + 3s_{11} + 3s_{33}}$$

$$\beta = \frac{s_{11} + s_{33} - 2s_{13} + s_{55}}{2s_{13} + s_{55} + 3s_{11} + 3s_{33}}$$

$$\gamma = \frac{4(s_{15} + s_{35})}{2s_{13} + s_{55} + 3s_{11} + 3s_{33}}$$

$$\delta = \frac{2(s_{15} - s_{35})}{2s_{13} + s_{55} + 3s_{11} + 3s_{33}}$$

λ and μ are two coefficients resulting from the integration of equation (12) and may be calculated in regard with boundary conditions at the point where F is applied. Equilibrium of the radial stress on a semi-circle from $-\pi/2$ to $\pi/2$ must be written for both projections of σ_r along the direction of F and along the normal of this direction.

This yields to the two following integrals :

$$\int_{-\pi/2}^{+\pi/2} \frac{\lambda\cos^2\theta + \mu\sin\theta\cos\theta}{1 + \alpha\cos 2\theta + \beta\cos 4\theta + \gamma\sin 2\theta + \delta\sin 4\theta}\,d\theta = -F \qquad (14)$$

$$\int_{-\pi/2}^{+\pi/2} \frac{\lambda\cos\theta\sin\theta + \mu\sin^2\theta}{1 + \alpha\cos 2\theta + \beta\cos 4\theta + \gamma\sin 2\theta + \delta\sin 4\theta}\,d\theta = 0$$

Using classical variable substitutions $t = \text{tg}\,\theta$, expression (14) yields

$$\int_{-\infty}^{+\infty} \frac{\lambda + \mu t}{A t^4 + B t^3 + C t^2 + D t + E}\, dt = -F$$

$$\int_{-\infty}^{+\infty} \frac{t(\lambda + \mu t)}{A t^4 + B t^3 + C t^2 + D t + E}\, dt = 0 \qquad (15)$$

with $A = 1 - \alpha + \beta$, $B = 2(\gamma - 2\delta)$, $C = 2(1 - 3\beta)$, $D = 2(\gamma + 2\delta)$, $E = 1 + \alpha + \beta$,

which can be resolved using residuals integral method. Note that solution of a 4th order algebric equation can be analytically found [10], but we used some numerical algorithm to obtain these solutions, with a good precision.

Let's have a look now to the problem of a quartz disk submitted to single in-plane diametrical compression (Fig. 3).

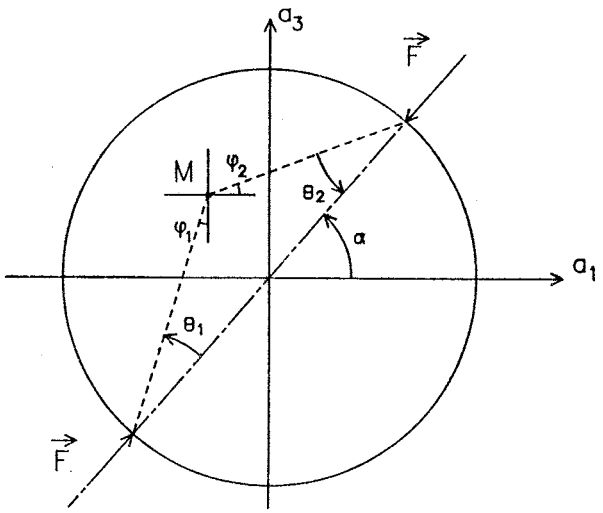


Fig. 3 : Single diametrical compression

In that case, the previous calculation is applied at the center of the disk, where the principal perturbation is supposed to take place. The condition of zero stress on the disk boundaries has to be taken into account. It is respected by adding a constant term at each non-zero stress component when the stress field is expressed at the disk boundaries. Then the stress distribution may be written at the center of the plate

$$\overline{T_{11}^o} = \frac{P}{\pi d t}\left(\frac{4\lambda}{1 + \alpha + \beta} - \frac{\lambda + \mu}{1 - \beta + \gamma}\right)$$

$$\overline{T_{33}^o} = -\frac{P}{\pi d t}\left(\frac{\lambda + \mu}{1 - \beta + \gamma}\right) \qquad (16)$$

$$\overline{T_{13}^o} = 0$$

Now the stress field can be expressed when the force is applied with an arbitrary angle Ψ in respect with the a_1 axis. This yields to the frequency shift equation versus Ψ

$$\left(\frac{\Delta F}{F_o}\right)^{(s)} = ({}^s a_{11}^{(s)} + {}^s a_{33}^{(s)})\left(\frac{\overline{T_{11}^o} + \overline{T_{33}^o}}{2}\right)$$

$$+ ({}^s a_{11}^{(s)} - {}^s a_{33}^{(s)})\left(\frac{\overline{T_{11}^o} - \overline{T_{33}^o}}{2}\right)\cos 2\beta$$

$$- {}^s a_{13}^{(s)}\left(\overline{T_{11}^o} - \overline{T_{33}^o}\right)\sin 2\beta \qquad (17)$$

Note that in the case of a double diametrical compression and if the force directions are perpendicular then the dependence of the frequency shift to angle Ψ reduces to

$$\left(\frac{\Delta F}{F_o}\right)^{(s)} = ({}^s a_{11}^{(s)} - {}^s a_{33}^{(s)})\left(\frac{\overline{T_{11}^o} + \overline{T_{33}^o}}{2}\right) \qquad (18)$$

that was already true in the isotropic case. There is still a residual dependence to angle Ψ due to the fact that $\overline{T}_{11}{}^o$ and $\overline{T}_{33}{}^o$ must be calculated for each step of Ψ.

When the coefficient ${}^s a_{11}{}^{(s)}$ is equal to ${}^s a_{33}{}^{(s)}$, the device will be theoreticaly insensitive to double diametrical compression. This criterium has been used to point out wich cuts are low sensitive to this kind of mechanical bias (Fig.4).

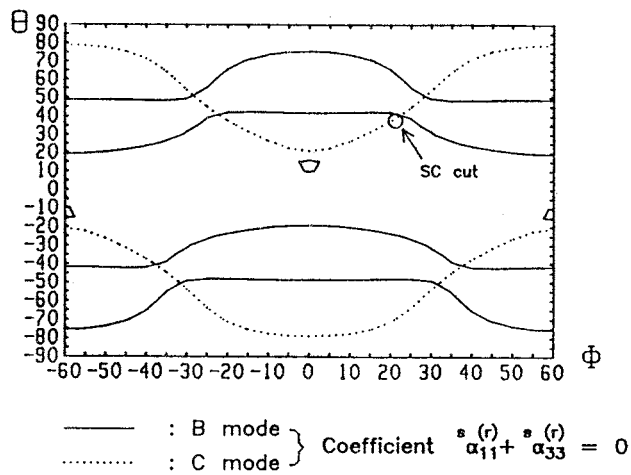
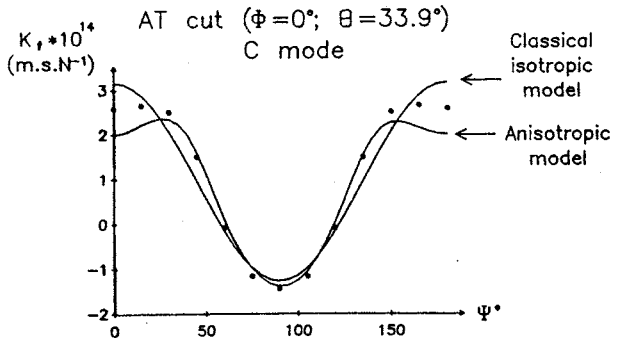


Fig. 4 : Quartz cuts insensitive to double diametrical compression

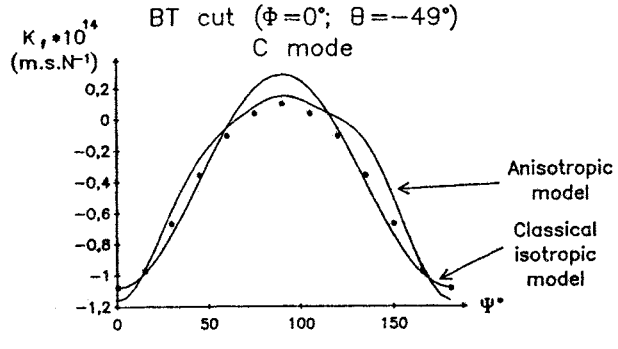
One can observe that there is no compensation effect for A mode. The SC cut determined by E.P. Eernisse is localized close to an intersection between B and C mode compensation effect. Now the present model can be computed and results may be compared to P.C.Y. Lee and M.S.H. Tang results obtained by using finite element analysis. This was made on a DATA GENERAL MV 1500/100 computer, and it takes almost a few second of CPU time to modelize one cut's behaviour when submitted to radial compression. Results appear on Fig. 5. They are given with the use of the classical force sensitivity coefficient K_F for BAW devices. This coefficient, $KF^{(s)}$ for the (s) mode, is related to the corresponding frequency shift as following

$$K_F^{(s)} = \frac{2d}{V_o^{(s)} P}\left(\frac{\Delta F}{F_o}\right)^{(s)} \qquad (19)$$

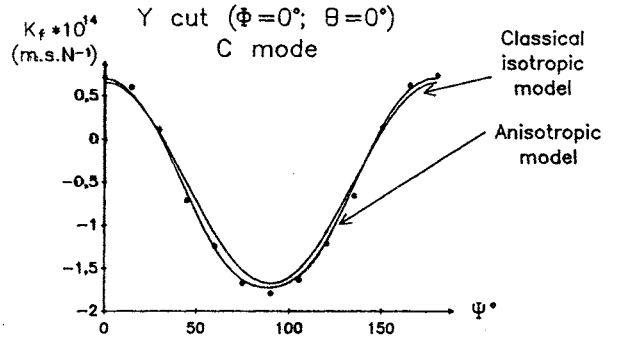
with $P = F/t$ the force by thickness unit and $V_o{}^{(s)}$ the wave velocity of the (s) mode for the unperturbed state.



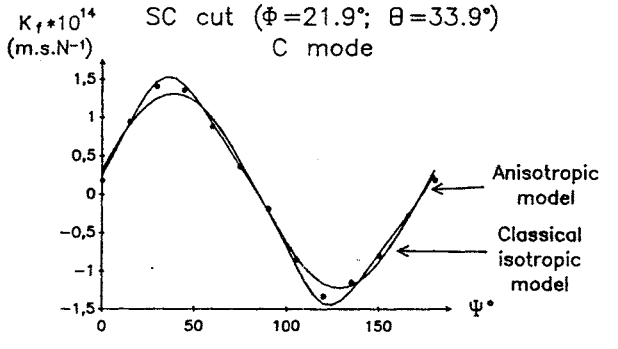
• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method



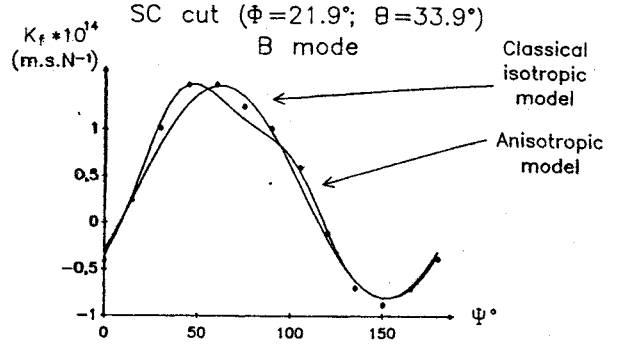
• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method



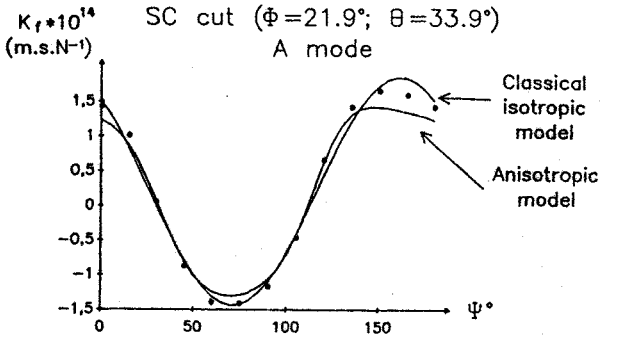
• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method



• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method



• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method



• : P.C.Y. LEE and M.S.H. TANG analysis using finite element method

Fig. 5 : Comparison between isotropic, anisotropic and finite element analysis results for AT, BT, Y and SC cuts

Classical isotropic model, Lee and Tang's finite element results and the present model have been superimposed on Fig.5. First, the good agreement between all the models i terms of phase and magnitude may be remarked. Nevertheless, significant differences between the three calculations are observed for particular azimuthal angle Ψ in the case of AT and BT cuts. Note that if the difference between the two analytical models reaches 25 % of magnitude, it reduces to 15 % when each of them is compared to finite element analysis (AT cut, Ψ = 0 to 50° ; BT cut, Ψ = 120 to 150°). In the case of cuts which present a quite isotropic behaviour, the fit between different results is fairly good. Calculations for SC cut, B and A modes, have been also computed. In the two cases, it appears the same difference between the models for particular Ψ values as in the AT and BT cuts. Even for A mode, we are close to a compensation effect for the double radial compression.

Conclusion

It has been shown that an anisotropic model permit to predict the existence of low sensitive cut to double diametrical compression, which is in good agreement with isotropic analysis and previous experiments. We point out differences between isotropic, finite element and present anisotropic results, but we can not establish at present time which calculation is closer to reality. Other analytical method should be investigated, and comparison with a statistically representative number of experiment could light our analysis.

Acknowledgments

The author would like to thank Mr. J.J. Gagnepain for its interest to this work, and instructive discussion about this problem and its physical aspects.

Bibliography

[1] J.P. Perdigues, "Etude de l'influence des forces sur la fréquence d'un résonateur à quartz", Thèse D.I. n° 58, Besançon, 1975.

[2] P.C.Y. Lee, Kuang-Ming Wu, "In-plane accelerations and forces on frequency changes in doubly-rotated quartz plates", J. Acoust. Soc. Am. 75 (4), 1105-1117, april 1984.

[3] R. Bourquin and B. Dulmet, "Force sensitivity of trapped energy vibrations in a contoured resonator", 41st Ann. Freq. Cont. Symp. 1987, pp. 289-294.

[4] E.P. Eernisse, "Calculation on the stress compensated (SC-cut) quartz resonator", Proc. of the 30th Ann. Freq. Cont. Symp. 1976.

[5] D. Janiaud, L. Nissim and J.J. Gagnepain, "Analytical calculation of initial stress effects on anisotropic crystals : application to quartz resonators", Proc. 32nd Ann. Freq. Cont. Symp. 1978, pp. 169-179.

[6] P.C.Y. Lee and M.S.H. Tang, "Thickness vibrations of doubly rotated crystal plates under initial deformations", IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control UFFC-34 (6), 659-666, 1987.

[7] H.F. Tiersten, J.C. Baumhauer, "Nonlinear electrostatic equations for small fields superposed on a bias", J. Acoust. Soc. Am. 54, 1017, 1973.

[8] D. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thèse, Université de Besançon, 1978.

[9] S.P. Timoshenko, J.W. Goodler, "Theory of elasticity", 3rd Edition, Mc Graw Hill Ed., 1970.

[10] J. Dieudonné, "Pour l'honneur de l'esprit humain (les mathématiques aujourd'hui)", Editions Pluriel, Hachette, 1987.

SESSION 3A

SAW DEVICES

Chairman : D. HAUDEN

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 3A Number 10

S.A.W. FILTERS FOR MOBILE RADIO

Philippe Defranould, Martine Doisy, Pierre Dufilié*, Jean-Michel Hodé

THOMSON-SINTRA ASM / D.T.A.S
399, Route des Crêtes - BP 38 - Parc de Sophia-Antipolis
06561 VALBONNE CEDEX FRANCE

* THOMSON-ICS CORP., P.O. Box 1088, Route 57, SOUTHWICK, MA 01077 USA

Emerging digital cellular radio systems (European 900 MHz and 1800 MHz D-nets) require high performance medium loss or low loss channel filtering. A highly selective filtering is required to keep a good bit error rate and part of this filtering must occur in the IF sections. Insertion loss does not appear as a key point but low or moderate insertion loss is prefered for both power consumption, circuit size and cost considerations. For use in mobile units these filters must be as small as possible and very cheap.

Surface Acoustic Wave (S.A.W.) filters are known for miniaturization, mass production capability and high reproducibility but exhibit fairly large insertion losses (typicaly 20 to 30 dB). Tremendous work has been done for many years all over the world to decrease the S.A.W. filter losses and various technologies are now available to keep insertion losses below 10 dB as required in many communication systems. All these points make the SAW technology a very attractive solution to the difficult problem of the IF channel filtering in mobile telephone systems.

However applications in consumer products such as mobile telephones require very small and inexpensive components. This implies that the low loss structures which are used to design the IF filters for these applications must match compact, single phase and single layer technologies.

This paper reports low loss SAW products which were specifically developed for use in the IF sections of cellular digital radio systems (European 900 MHz D-net also called GSM). Efficient low loss technologies based on Single Phase UniDirectional Transducers (SPUDT's) are used for optimum design of these filters in terms of performance to cost ratio. Insertion losses range from 3 to 10 dB, bandwidths from 130 to 500 kHz and center frequencies from 45 to 130 MHz. Rejections are as high as 50 dB.

Session 3A Number 11

LOW LOSS SAW FILTERS USING MULTIPLE TRANSDUCERS

Abbott B.P, Heighway J, Morgan D.P & Thorvaldsson T.

ASCOM *Microsystems*, Ch. Chapons des Prés, CH-2022, Bevaix

1. ABSTRACT

The use of multiple transducer arrangements with interleaved reflective gratings has been investigated in order to achieve low loss high performance filters at frequencies of around 500 MHz. In particular, a 6 transducer scheme with 3 samples per wavelength has been operated on Lithium Niobate and the results compared to theory. The performance achieved is compared to results obtained for a new NSPUDT cut on Lithium Tantalate for a similar filter. Detailed analyses of both designs are presented that include a variety of second order effects.

2. INTRODUCTION

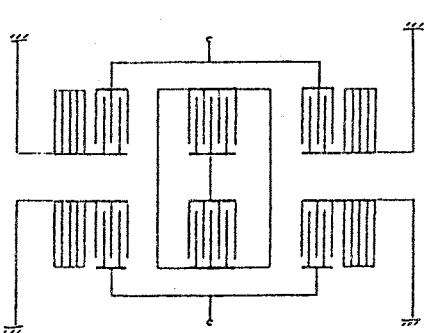
The use of low loss filters in the radio frequency filtering, particularly in mobile systems, has grown significantly in recent years. The emphasis in all of the new applications is on loss of less than 8 dB, high rejection of key frequencies (particularly harmonics) and high operating frequencies. These requirements have led Surface Acoustic Wave (SAW) designers to use novel techniques and new substrates.

We report here on two approaches to a 460 MHz filter that employ the combination of several known SAW techniques plus unusual substrate properties allowing the implementation of miniature, low loss devices using single phase unidirectional transducer (SPUDT) schemes.

3. SIX TRANSDUCER FILTERS

These devices comprise two identical 3-transducer devices occupying parallel tracks. The two central transducers are connected, giving efficient coupling of SAWs from one track to the other at the centre frequency and also helping to suppress the stop-band levels. To do this the transducers must be self-resonant which limits the bandwidth to about 3% on YZ lithium niobate. The outer transducers on each track are connected in parallel to the device input and output (see Fig 1). To reduce the loss, reflecting gratings have been added beyond the outer transducers; the gratings using either short-circuit or open-circuit metal strips. In exchange for reducing the loss, the gratings reduce the device bandwidth - for example bandwidths in the region of 1.5% are theoretically predicted for a device insertion loss of 2 dB.

Fig 1 Six transducer layout schematic



Analysis software has been developed to predict the device response and to serve as a design aid. The analysis takes full account of the multiple SAW reflections within the device and predicts the response fully allowing for matching circuitry. For the central track-coupling transducers, the analysis can, and does, include interactions due to electrode reflections. These are included because they occur in single-electrode transducers, which are easier to fabricate than types which avoid interactions. It is also known, that interactions distort the transducer response in such a way as to improve the track coupling, giving a somewhat larger bandwidth.

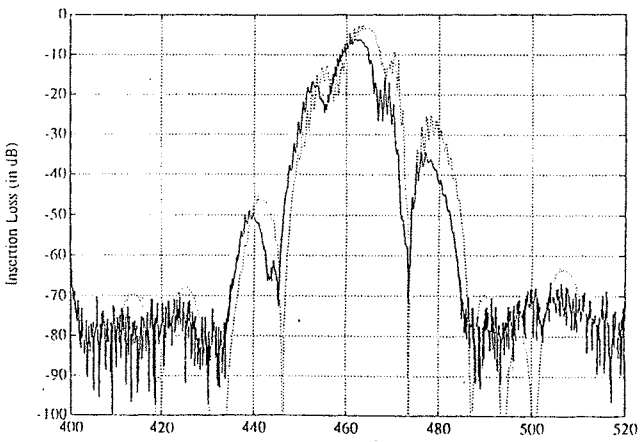
Because the effects of interactions and multiple internal reflections would be more significant (and because the effects are extremely complex) the outer transducers have been laid out as types with minimal interactions i.e. 3 electrodes per period. The transducers were all initially unweighted, although results that incorporate withdrawal weighting in the central transducers are expected. The software can be used to assess the effects of design parameter changes so as to optimise the response - for example, the outer transducers have centre frequencies raised a little because the track coupling was found to be centred at a frequency somewhat above the transducer centre frequency. Devices were designed for both YZ and 128 degree rotated lithium niobate substrates, using a centre frequency of around 460 MHz.

3.1 YZ NIOBATE DESIGN

The central track-coupling transducers for this design had 67 electrodes giving about 3% bandwidth, a value determined approximately by the coupling constant of the substrate and the transducer structure (2 electrodes per period). The outer transducers had 15 periods with 463 MHz centre frequency, and used 3 electrodes per period to avoid interaction effects. The gratings on the outer transducers have 50 metal strips with a half-wavelength period; the strips are shorted together. The strip reflection coefficient is negative imaginary for this case, implying that the grating must be displaced $\lambda/8$ away from the transducer to give the correct phase relationship. The strip reflection coefficient was estimated at -1.5%, based on electrical loading only. In practice, the film thickness used (nominally 2000 Angstroms) implies that the magnitude may be somewhat higher due to additional mechanical loading.

The outer transducers, with their gratings, have a response giving two close-in sidelobes in addition to the main peak. The sidelobes are a consequence of the rather long grating structure, occuring when the distance the centres of the transducer and grating has increased by $\lambda/2$. The sidelobes would be less detrimental to the device response if a shorter grating were used, but this would result in an increased device loss and for this design the loss was given higher priority. The outer transducers can be matched to give either minimum loss or minimum reflection (giving lower ripple in the device response). Again, loss was given priority here, and the aperture (0.4 mm) was chosen for minimum loss with a 50 ohm impedance. This required a parallel inductor of 19 nH. Parallel tuning was chosen in preference to series because it enables any stray capacitances to be compensated for by changing the inductor value. The design was carefully laid out using wide bus bars to minimise losses due to film resistance.

Fig 2 YZ Niobate comparison of experiment (___) and theory(.....)



Devices were fabricated and assembled in TO8 packages and comparison between experiment and theory are shown in Fig 2. The close-in sidelobes, due to the outer transducers and gratings, were 10-15 dB below the peak and at about +/- 8 MHz from the centre frequency. More remote sidelobes were at -25 and -40 dB. The agreement with the theoretical prediction (with no inductors), is considered good. The preliminary model does not include a number of effects but these have subsequently been added. These have included electrical resistance in the track coupling transducers, variation of parameters in these transducers and variation of the electrode reflection coefficient in transducers and gratings; track resistance of 7ohms and a stray C of 4 pF has been assessed for the current layout.

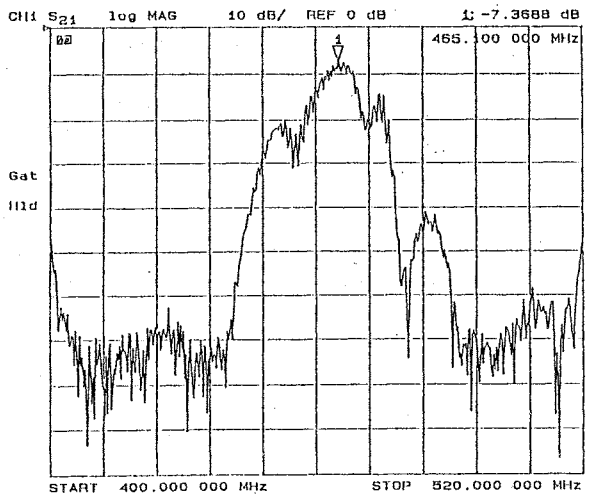
3.2 128° NIOBATE DESIGN

There are two important distinctions between YZ and 128° niobate. Firstly, the piezoelectric coupling is higher, implying that fewer electrodes are needed in the track-coupling transducers and hence giving wider band track coupling. Secondly, the reflection coefficient of aluminium strips is strongly affected by mechanical loading. Mishima [1] has shown that for shorted strips the mechanical loading has the effect of reducing the reflection coefficient, by an amount depending on metal thickness. In fact, the reflection coefficient is zero for a film thickness of 2.6% of the wavelength. For open-circuit strips the electrical reflection coefficient reverses sign, so the mechanical loading is expected to increase the reflection. This is confirmed by Hartmann's measurements [2] on both open and shorted strips. For a 2.15% film thickness the experimental reflection coefficient was 3% (positive imaginary), twice the value for YZ niobate. This implies that, for the same loss, the gratings can have fewer electrodes; a second factor increasing the bandwidth obtainable. In view of the limited information on mechanical loading effects a film thickness of 2.15% (about 1900 Angstroms) was chosen. To our knowledge, this is the first time the enhanced reflectivity of 128 niobate has been exploited in an experimental device.

Other features of the design were similar to the YZ design above. The track-coupling transducers had 45 single electrodes with a reflection coefficient assumed for analysis to be -0.7% though the precise accuracy is uncertain. The aperture was 0.4 mm. The outer transducers had 15 periods with 465 MHz centre frequency and 3 electrodes per period. The gratings had 25 open-circuit strips, and because the strip reflection coefficient should be positive imaginary, were displaced towards the transducers. The aperture was designed such that the outer transducer gave minimum loss (rather than minimum reflection) for a 50 ohm load, requiring a 24 nH parallel inductor on each port.

As in the YZ devices, no inductors were used and as expected a 9 MHz bandwidth was obtained. This is significantly larger than the 5 MHz bandwidth of YZ devices, though the 128° devices showed a greater ripple. The results agree quite well with the theory, that the agreement suggesting that the open-circuit strips in the gratings are operating as intended, with a high reflection coefficient.

Fig 3 128° Niobate result



3.3 DESIGN MODIFICATION

The results contained in Figs 2 & 3 show that the initial design contained high first sidelobes as a result of the responses of the outer transducers. These sidelobes can be substantially reduced by the use of uniform group type SPUDT structures. Further out sidelobes result from the combination of effects from the outer and the track coupling function which can be reduced by incorporating withdrawal weighting in the track coupling structures.

New techniques have been used to optimise the design of withdrawal weighted transducers which have yielded close in sidelobes of greater than 45 dB and devices incorporating these structures are in process.

4. NATURAL SINGLE PHASE UNIDIRECTIONAL TRANSDUCER STRUCTURES (NSPUDT)

The use of NSPUDT structures has not found wide application, partially because there is a lack of knowledge of the fundamental NSPUDT parameters for substrates. The unidirectional behaviour of the NSPUDT is achieved by taking advantage of the distributed internal finger reflections on unconventional asymmetric crystal orientations [3], [4].

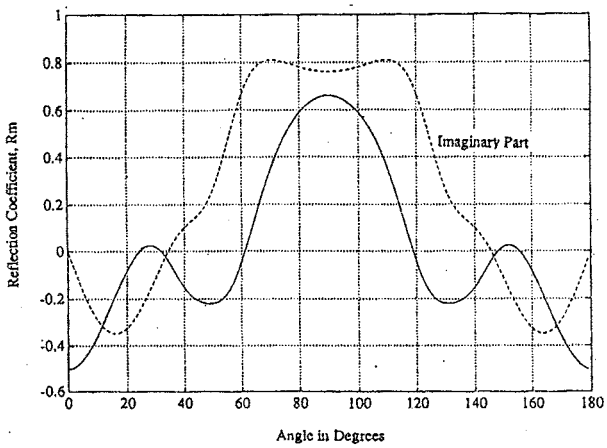
It is necessary to know the magnitude and phase of the finger reflection when designing NSPUDT filters or when searching for new NSPUDT orientations and hence, a theoretical model for the reflections as well as other key model parameters has been developed.

As explained in [3] the requirement is a 45° phase shift between the reflection and transduction centres and the NSPUDT cut is searched for via a parameter R_m. By keeping two Euler angles fixed and plotting the real and imaginary part of the parameter R_m as a function of the third angle, a search can be made.

As an example Fig 4 shows R_m as a function of the propagation angle θ for ST- Quartz. An angle where R_m is almost pure imaginary (i.e., where the real part of R_m is approximately zero) is looked for. A NSPUDT orientation in this case is where $\theta = 25°$ or $155°$. If this condition is being fulfilled, the coupling coefficient, phase velocity and power flow angle at this orientation (a power flow angle close to zero is preferred) are determined. There are other NSPUDT cuts at $\theta = 60°$ and $\theta = 120°$ but these suffer from a low coupling coefficient.[5]

$LiTaO_3$ is an interesting substrate because it has good coupling and moderate temperature stability. Fig 5 - Fig 8 show the various NSPUDT model parameters on YZ-$LiTaO_3$ as a function of the propagation angle θ. Firstly, the first two Euler angles are kept fixed and the third is varied (λ, μ, θ) - ($0°$, $90°$, θ). By investigating R_m, k^2 and the power flow angle, new cuts which

Fig 4 Real and imaginary parts of the mechanical perturbation term R_m as a function of angle θ for ST-Quartz with Euler angles:$(\lambda, \mu, \theta) = (0^o, 132.75^o, \theta)$



have interesting properties for NSPUDT applications can be located. These orientations are given by the Euler angles: $(\lambda, \mu, \theta) = (0^o, 90^o, 141.25^o)$ and $(0^o, 90^o, 218.75^o)$.

Fig 6 demonstrates that the power flow is zero and Fig 8 that the coupling coefficient is high for these cuts. In fact, there is a wide range of angles θ on YZ-LiTaO3 where NSPUDTs can be realized (e.g. $135^o \leq \theta \leq 105^o$) but these orientations do not have a zero power flow angle.

Fig 5 Mechanical perturbation term R_m for YZ-LiTaO3

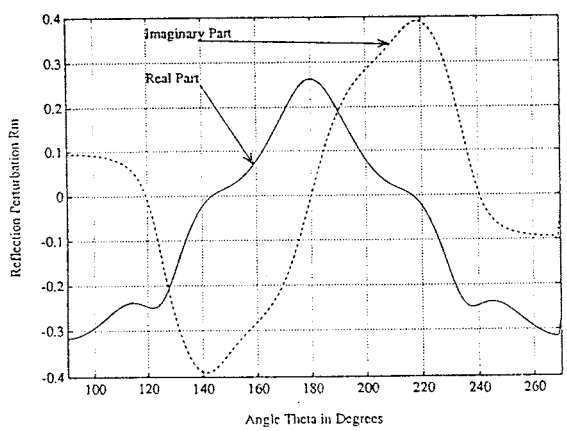


Fig 6 Power flow angle for YZ-LiTaO3

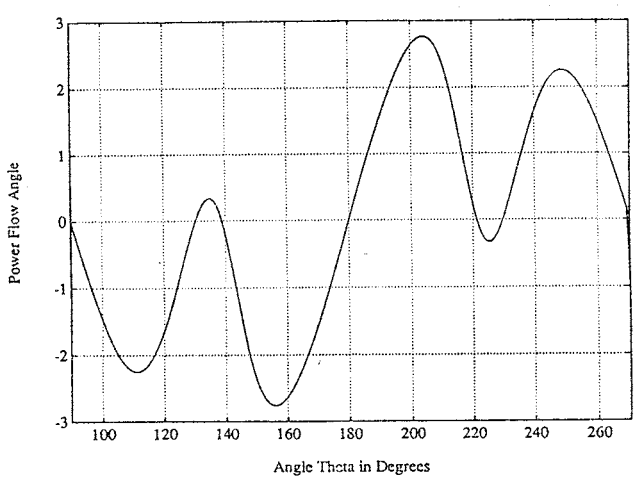


Fig 7 Velocities v_f (free surface) and v_m (metallized surface) for YZ-LiTaO3

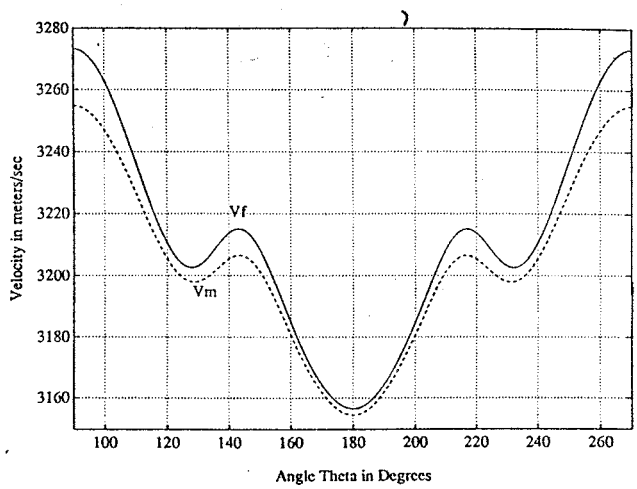
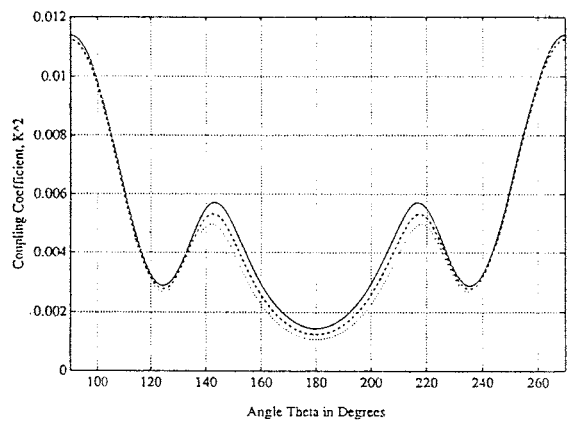


Fig 8 Coupling Coefficient k^2 for YZ-LiTaO3. The different curves are for 3 different ways of calculating k^2 as explained in [7].

(---) : using $2(v_f - v_m)/v_f$

(---) : using Γ_α (...) : using Γ_m



It is worth noting that a considerable difference was found using different sets of elastic stiffness (c_{ijkl}) piezoelectric (e_{ijk}) and dielectric ($\in_{ijf}$) constants of the LiTaO3. The parameters used come from [8] which agree better with experimental results than those given in [9].

If R_m is not pure imaginary, an asymmetry in the transducer response is introduced. Depending on the signs of the mechanical and electrical loading terms, this asymmetry can be compensated by choosing the appropriate metal film thickness to make κ pure imaginary. For YZ+51.25° LiTaO3 this compensation can be achieved because R_m has a positive real part and the electrical term is always negative. To demonstrate the asymmetry discussed above a measured conductance and susceptance of NSPUDT filter on ST-Quartz (X+25°) is shown in Fig 9.

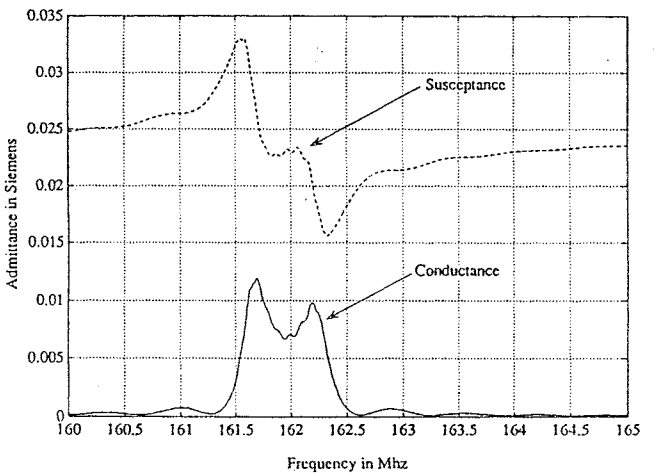
A summary of the calculated NSPUDT perturbation parameters and the wave velocities for the X + 25° propagation on Quartz and the new NSPUDT orientation on YZ-LiTaO3 -Euler angles (0°, 90°, 141.25°) are given below. These were found by applying the theory described above.

Table 1 Calculated Loading Parameters and Velocities

	Shorted Al strips with $\eta = 0.5$	
PARAMETER	ST-Quartz X+25°	$LiTaO_3$ YZ+51.25°
$D\kappa$	-0.7285	-0.7285
D_m †	- 0.2554	- 0.0511
$R\kappa$	- 0.7178	-0.7178
R_m	-0.0086 - *j* 0.2384	-0.0233 - *j* 0-3892
v_f	3247.81 (m/s)	3213.8 (m/s)
v_m	3245.68 (m/s)	3205.4 (m/s)

† In [5] D_m is defined differently (factor η)

Fig 9 Measured Admittance for a device on ST-Quartz (X+25°). The series resistance has been extracted.



The magnitude and phase of SAW transduction, the finger reflection coefficients and other key model parameters were measured using a 5 transducer measurement structure. The measurement setup and the data manipulation software for the parameter extraction is described in detail in [10], [11]. Fig 10 shows the measurement setup, the centre transducer is the NSPUDT, the other IDTs are 20λ long split fingers IDTs. Results of extracted data from measurements are shown in Tables 2 & 3. When κ is measured the combined effect of the electrical and mechanical contributions to the reflection coefficient are measured. In order to separate these contributions the test structures for at least three metal film thicknesses have to be measured. This has not been done however, and therefore the values for κ in the tables are specific to the metal film thickness on the test devices. The calculated values in Table 2 were based on material parameters in [9] but for those in Table 3 the parameters from [8] were used.

Fig.10 Measurement setup (5-IDTs)

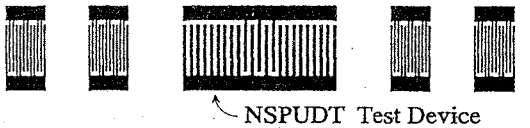


Table 2 Comparison with Measurement

	ST-Quartz (X + 25°)	
PARAMETER †	Calculated	Measured
v_{ac} (m/s)	3242.1	3239.7
Reflectivity $\|\kappa\|$	0.00245	0.00292
Phase of κ	103.17°	95.68°
coupling coeff. κ^2	0.001315	0.001442
C_o pF/m	49.27	50.77

† *(for single finger SPUDT with η = 0.5)*

† *(κ is for a device having parameters given below)*

Table 3 Comparison with Measurement

	$LiTaO_3$ - YZ + 51.25°	
PARAMETER †	Calculated	Measured
v_{ac} (m/s)	3206.7	3199.8
reflectivity [κ]	0.0053	0.0049
Phase of κ	113.91°	101.6°
coupling coeff. κ^2	0.0052	0.0065
Co pF/m	446.6	455

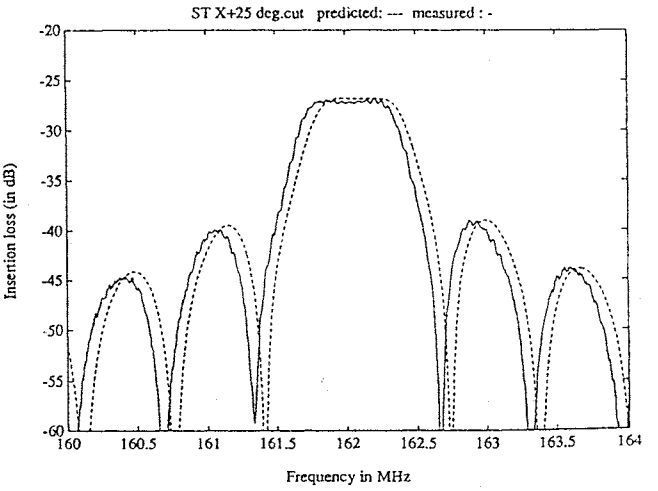
† *(for single finger NSPUDT with η = 0.5)*

† *(device with λ = 16μm and H_m = 2000 Angstroms)*

A 250 wavelength long uniform NSPUDT was built on ST(X+25°)Quartz. The acoustic wavelength was 20μm, the beamwidth was 80 λ and the metal film thickness was 2000Å. For the second IDT a 20 wavelength split finger IDT was used.

In Fig 11 a coupling of modes (COM) model prediction of the insertion loss is compared to measurement. The theoretical parameters from Table 1 were used for the analysis. The shift in frequency observed is attributed to energy storage effects not included in the analysis.

Fig 11 Comparison : Measurement (--) COM theory (...)



The parameters presented in Table 1 were used to design a NSPUDT filter for use on the new YZ+51.25 degree $LiTaO_3$ orientation. Due to the basic principle of NSPUDT operation, only one of the two transducers in the filter can be a NSPUDT. One of the transducers was a withdrawal weighted NSPUDT combined with non-reflecting transduction sections of a 3 group Hopscotch [12] type. This was done in order to obtain more broadband

directivity. The other transducer was a single level metallisation, reflectionless 2 group Hopscotch transducer with Hamming apodization. The Hopscotch transducer consists of groups of electrodes which are offset a quarter wavelength. This places the transduction weighting in quadrature and simultaneously cancels finger reflections. Hopscotch transducers exhibit sub-harmonic as well as higher harmonic responses. In the broadband measured response shown in Fig 13 some of these harmonics can clearly be observed. Fig 12 shows measured insertion loss of the unmatched filter.

Fig 12 Measured insertion loss of the unmatched filter

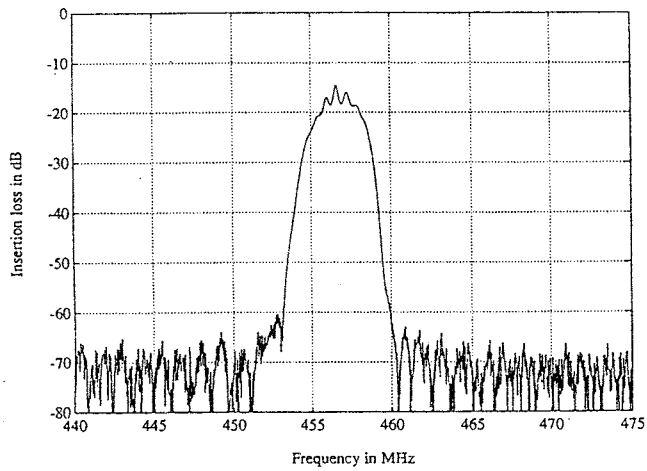
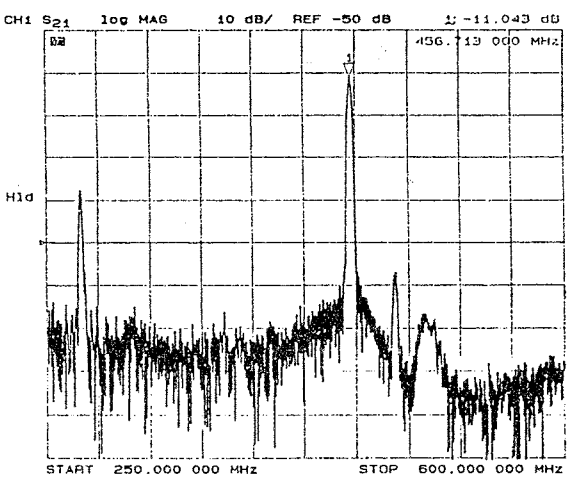


Fig 13 Measured broadband frequency response



As discussed in [13] single phase transducers should be matched at the acoustic port since a matched electric port will usually cause severe triple transit distortions. Fig 14 shows a measured frequency response after matching both transducers at the electrical ports (matching for minimum loss). A minimum measured insertion loss of 7.1 dB was achieved. Fig 15 shows the response after matching the NSPUDT transducer at the acoustic port (matching for minimum ripple) but the other IDT at the electrical port. In Fig 16 the minimum insertion loss is 8.5 dB. It should be kept in mind that one of the transducers has a 3 dB bidirectional loss. The 3 dB bandwidth is 2.5 MHz and the measured triple transit suppression was found to be 34 dB. Fig 16 shows the predicted response of the matched filter.

Fig 14 Measured frequency response after matching for minimum loss (both transducers matched at the electrical port)

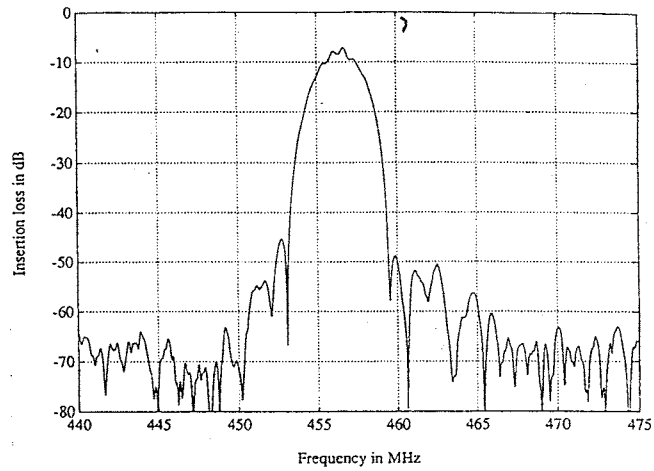


Fig 15 Measured frequency response after matching for minimum ripple (NSPUDT matched at acoustic port but the Hopscotch at electric port)

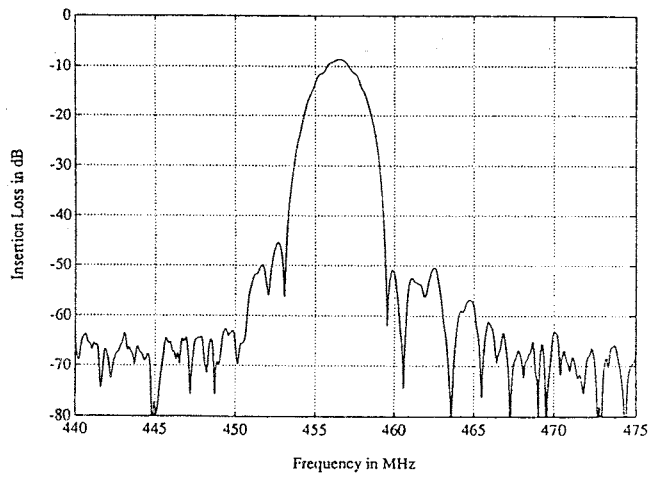
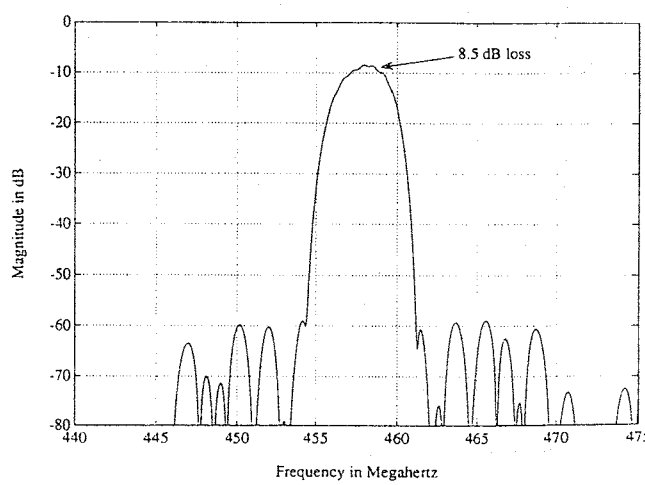


Fig 16 Predicted frequency response of the matched filter



5. CONCLUSIONS

Low loss filters at 500 MHz have been demonstrated on two cuts of $LiNbO_3$ and on a new orientation of $LiTaO_3$. Insertion losses of less than 8 dB with far out sidelobe levels of 65 dB have been obtained for units with more than

5 MHz bandwidth.
Theoretical analyses using both the COM technique and a distributed delta function technique compare very well with the measured performances.

6. ACKNOWLEDGEMENTS
The authors would like to express their thanks to Dr.H. Choffat for his process support, to H. Madorin for the assembly of units and to Dr. F. Nyffeler, for many helpful discussions. The work was performed at ETH Zurich and Ascom Microsystems and was funded by the Swiss PTT and Ascom.

7. REFERENCES

1. N. Mishima, M. Takase, S. Mitobe & Y. Ebata
"IIDT Type SAW Filter Using Acoustic Reflection Cancel Condition with Solid IDT"
IEEE 1989 Ultrasonics Symposium Proceedings, pp.31-35
2. S. Jen & C.S. Hartmann
"An Improved Model for Chirp Slanted SAW Devices"
IEEE 1989 Ultrasonics Symposium Proceedings, pp.1-13
3. P. V. Wright
"The Natural Single-Phase Unidirectional Transducer: A New Low-loss SAW Transducer"
IEEE 1985 Ultrasonics Symposium Proceedings, pp.58-63
4. P.V. Wright
"Low-Cost High Performance Resonator and Coupled-Resonator Design: NSPUDT and other Innovative Structures"
43rd Annual Frequency Control Symposium, May 1989
5. T. Thorvaldsson
"Analysis of the Natural Single Phase Unidirectional SAW Transducer"
IEEE 1989 Ultrasonics Symposium Proceedings, pp. 91-96
6. D.P. Morgan
" Surface-Wave Devices for Signal Processing"
Elsevier, Amsterdam, 1985
7. B.P. Abbott & D.C. Malocha
"Closed Form Solutions for Multistrip Coupler Operation Including the Effects of Electrode Resistivity"
IEEE 1990 Ultrasonics Symposium Proceedings
8. R.T. Smith & F.S. Welch
Journal of Applied Physics Vol.42
pp 2219 - 2230, 1971
9. A.J. Slobodnik, E.D. Conway & R.T. Delmonico
"Microwave Acoustics Handbook"
Air Force Cambridge Research Laboratories, AFCRL-TR-73-0597, Oct'73
10. C.S. Hartmann & B.P. Abbott
"Experimentally Determining the Transduction Magnitude and Phase and The Reflection Magnitude and Phase of SAW SPUDT Structures"
IEEE 1990 Ultrasonics Symposium Proceedings
11. C.S. Hartmann & R.T. Hartmann
"Software for Multiport RF Network Analysis with a Large Number of Frequency Samples and Applications to 5 Port SAW Device Measurements"
IEEE 1990 Ultrasonics Symposium Proceedings
12. J.C. Andle & M.B. King
"A Novel Group-Type SPUDT"
IEEE 1987 Ultrasonics Symposium Proceedings, pp 189-192
13. B.P. Abbott, C.S. Hartmann & D.C. Malocha
" Matching of Single Phase Unidirectional SAW Transducers"
IEEE 1990 Ultrasonics Symposium Proceedings

Session 3A Number 12

SAW SENSITIVITY TO STRESSES AND TEMPERATURE GRADIENTS NEW THEORETICAL AND EXPERIMENTAL RESULTS

S. Ballandras and E. Bigler

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

Abstract

Since anisotropic model for radial in-plane stresses have been implemented successfully, we try to increase the precision of our previsions by using finite element methods. These new calculations are compared to previous analytical ones and to experimental results to explain discrepancy between predictions and measurements. Conclusion can be made upon stress sensitivity of quartz devices. Experiments have been made on cuts insensitive to both static temperature and compression effects, to evaluate sensitivity of these cuts to dynamic temperature effects. Results are shown for cuts close to B.K. Sinha's STC cut and to TG ($\phi=0$, $\theta=-10$, $\psi=30$) cut. Interpretation of experimental results is proposed and may yield to new theoretical research.

1. Introduction

Since piezoelectric crystals are used to build high stability oscillators using Bulk (BAW) or Surface (SAW) acoustic waves, a great effort has been made to minimize the influence of external parameters on the resonnant frequency of the device. Different studies on BAW resonators have shown that such a problem may be solved by using two different and complementary methods :
- insulating the device from external influences by optimizing fixation and packaging conditions of the crystal.
- using physical properties of the crystal (anisotropy) adapted to a particular shape of the substrate where the waves propagate.

The second approach yields to the discovery of the SC-cut introduced by E.P.Eernisse [1] in the case of BAW propagating in a thin circular quartz plate held in two diametricaly opposed points. This cut is compensated for both temperature and planar isotropic stresses effects.

This method, applied on SAW devices in quartz, permits to point out the existence of two singly rotated quartz cuts exhibiting low sensitivity to 1st order temperature effects and to planar stresses effects [2, 3, 4], in the case of a symmetrical distribution of stresses inside the crystal. This situation occurs, for example, in a circular quartz plate with suitable symmetrical fixation conditions. Experiments have been implemented and have confirmed compensation effects predicted theoretically for the cut ($\Phi=0$; $\theta=41.8$; $\psi=48.2$), close to B.K. Sinha's STC-cut [5], and for TG-cut ($\Phi=0$; $\theta=-10$; $\psi=30$).

In the present work, the case of planar symmetrical stress compensation is considered, using finite element analysis in order to increase the precision obtained on predicted frequency shifts.

Experimental results on the dynamical thermal behavior of SAW delay lines built on STC-cut and TG-cut substrates are presented. It is shown that both cuts exhibit a very low sensitivity to dynamical thermal effects, with an improvement of one order of magnitude over the classical ST,X cut tested under the same conditions.

2. Saw quartz cuts submitted to diametrical compression

a) Experimental measurements

SAW propagating substrate is a quartz disk with dimensions of 22 mm in diameter and 1 or 2 mm in thickness. Two interdigital transducers are deposited on one side, having respectively 100 and 150 finger pairs. Transducer period is 34.4 microns. Like this, a quartz delay line is obtained with time delays in the range of 1.1 to 1.6 microseconds, according to Rayleigh wave velocity. This device is used in an oscillator operating at a frequency close to 100 MHz. The force-frequency effect is measured for each azimuthal angle β between the SAW propagation direction and the direction of the force. (Fig.1)

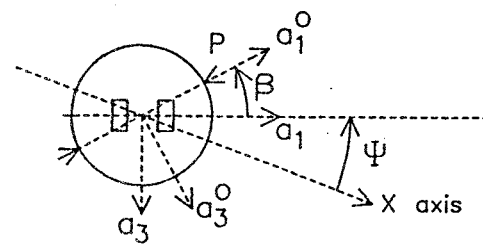


Fig. 1 : Definition of azimuthal angle β

b) Comparison between experimental data and theoretical results obtained by finite element analysis

In the case of an anisotropic disk, the force-frequency effect can be calculated using the model of stress-sensitivity coefficients as defined in [6] ; an exact calculation of stresses at the center of the plate can be obtained following Janiaud's calculation [7,8].

To increase the precision of theoretical predictions, a finite element 2-D model has been applied to a circular quartz disk submitted to a single diametrical compression. The model takes fully into account anisotropy of the crystal and yields a distribution of stresses, strains and displacement gradients, tabulated for each element of the model representing the plate. A perturbation method [9] is used to compute frequency shifts induced on the delay line oscillator. Starting from the integral expression given by

$$\frac{\Delta F}{F_o} = \frac{\iiint_v \frac{\partial u_i^{o*}}{\partial x_k} \overline{H_{ikjm}} \frac{\partial u_j^o}{\partial x_m} dv}{2\rho_o \omega_o^2 \iiint_v u_\ell^{o*} u_\ell^o dv} \qquad (1)$$

where

$$\overline{H_{ikjm}} = \delta_{ij} \overline{T_{km}} + C_{pkjm} \frac{\overline{\partial u_i}}{\partial a_p} + C_{ikqm} \frac{\overline{\partial u_j}}{\partial a_q} + C_{ikjmuv} \overline{S_{uv}} \qquad (2)$$

the digitized expression of the perturbation method can be derived, in the particular case of elements with equal surfaces as

$$\frac{\Delta F}{F_o} = \frac{\sum_{r,s}^{4} A_r^* A_s \, {}^{o}u_i^{o*(r)} \, {}^{o}u_j^{o(s)} \, n_k^{*(r)} \, n_m^{(s)} \sum_e \left(\frac{\overline{H_{ikjm}}(e)}{n_2^{*(r)} - n_2^{(s)}} \right)}{2\rho_o V_o^2 \sum_{r,s}^{4} A_r^* A_s \, {}^{o}u_\ell^{o*(r)} \, {}^{o}u_\ell^{o(s)} \sum_e \left(\frac{1}{n_2^{*(r)} - n_2^{(s)}} \right)} \quad (3)$$

It should be noted here that instead of using the complete expansion of eq. (3), the approximation of stress-sensitivity coefficients can be combined with finite element analysis to compute frequency shifts, and yields

$$\frac{\Delta F}{F_o} = \frac{\sum_e {}^{s}a_{ij} \cdot \overline{T_{ij}}(e)}{N} \quad (4)$$

where e represents the current element and N the total number of elements in the model.
Calculations have been performed on the model represented on Fig. 2. Values for stresses, strains and displacement gradients have been taken from the eight triangles forming a regular octagon at the center of the plate. This method has been applied to diametrical compression of a (ST,X) subtrate, and results are shown on Fig. 3, where the predicted theoretical values are superimposed on the curves given by the approximation of isotropic stress calculation.

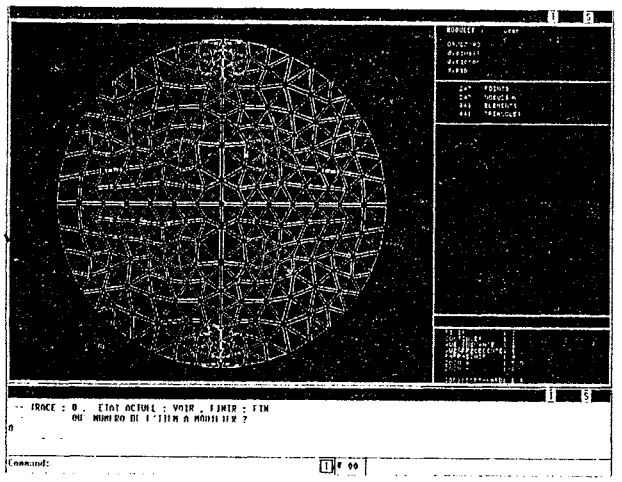


Fig. 2 : Two dimension finite element model

Fig. 3.a is obtained using eq. (3) and Fig. 3b corresponds to the simplified form of eq. (4). Agreement between theoretical models and experimental results is satisfactory on Fig. 3.b, but calculation using equation (3) presents certain discrepancies in terms of the magnitude of $\Delta F/F_0$ for particular angles β . This could be explained by the choice of interpolation polynomials (Lagrange type of the 1st order) used for the finite element model [10], introducing significant errors. The exact shape of SAW energy flow should also be taken into account in this problem where actual dimensions of the plate are not very large compare to the SAW propagating volume. This will be the next step introduced in our finite element model.

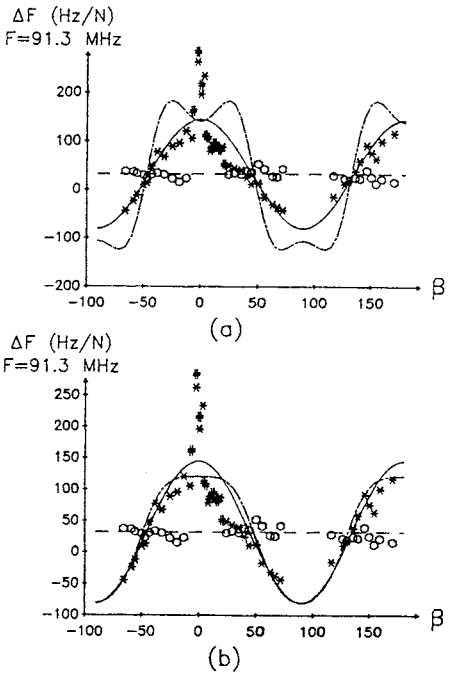


Fig. 3 : Frequency shifts due to single diametrical compression calculated using finite element analysis
a) with eq. (3) b) with eq. (4)

3. SAW quartz cuts submitted to dynamical thermal perturbations

a) Perturbation method for dynamical temperature effects

According to previous models introduced to predict dynamical temperature effects [11, 12], variations of temperature Θ inside a circular substrate submitted to external temperature changes applied through a couple of diametrical points, can be written as

$$\theta(a_\ell, t) = f_1(a_\ell)\,\Phi(t) + f_2(a_\ell)\,\frac{\partial \Phi(t)}{\partial t} \quad (5)$$

where a_ℓ represent material coordinates ($\ell = 1,3$) in the plane of the subtrate, and $\Phi(t)$ is the thermal excitation function (heat flux) applied simultaneously on both diametrical fixation points.

In this case of thermoelastic effects, the perturbation tensor $\overline{H_{ikjm}}$ will be given by

$$\overline{H_{ikjm}} = \delta_{ij}\,\overline{T_{km}}(a_\ell, t) + C_{pkjm}\,\frac{\partial \overline{u_i}(a_\ell, t)}{\partial a_p} + \quad (6)$$

$$C_{ikqm}\,\frac{\partial \overline{u_j}(a_\ell, t)}{\partial a_q} + C_{ikjmuv}\,\overline{S_{uv}}(a_\ell, t) + \frac{dC_{ikjm}}{dT}(\theta(a_\ell, t) - \theta_o)$$

Following a derivation given in [13], resulting frequency shifts will take the following form

$$\frac{\Delta F}{F_o} = \bar{a}\,\Phi(t) + \bar{a}\,\frac{\partial \Phi(t)}{\partial t} \quad (7)$$

where $\bar{a}$ and $\hat{a}$ are two coefficients defining static and dynamic temperature sensitivities ; these coefficients depend on thermoelastic properties of the crystal. Since in fact no experimental access is possible to estimate the heat flux excitation function $\Phi(t)$, effective coefficients $\bar{a}_e$ and $\hat{a}_e$ will be defined with respect to temperature $\Theta(M_0,t)$ measured at a particular point M_0

$$\frac{\Delta F}{F_0}(M_0,t) = \bar{a}_e \theta(M_0,t) + \hat{a}_e \frac{\partial \theta(M_0,t)}{\partial t} \qquad (8)$$

If around some temperature Θ (measured at point M_0) coefficient $\bar{a}_e$ vanishes, then coefficient $\hat{a}_e$ will be representative of dynamical thermal effects. If the excitation function is sinusoidal, dynamical temperature effects (in the limit of small amplitudes of thermal perturbations) will also be sinusoidal ; dynamical frequency vs temperature curves will exhibit an elliptic shape due to a phase shift of $\pi/2$ between excitation and detection.

b) Experimental device

An experimental procedure has been established to test the behaviour of STC and TG cuts submitted to dynamical temperature variations, with the assumption that thermal perturbations propagate mainly through fixation points of the quartz plate. SAW substrate is contained inside a brass enclosure. The circular quartz plate is held in two diametrical points by a pair of copper ribbon holders acting both as supports and thermal link to a copper base plate. Under this plate, a NPN-type power transistor is attached, acting as the heating source. (Fig. 4).

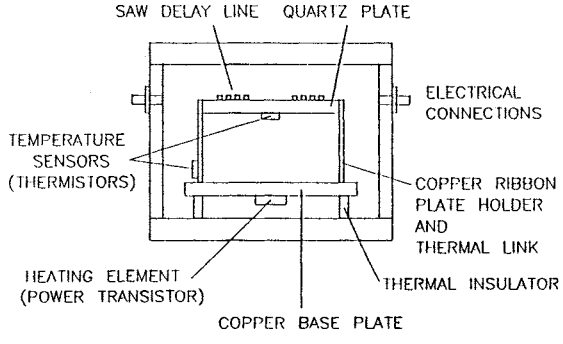


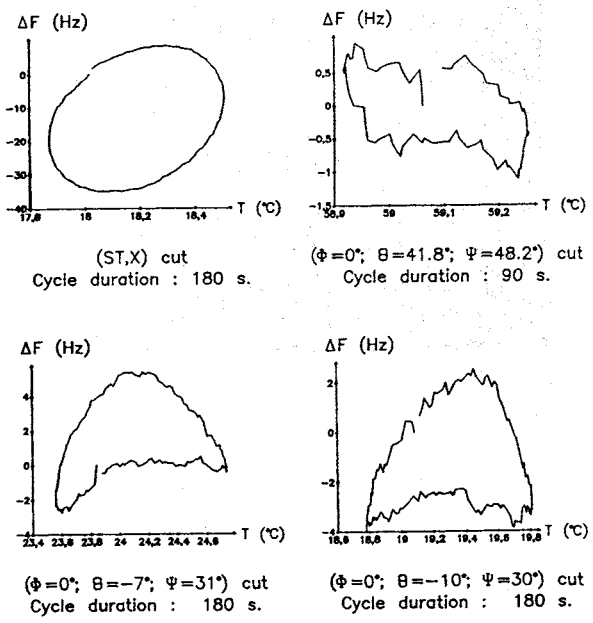
Fig. 4 : Experimental device for testing dynamical thermal behaviour of SAW quartz cuts

Two resistive sensors allow measurements of temperature variations under the quartz plate and on one of the holders.

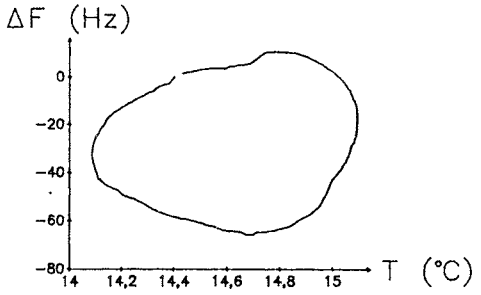
The device shown on Fig. 4 and the loop amplifier are at an initial temperature of -30°C. Thermal cycles around ambiant temperature are obtained by applying a sinusoidal voltage to the heating element. The whole experiment is controlled by a computer that commands excitation of the heating transistor and stores temperature and frequency variations. For each cut, a representative number of cycles is recorded and can be averaged to improve signal-to-noise ratio.

c) Experimental results

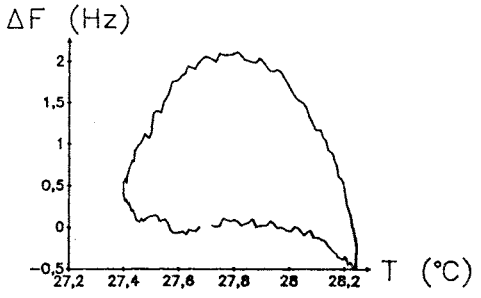
Experiments have been performed for (ST,X), STC and TG cuts. Another cut, ($\Phi=0$; $\theta=-7$; $\psi=31$), close to TG has also been tested to point out the fact that compensation effects can be obtained with a good tolerance on cut angles ; this property has already be shown in the case of diametrical compression [6]. Results appear on Fig. 5 for the different cuts. When compared with (ST,X), sensitivity of STC and TG cuts appears to be very low. It should be noted that if static compensation takes place around ambiant temperature for (ST,X) and TG cuts, STC turnover point has been found in the vicinity of 60°C in our experimental device. A parabolic shape, a residue of quasi-static temperature effects, appears on experimental frequency-temperature curves for TG cut. From this result, it can be inferred that the sensitivity of TG cut to dynamical temperature effects is actually lower than measured. To increase precision of measurements and avoid some parasitic effects, a thermal insulating foam plate has been inserted between quartz and copper (to reduce thermal radiation and convection effects). In the case of TG and (ST,X) cuts, results show that in this new configuration, designed to bring the experiment closer to theory, the sensitivity of TG-cut has been reduced by a factor of 2, while (ST,X) behaviour does not change significantly. Values for $\hat{a}_e$ are reported in table I.



with experimental device of Fig. 4



(ST,X) cut
Cycle duration : 180 s.



(Φ=0°; Θ=−10°; Ψ=30°) cut
Cycle duration : 180 s.

with an insulating foam between quartz and copper plates

Fig 5 : Experimental results of SAW quartz cuts submitted to dynamical temperature variations

Cuts / a_c values in s.K^{-1}	(ST,X) Φ=0° Θ=42.75° Ψ=0°	STC Φ=0° Θ=41.8° Ψ=48.2°	Close to TG Φ=0° Θ=−7° Ψ=31°	TG Φ=0° Θ=−10° Ψ=30°
without insulating foam	$20.92*10^{-6}$	$1.71*10^{-6}$	$2.05*10^{-6}$	$1.75*10^{-6}$
with insulating foam	$20.30*10^{-6}$	—	—	$0.86*10^{-6}$

Table 1

Conclusion

It has been shown that TG and STC cuts are less sensitive to diametrical compression than the classical (ST,X) cut. All these cuts are also 1st order temperature-compensated. An efficient finite element method has been introduced to handle anisotropic model of compression effects in circular SAW devices. Experimental evidence has been found that TG and STC cuts are much less sensitive to dynamical thermal effects than (ST,X) cuts. Greater precison could be obtained if SAW resonators were used instead of delay lines, and further investigations will be required to check if a resonator design is possible (they have non-zero power flow angle) on the new stress- and temperature-compensated SAW quartz cuts presented here.

Acknowledgements

- This work is supported by Rome Air Development Center, Hanscom Field, Air Force Base, Bedford, MA 01731 (USA) under AFOSR grant # 87-0241.
- The authors are grateful to Mrs Coquerel, Bonjour and Magnin for the realisation of experimental devices.and to Mr Crolet who made possible calculations on the finite element program MODULEF.

References

[1] E.P. Eernisse, "Calculations on the stress-compensated (SC-cut) quartz resonator", Proc. 30th Ann. Freq. Cont. Symp., 1976, pp. 8-11.

[2] G. Théobald, E. Bigler, J.J. Gagnepain, "Stress-compensated cuts for surface acoustic waves", Proc. 1986 Ultrasonics Symp., IEEE cat. n° 86CH2375-4, pp. 209-212.

[3] E. Bigler, G. Théobald and D. Hauden, "SAW quartz cuts with low stress and temperature sensitivity", Electron. Lett., vol. 23, n° 10, pp. 514-516, (may 1987).

[4] E. Bigler, R. Coquerel, D. Hauden, "Experimental temperature and stress sensitivities of surface acoustic wave quartz cuts", Proc. 42nd Ann. Freq. Cont. Symp., IEEE cat. n° 88 CH 2588-2, pp. 224-229.

[5] B.K. Sinha, "A stress and temperature compensated orientation and propagation direction for surface acoustic wave devices", IEEE Trans. Ultrason. Ferro. Freq. Cont., vol. UFFC-34,n° 1, pp. 64-74 (jan. 1987).

[6] S. Ballandras, "Sensibilité des oscillateurs à quartz à ondes de surface aux contraintes et gradients thermiques", Thèse de l'Université de Franche-Comté, 22 mars 1991, Besançon.

[7] D. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thèse, Université de Besançon, 1978, pp. 63-67.

[8] E.Bigler, S. Ballandras, D. Hauden, "Experimental study of planar stresses effects on surface acoustic wave devices", 3rd European Frequency and Time Forum, Besançon, France (march 1989).

[9] H.F. Tiersten and J.C. Baumhauer, "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., vol. 54, n° 4, pp. 1017-1074 (oct. 1973).

[10] O.C. Zienkiewicz, "La méthode des éléments finis", Mac Graw Hill, 1979.

[11] D. Hauden, G. Théobald, "Dynamic thermal sensitivity of SAW quartz oscillators", Ultrasonics Symposium, Boston, 1980.

[12] J.P. Valentin, G. Théobald, J.J. Gagnepain, "Frequency shifts arising from in-plane temperature gradient distribution in quartz resonators", Proc. 38th Ann. Freq. Cont. Symp., Philadelphia, 1984.

[13] G. Théobald, D. Hauden, "Theoretical analysis of dynamic thermal effects in SAW devices", Proc. 38th Ann. Freq. Cont. Symp., Philadelphia, 1984.

Session 3A Number 13

Propagation of Piezoelectric Waves in Singly- and Doubly- Periodic Metallic Gratings

A.R. Baghai-Wadji and Ch. Mecklenbräuker

Technische Universität Wien, Institut für Allgem. Elekt. und Elek.
Gußhausstr. 27-29, A-1040, Wien, Austria

Abstract: A rigorous theory for wave propagation of bounded waves and guided modes in singly- and doubly-periodic gratings in piezoelectric environment is presented. Highly conductive, and negligibly massive, metallic electrodes are assumed to be located on the surface of a piezoelectric substrate or on the interface of two substrates (in the latter case, at least one of the substrates must be piezoelectric). Half-space configurations and substrates with finite thickness are considered.
Starting with the governing piezoelectric equations and the constitutive relations, we derive the full wave Green's functions of the corresponding periodic problem. The proposed solution procedure relies on the Floquet's theorem in 1D- and 2D-representation and on the concept of dual series.

I. INTRODUCTION

Around the end of 19th century, the propagation behaviour of waves in periodic structures, as models for crystal lattices, attracted many scientists. As early as 1881 Cauchy, et. al [1] discussed the properties of simplest periodic lattices consisting of identical particles. A second relevant impulse to the research of periodic structures has its origin in signal-processing applications. From an engineering point of view, periodic structures seemed to offer a useful and simple technical tool to produce signal-shaping devices. At the beginning of the twentieth century periodic networks were used to develop electric filters [2], and in the thirties and forties optical multilayers were proposed to fabricate polarizers, beam splitters, and filters [3]. In the post-war period the microwave tubes dictated the development and research of periodic structures. The periodic structures were used to slow the microwaves propagating in tubes and, consequently, to increase with relatively slow electron beams, their coupling efficiency.

In 1967 White and Voltmer [4] invented the interdigital transducer to excite electrically surface acoustic waves (SAW) in piezoelectric substrates. Meanwhile, in the frequency range from a few MHz to some GHz, SAW devices have become established for signal-processing and signal-shaping purposses. From a modelling and simulation point of view these devices benefited from the theory of periodic structures. Many concepts developed from this theory at the end of the 60's and early 70's are still of importance in SAW technology. To complete the picture can be mentioned another relevant application of periodic structures. Structures, constructed by one-, two-, and three-dimensional periodic replica of cavities in elastic media, are used as models for the investigation of porosity of materials.

The theories of differential equations with periodic coefficients, of the wave propagation in media with periodic boundary conditions [5], and of the wave propagation in media with periodic material constitutions [6] rely on the pioneering work of Floquet (see for example [7]) and Bloch's generalization [8] of these results.

In many micro acoustic devices the SAW structures can be modelled as a set of periodic, infinitely long parallel electrodes mounted on a piezoelectric substrate, preferably Quarz, $LiNbO_3$ or $LiTaO_3$. In this model we obviously lose any information regarding the longitudinal- and transversal- end effects which reflect, respectively, the finiteness of the electrode number and the finiteness of the aperture of the interdigital transducers. Nevertheless, the field analysis results, even under physically limiting conditions, are valuable for design of SAW devices. On the other hand, for many other applications; i.e., transducers with rather a small number of electrodes and/or transducers with small apertures, can not be satisfactorily modeled as periodic structures. In a series of publications we discussed the phenomena in connection with longitudinal and transversal finiteness of transducers ([9] and its references). However, three-dimensional calculations are generally very time-consuming, and the required computer resources are considerable. Not to forget that the field distributions are frequency-dependent, and thus must be calculated independently and individually for every frequency value. On the contrary, the field solutions of periodic structures, once formulated, are incomparably more efficient, both with respect to time as well as to the computer resources.

In his important early work, Engan treated a periodic sequence of metallic electrodes on an anisotropic dielectric half-space [10]. Formulating by dual series, the associated boundary conditions he impressively found analytical solutions for the coefficients appearing in the Fourier expansions of the electric potential distribution in the electrostatic approximation.

In a recent paper, Xue, and Shui [5] developed an elegant procedure for solution of dual series arising in SAW devices, including the piezoelectricity of the substrate. Here, in contrast to Engan's work, the determination of the coefficients in the Fourier expansions are to be calculated numerically. Nevertheless, despite the complexity of the problem they studied, and at the same time the rigorosity of their method, the paper of Xue and Shui can be regarded as one of the important works in the SAW literature of the 1980's.

Our purpose in this paper is manifold extensions of the work of Xue and Shui to embrace a larger class of problems: (i) We show that [5] can be extended so as to include three dimensional problems; i.e., doubly periodic problems and singly periodic structures with finite aperture. (ii) It is shown that the package influence on the field distribution, simulated as grounded parallel plates above and below the SAW structures, can be very easily accounted for. (iii) The structure in the fundamental period can have a rather general topology, and electrically can be arbitrary. Even electrically disconnected floating electrodes can be analysed as simply as first proposed for transducers which consist of finite number of electrodes [11]. (iv) In rather an obvious way we here, develop and present one- and two-dimensional dynamic (frequency dependent) periodic dyadic Green's functions. These functions can be regarded as extensions of our previously discussed Green's functions in [12]-[15]. (v) A theoretical discussion will here be presented for the evaluation of dynamic element factors. Until now, for half space dielectrics, only a single element factor in electrostatic approximation is known [16]. And finally, from a numerical point of view, we wish to emphasize one important aspect: Similar to the nonperiodic case [17], we convert the underlying homogeneous system of basic partial differential equations into numerically a much more stable equivalent eigenvalue system.

This paper is organized as follows: In Section II the statement of the problem is followed by a theoretical part introducing two matrix differential operators. In Section III we introduce and discuss the dyadic periodic Green's functions and the concept of dynamic element factors. Finally in Section IV a glimpse of the numerical results is presented.

II. THEORY

II.a Statement of Problem: Consider a piezoelectric plate bounded by an interface- and by a back- plane located at $x_3 = 0$ and $x_3 = -d_l$. The backplane may or may not be metallized and grounded. The adjacent medium above the interface is either free space, isotropic or anisotropic elastic medium, or an anisotropic piezoelectric. At a distance d_u above the interface a grounded metallic plate may or may not exist. On the interface let us consider the following electrode geometries. (i) Singly periodic structure (p-periodic in x_1-direction) consisting of infinitely long, massless, negligibly thin x_2-parallel electrodes, Fig.1.

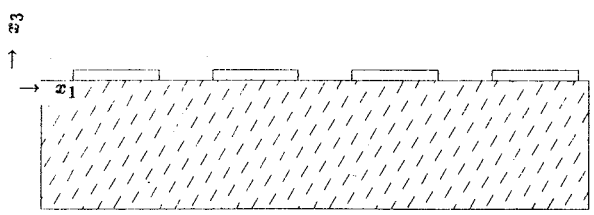


Fig.1 A singly periodic structure

(ii) The same structure as in the first case, but with electrodes possessing arbitrary cross sections. (iii) The same conditions as in (i) or in (ii), except with finite aperture length. (iv) Doubly periodic structure (p-periodic in x_1-direction and q-periodic in x_2-direction), Fig.2.

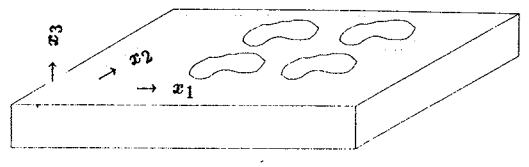


Fig.2 A doubly periodic structure

The geometry in the periodic cells in the $(x_1, x_2)-$plane may be arbitrary. Moreover, we can include in the analysis metallic, electrodes with finite thickness.

Our purpose is to establish a rather general theory which is well adopted to the solution of the above mentioned boundary value problems. To be as complete as possible, we develope the theory and outline the solution procedure in three dimensions.

II.b Operators $\underline{\underline{\mathcal{L}}}$ and $\underline{\underline{\tau}}$ Let us start with the basic system of partial differential equations in Auld's compact notation

$$\underline{\underline{\nabla}}^T \underline{\underline{c}}\, \underline{\underline{\nabla}}\, \vec{u} - \rho \frac{\partial^2 \vec{u}}{\partial t^2} = -\underline{\underline{\nabla}}^T \underline{\underline{e}}\, \vec{\nabla}\, \varphi, \tag{1a}$$

$$\vec{\nabla}^T \underline{\underline{e}}\, \vec{\nabla}\, \varphi = \vec{\nabla}^T \underline{\underline{e}}^T \underline{\underline{\nabla}}\, \vec{u}. \tag{1b}$$

Here, the superscript T signifies transposition and the differential operators $\underline{\underline{\nabla}}$ and $\vec{\nabla}$ are

$$\underline{\underline{\nabla}}^T = \begin{pmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{pmatrix}, \tag{2}$$

$$\vec{\nabla}^T = (\partial_1, \partial_2, \partial_3). \tag{3}$$

Further, $\underline{\underline{c}}$, $\underline{\underline{e}}$ and $\underline{\underline{\varepsilon}}$ are the elastic, piezoelectric, and permittivity tensors of the solid, respectively. $\vec{u}^T = (u_1, u_2, u_3)$ and φ denote the mechanical displacement and the electrical potential. In the following $\vec{U}^T = (u_1, u_2, u_3, \varphi)$. For further development it is instructive to define the following two differential-operators:

$$\begin{aligned} \underline{\underline{\mathcal{L}}} =& \left[\underline{\underline{M}}^{(a)} - \{\underline{\underline{M}}^{(ac)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \{\underline{\underline{M}}^{(ac)}\}\right] \frac{\partial^2}{\partial x_1^2} + \\ &+ \left[\underline{\underline{M}}^{(b)} - \{\underline{\underline{M}}^{(bc)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \{\underline{\underline{M}}^{(bc)}\}\right] \frac{\partial^2}{\partial x_2^2} + \\ &+ \Big[\underline{\underline{M}}^{(ab)} + \{\underline{\underline{M}}^{(ab)}\}^T - \\ &\quad - \{\underline{\underline{M}}^{(ac)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \{\underline{\underline{M}}^{(bc)}\} - \\ &\quad - \{\underline{\underline{M}}^{(bc)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \{\underline{\underline{M}}^{(ac)}\}\Big] \frac{\partial^2}{\partial x_1 \partial x_2} + \\ &+ \{\underline{\underline{M}}^{(ac)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \frac{\partial}{\partial x_1} \underline{\underline{\tau}} + \\ &+ \{\underline{\underline{M}}^{(bc)}\}^T \{\underline{\underline{M}}^{(c)}\}^{-1} \frac{\partial}{\partial x_2} \underline{\underline{\tau}} + \\ &+ \frac{\partial}{\partial x_3} \underline{\underline{\tau}} - \\ &- \underline{\underline{\rho}}\, \frac{\partial^2}{\partial t^2}, \end{aligned} \tag{4a}$$

$$\underline{\underline{\tau}} = \underline{\underline{M}}^{(ac)} \frac{\partial}{\partial x_1} + \underline{\underline{M}}^{(bc)} \frac{\partial}{\partial x_2} + \underline{\underline{M}}^{(c)} \frac{\partial}{\partial x_3}. \tag{4b}$$

The material matrices $\underline{\underline{\rho}}$, $\underline{\underline{M}}^{(a)}$, $\underline{\underline{M}}^{(b)}$, $\underline{\underline{M}}^{(c)}$, $\underline{\underline{M}}^{(ab)}$, $\underline{\underline{M}}^{(bc)}$, $\underline{\underline{M}}^{(ac)}$ are defined as follows:

$$\underline{\underline{\rho}} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\underline{\underline{M}}^{(a)} = \begin{pmatrix} c_{11} & c_{16} & c_{15} & e_{11} \\ c_{61} & c_{66} & c_{65} & e_{61} \\ c_{51} & c_{56} & c_{55} & e_{51} \\ e_{11} & e_{16} & e_{15} & -\varepsilon_{11} \end{pmatrix},$$

$$\underline{\underline{M}}^{(b)} = \begin{pmatrix} c_{66} & c_{62} & c_{64} & e_{62} \\ c_{26} & c_{22} & c_{24} & e_{22} \\ c_{46} & c_{42} & c_{44} & e_{42} \\ e_{26} & e_{22} & e_{24} & -\varepsilon_{22} \end{pmatrix},$$

$$\underline{\underline{M}}^{(c)} = \begin{pmatrix} c_{55} & c_{54} & c_{53} & e_{53} \\ c_{45} & c_{44} & c_{43} & e_{43} \\ c_{35} & c_{34} & c_{33} & e_{33} \\ e_{35} & e_{34} & e_{33} & -\varepsilon_{33} \end{pmatrix},$$

$$\underline{\underline{M}}^{(ab)} = \begin{pmatrix} c_{16} & c_{12} & c_{14} & e_{12} \\ c_{66} & c_{26} & c_{64} & e_{62} \\ c_{56} & c_{52} & c_{45} & e_{52} \\ e_{16} & e_{12} & e_{14} & -\varepsilon_{12} \end{pmatrix},$$

$$\underline{\underline{M}}^{(ac)} = \begin{pmatrix} c_{15} & c_{56} & c_{55} & e_{51} \\ c_{41} & c_{46} & c_{45} & e_{41} \\ c_{31} & c_{36} & c_{35} & e_{31} \\ e_{31} & e_{36} & e_{35} & -\varepsilon_{13} \end{pmatrix},$$

$$\underline{\underline{M}}^{(bc)} = \begin{pmatrix} c_{56} & c_{52} & c_{54} & e_{52} \\ c_{46} & c_{24} & c_{44} & e_{42} \\ c_{36} & c_{32} & c_{34} & e_{32} \\ e_{36} & e_{32} & e_{34} & -\varepsilon_{23} \end{pmatrix}.$$

We now define $\vec{T}^T = (T_5, T_4, T_3, D_3)$ and form the vector

$$\vec{X} = \begin{pmatrix} \vec{U} \\ \vec{T} \end{pmatrix}. \tag{5}$$

One can show that the set of equations

$$\underline{\underline{\mathcal{L}}}\,\vec{X} = \vec{0}, \tag{6a}$$

$$\underline{\underline{\tau}}\,\vec{X} = \vec{T}, \tag{6b}$$

corresponds to Eqs. (1). After some algebra we obtain the system of differential equations

$$\underline{\underline{A}}\,\vec{X} = \frac{\partial}{\partial x_3}\vec{X}. \tag{7}$$

II.c Solutionansatz for doubly periodic structures: Choosing for the components of $\vec{X}$ solutions of the form

$$\Psi(x_1, x_2, x_3) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \Psi_{mn} e^{-jk_{1n}x_1} e^{-jk_{2m}x_2} e^{j\kappa_{mn}x_3}, \tag{8a}$$

truncating the double sum

$$\Psi = \sum_{n=-N}^{N} \sum_{m=-M}^{M} \{\cdots\}, \tag{8b}$$

and inserting in (7) yields the eigenvalue system for the determination of κ and the corresponding eigenvectors:

$$\underline{\underline{A}}\vec{X} = \kappa\vec{X}. \tag{9}$$

III. Dynamic, Periodic Green's Functions

III.a A Constructive Reasoning: According to Floquet's theorem the propagation of waves in p–periodic structures can be described by

$$f(x_1) = \sum_{n=-\infty}^{\infty} \bar{a}_n e^{-jk_n x_1}, \tag{10a}$$

where

$$k_n = k_0 + K_n = k_0 + \frac{2\pi n}{p}. \tag{10b}$$

Substituting (10b) in (10a) we obtain

$$\begin{aligned} f(x_1) &= \sum_{n=-\infty}^{\infty} \bar{a}_n e^{-jk_0 x_1} e^{-jK_n x_1} = \\ &= e^{-jk_0 x_1} \sum_{n=-\infty}^{\infty} \bar{a}_n e^{-jK_n x_1} = \\ &= e^{-jk_0 x_1} f_p(x_1). \end{aligned} \tag{11}$$

Remarks: (i) Obviously $f_p(x_1 + p) = f_p(x_1)$ is valid. (ii) Without any lose of generality k_0 can be restricted to the closed interval $-\frac{\pi}{p} \le k_0 \le \frac{\pi}{p}$. (iii) The $\bar{a}_n$ are related to the Fourier expansion coefficients of the periodic function $f_p(x_1)$

$$\bar{a}_n = \frac{1}{p} \int_{-\frac{p}{2}}^{\frac{p}{2}} f_p(x_1) e^{jK_n x_1} dx_1. \tag{12}$$

Regarding $f(x_1)$ as the field response, we now look for the simplest possible source function which can excite a wave according to $f(x_1)$. Analogous to the non-periodic wave excitation by single δ–functions, in the present case we suggest a periodic sequence of δ–functions,

$$\rho(x_1) = \pi \sum_{n=-\infty}^{\infty} e^{-jk_0 x_1} \delta(x_1 - np). \tag{13}$$

The introduction of the scaling factor π will soon be clear. Employing the equality

$$\sum_{n=-\infty}^{\infty} \delta(x_1 - np) = \frac{1}{p} \sum_{n=-\infty}^{\infty} e^{-jK_n x_1}, \tag{14}$$

Eq.(13) reads

$$\rho(x_1) = \frac{\pi}{p} \sum_{n=-\infty}^{\infty} e^{-jk_0 x_1} e^{-jK_n x_1}. \tag{15}$$

The latter expression for $\rho(x_1)$ compared with (11) shows that $\rho(x_1)$ also has formally the properties of a travelling wave in a periodic structure. Nevertheless we give a second interpretation, which aside from a physical significance, is amenable in describing a series of practically important devices.

The exponential $e^{-jk_0 x_1}$ is a non-singular smooth function, thus Eq. (13) equivalently reads

$$\rho(x_1) = \pi \sum_{n=-\infty}^{\infty} e^{-jk_0 np} \delta(x_1 - np). \tag{16}$$

The latter formulation allows a simple interpretation: Eq.(16) represents a periodic sequence of phased Dirac's impluses, which spatially are separated by a distance p and where the phase of two successive lines differ by $e^{-jk_0 p}$, Fig.3.

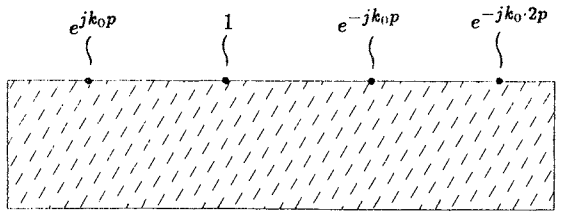


Fig.3 Phased line array

For further clarification of the subject let us consider some illustrative examples: (i) Assume $k_0p = \pi$. Then $e^{-jk_0np} = e^{-jn\pi}$. This corresponds to an alternating $+/-$ sequence of line sources (bidirectional two-phase array). (ii) For $k_0p = \frac{2\pi}{3}$ we obtain $e^{-jk_0np} = e^{-j\frac{2\pi}{3}n}$, which is a three-phase array. (iii) In a similar manner selecting $k_0p = \frac{2\pi}{4}$ and $k_0p = \frac{2\pi}{n}$, four- and $n-$phase arrays [16] can be discussed. A harmonic time-dependence is assumed, and the corresponding exponential function $e^{j\omega t}$ is omitted. Analogous to the non-periodic case, the response of the medium under consideration to a phased line array will be denoted as periodic Green's function.

Remark: (i) p characterizes the topological periodicity. (ii) In the case of $n-$phase arrays; i.e. $k_0p = \frac{2\pi}{n}$, the length np manifests a second charateristic quantity for periodicity (excitation periodicity). (iii) Finally let the line sources operate at a frequency ω_0, and let a surface wave, with a material dependent velocity v_s, be excited. The wavelength $\lambda_s = \frac{v_s}{\omega_0}$ is a third important quantity in our problem. The set of parameters p, np and ω_0 uniquely determines the system.

III.b Phased Arrays and the Dynamic-, Dyadic- Periodic- Green Functions: As a next step we wish to construct the Green's functions associated with our problem. To easily distinguish the nature of the various Green's functions appearing in the subsequent analysis, we preferably excite the medium by a phased electrode array, rather than by a phased line array, Fig.4.

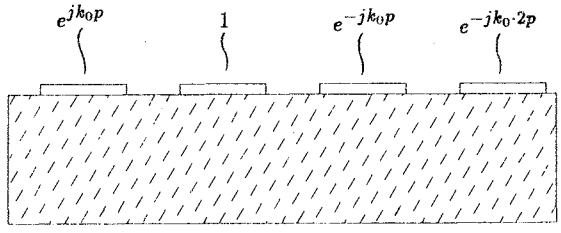


Fig.4 A phased electrode array

Thus let us begin considering a phased array of infinitely long thin parallel metallic electrodes located on the surface of a piezoelectric substrate (to facilitate an analytical discussion, and to compare our results with our previous work regarding single line excitation, we assume a piezoelectroic half-space). The neighbouring electrodes have a distance p and when passing along the positive x_1- axis, the phase of successive electrodes changes according to $e^{-jk_0x_1}$. We wish to denote the interval $[-\frac{p}{2}, \frac{p}{2}]$ as the fundamental cell. Further, by defintion the phase of the electrode in the fundamental cell is set to be equal to unity. Assume a certain charge density distribution in the fundamental cell to which the Fourier coefficients $\bar{\rho}_n$ are related, according to (12). Thus we write

$$\rho(x_1) = \pi \sum_{n=-\infty}^{\infty} \bar{\rho}_n e^{-jk_0x_1} e^{-jK_nx_1}. \tag{17}$$

In the special case $\bar{\rho}_n = 1$ we obtain our phased line array. To make possible an analytical discussion let us now assume that the material constitution is such as to excite only shear horizontal waves [13]. In this case the independent field variables are the mechanical displacement $u_2(x_1, x_3)$ and the electrical potential $\varphi(x_1, x_3)$. Similar to [14] we refer to the frequency-dependent field-responses to the periodic phased line array, periodic dynamic Green's functions $G^{(u,\rho)}(x_1, x_3)$ and $G^{(\varphi,\rho)}(x_1, x_3)$. A piezoelectric material can be excited both electrically as well as mechanically. Thus, similar to the above charge line arrays, the piezoelectric substrate can also be excited by a train of phased line forces. Let us denote the response of the medium to this excitation mechanism by periodic Green's functions $G^{(u,\tau)}(x_1, x_3)$ and $G^{(\varphi,\tau)}(x_1, x_3)$. In [14] and [15] we discused the functional properties of the Green's functions resulting from a single-line charge source and a single-line force. The question arises whether or not the same relations are valid in periodic case. A second question concerns a possible relationship between the Green's functions in periodic and non-periodic cases. Let us now consider a semi-infinite piezoelectric being forced simultaneously by the charge distribution $\rho(x_1)$, Eq.(17), and a stress distribution $\tau(x_1)$. As argued above, we take for $\tau(x_1)$ a similar function as for $\rho(x_1)$

$$\tau(x_1) = -j\pi \sum_{n=-\infty}^{\infty} \bar{\tau}_n e^{-jk_0x_1} e^{-jK_nx_1}. \tag{18}$$

The field distribution in the substrate can be written as

$$\psi(x_1, x_3) = \sum_{n=-\infty}^{\infty} \{\bar{\alpha}_n^{(1)} \bar{\psi}_n^{(1)} e^{j\kappa_n^{(1)} k_n x_3} + \\ + \bar{\alpha}_n^{(2)} \bar{\psi}_n^{(2)} e^{j\kappa_n^{(2)} k_n x_3}\} e^{-jk_nx_1}. \tag{19}$$

The quantity ψ represents any field quantity in the substrate. In the free space, using the potential continuity condition at the interface, we have

$$\varphi^{(f)}(x_1, x_3) = \sum_{n=-\infty}^{\infty} \{\bar{\alpha}_n^{(1)} \bar{\varphi}_n^{(1)} e^{-|k_n|x_3} + \\ + \bar{\alpha}_n^{(2)} \bar{\varphi}_n^{(2)} e^{-|k_n|x_3}\} e^{-jk_nx_1}. \tag{20}$$

For any given value k_n, Eq.(20) satisfies the Laplace equation in the free space. From the relation $D_3^{(f)}(x_1, x_3) = -\varepsilon_0 \frac{\partial \varphi^{(f)}}{\partial x_3}$ we obtain

$$D_3^{(f)}(x_1, x_3) = \sum_{n=-\infty}^{\infty} [\varepsilon_0 |k_n|] \{\bar{\alpha}_n^{(1)} \bar{\varphi}_n^{(1)} e^{-|k_n|x_3} + \\ + \bar{\alpha}_n^{(2)} \bar{\varphi}_n^{(2)} e^{-|k_n|x_3}\} e^{-jk_nx_1}. \tag{21}$$

If we now substitute (10b) and impose the remaining boundary conditions; i.e.,

$$D_3^{(f)}(x_1, 0) - D_3(x_1, 0) = \rho(x_1), \tag{22a}$$

$$T_4(x_1, 0) = \tau(x_1), \tag{22b}$$

we arrive at

$$\sum_{n=-\infty}^{\infty} \varepsilon_0 |k_n| \left[\bar{\alpha}_n^{(1)}\bar{\varphi}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{\varphi}_n^{(2)}\right] e^{-jK_n x_1} -$$
$$- \sum_{n=-\infty}^{\infty} \left[\bar{\alpha}_n^{(1)}\bar{D}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{D}_n^{(2)}\right] e^{-jK_n x_1} =$$
$$= \frac{\pi}{p} \sum_{n=-\infty}^{\infty} \bar{\rho}_n e^{-jK_n x_1}, \qquad (23a)$$

$$\sum_{n=-\infty}^{\infty} \left[\bar{\alpha}_n^{(1)}\bar{T}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{T}_n^{(2)}\right] e^{-jK_n x_1} =$$
$$= \frac{-j\pi}{p} \sum_{n=-\infty}^{\infty} \bar{\tau}_n e^{-jK_n x_1}. \qquad (23b)$$

In both equations the exponential term $e^{-jk_0 x_1}$ has been eliminated. However, it is worth mentioning that the term $|k_n|$ explicitly and the quantities $\bar{\varphi}^{(i)}$, $\bar{D}^{(i)}$,$\bar{T}^{(i)}$ implicitly depend on k_0. Moreover they depend on the frequency ω too. Now we multiply both sides of (23a) and (23b) by $e^{jK_m x_1}$ and integrate from $-p/2$ to $p/2$. Using the orthogonality property

$$\frac{1}{p}\int_{-\frac{p}{2}}^{\frac{p}{2}} e^{-jK_n x_1} e^{jK_m x_1} dx_1 = \delta_{mn}, \qquad (24)$$

and rearrangement yields

$$\bar{\alpha}_n^{(1)}\left[\varepsilon_0|k_n|\bar{\varphi}_n^{(1)} - \bar{D}_n^{(1)}\right] + \bar{\alpha}_n^{(2)}\left[\varepsilon_0|k_n|\bar{\varphi}_n^{(2)} - \bar{D}_n^{(2)}\right] = \frac{\pi}{p}\bar{\rho}_n, \qquad (25a)$$

$$\bar{\alpha}_n^{(1)}\bar{T}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{T}_n^{(2)} = \frac{-j\pi}{p}\bar{\tau}_n, \qquad (25b)$$

with (10b) and after some simple algebra and defining $\bar{\bar{A}}_n^{(1)}$, $\bar{\bar{A}}_n^{(2)}$, $\bar{\bar{T}}_n^{(1)}$ and $\bar{\bar{T}}_n^{(2)}$ we obtain

$$\bar{\alpha}_n^{(1)}\bar{\bar{A}}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{\bar{A}}_n^{(2)} = \bar{\rho}_n, \qquad (26a)$$
$$\bar{\alpha}_n^{(1)}\bar{\bar{T}}_n^{(1)} + \bar{\alpha}_n^{(2)}\bar{\bar{T}}_n^{(2)} = \bar{\tau}_n. \qquad (26b)$$

Solving for $\bar{\alpha}_n^{(1)}$ and $\bar{\alpha}_n^{(2)}$ we come to the result

$$\bar{\alpha}_n^{(1)} = \frac{\bar{\bar{T}}_n^{(2)}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}}\bar{\rho}_n + \frac{-\bar{\bar{A}}_n^{(2)}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}}\bar{\tau}_n,$$
$$\bar{\alpha}_n^{(2)} = \frac{-\bar{\bar{T}}_n^{(1)}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}}\bar{\rho}_n + \frac{\bar{\bar{A}}_n^{(1)}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}}\bar{\tau}_n. \qquad (27a,b)$$

According to (19) let us now write

$$u_2(x_1,x_3) = \sum_{n=-\infty}^{\infty} \left[\bar{\alpha}_n^{(1)}\bar{U}_n^{(1)} e^{j\kappa_n^{(1)}x_3} + \right.$$
$$\left. + \bar{\alpha}_n^{(2)}\bar{U}_n^{(2)} e^{j\kappa_n^{(2)}x_3}\right] e^{-jk_n x_1}, \qquad (28a)$$
$$\varphi(x_1,x_3) = \sum_{n=-\infty}^{\infty} \left[\bar{\alpha}_n^{(1)}\bar{\varphi}_n^{(1)} e^{j\kappa_n^{(1)}x_3} + \right.$$
$$\left. + \bar{\alpha}_n^{(2)}\bar{\varphi}_n^{(2)} e^{j\kappa_n^{(2)}x_3}\right] e^{-jk_n x_1}. \qquad (28b)$$

Substituting for $\bar{\alpha}_n^{(1)}$ and $\bar{\alpha}_n^{(2)}$ and rearranging we obtain

$$u_2(x_1,x_3) = \sum_{n=-\infty}^{\infty} \left[\bar{G}_n^{(u,\tau)}\bar{\tau}_n + \bar{G}_n^{(u,\rho)}\bar{\rho}_n\right] e^{-jk_n x_1}, \qquad (29a)$$
$$\varphi(x_1,x_3) = \sum_{n=-\infty}^{\infty} \left[\bar{G}_n^{(\varphi,\tau)}\bar{\tau}_n + \bar{G}_n^{(\varphi,\rho)}\bar{\rho}_n\right] e^{-jk_n x_1}, \qquad (29b)$$

where

$$\bar{G}_n^{(u,\tau)}(x_3|k_0,\omega) = \frac{-\bar{\bar{A}}_n^{(2)}\bar{U}_n^{(1)}e^{j\kappa_n^{(1)}x_3} + \bar{\bar{A}}_n^{(1)}\bar{U}_n^{(2)}e^{j\kappa_n^{(2)}x_3}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}},$$
$$\bar{G}_n^{(u,\rho)}(x_3|k_0,\omega) = \frac{-\bar{\bar{T}}_n^{(2)}\bar{U}_n^{(1)}e^{j\kappa_n^{(1)}x_3} - \bar{\bar{T}}_n^{(1)}\bar{U}_n^{(2)}e^{j\kappa_n^{(2)}x_3}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}},$$

$$\bar{G}_n^{(\varphi,\tau)}(x_3|k_0,\omega) = \frac{-\bar{\bar{A}}_n^{(2)}\bar{\varphi}_n^{(1)}e^{j\kappa_n^{(1)}x_3} - \bar{\bar{A}}_n^{(1)}\bar{\varphi}_n^{(2)}e^{j\kappa_n^{(2)}x_3}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}},$$
$$\bar{G}_n^{(\varphi,\rho)}(x_3|k_0,\omega) = \frac{-\bar{\bar{T}}_n^{(2)}\bar{\varphi}_n^{(1)}e^{j\kappa_n^{(1)}x_3} - \bar{\bar{T}}_n^{(1)}\bar{\varphi}_n^{(2)}e^{j\kappa_n^{(2)}x_3}}{\bar{\bar{A}}_n^{(1)}\bar{\bar{T}}_n^{(2)} - \bar{\bar{A}}_n^{(2)}\bar{\bar{T}}_n^{(1)}}.$$

$(30a-d)$

Figs. 5 and 6 show examples for phased point- and for two-dimensional phased electrode-arrays.

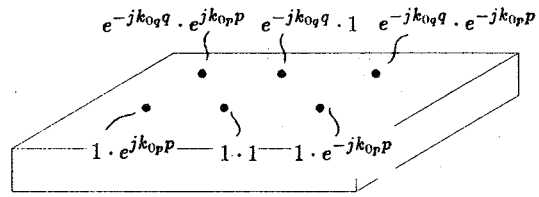


Fig.5 Example for a phased point array

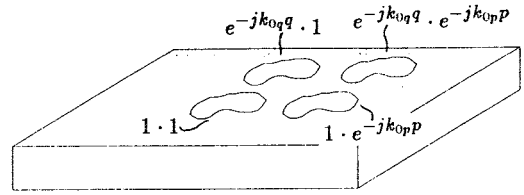


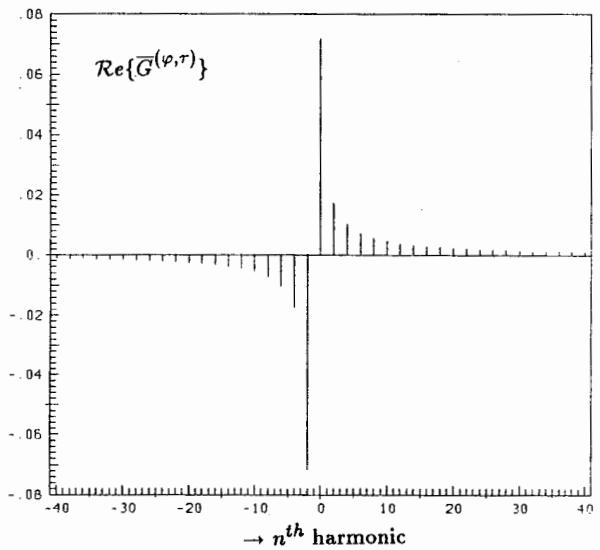
Fig.6 Example for a two dimensional phased electrode array

III.c Element Factors: Analogous to the electrostatic limit [16], we consider the following problem: In a $p-$periodic multi-phase electrode array we variate the exponential $e^{-jk_0 x_1}$ by continuously varying k_0 in the closed interval $[-\frac{\pi}{p}, \frac{\pi}{p}]$. In this process, to every value of k_0, an individual charge distribution can be assigned. We can show that the superposition of all the charges on the electrodes, leads to a frequency-dependent fundamental charge distribution (element factor). Moreover, in contrast to the electrostatic case, here it can be shown that, the definition of three additional fundamental stress distributions is meaningful.

IV. Results

We now apply our theory to phased line- and phased electrode-arrays on piezoelectric substrates. In our examples the piezoelectric substrate is Quarz. As mentioned above, the periodic field variables depend on the period p, the normalized excitation parameter $\xi(=\frac{k_0}{\frac{\pi}{p}})$ and the normalized frequency $\eta(=\frac{\omega p}{\pi v_0}$, here $v_0 = 5 \cdot 10^3 \mathrm{ms}^{-1})$.

Example 1: A periodic sequence of line forces excite the medium. A bi-phase line array is assumed (the electrical period is twice long as the geometrical period p). Figs.7a and 7b, respectively, show the real- and the imaginary- part of the Fourier coefficients for the potential response; i.e. the quantities $\overline{G}_n^{(\varphi,\tau)}$.



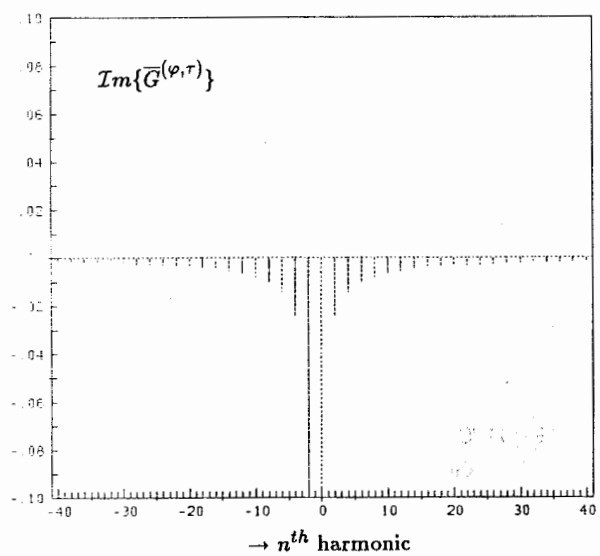


Fig.7 Fourier coefficients of $G^{(\varphi,\tau)}(x_1)$.

Example 2: Here, we consider a multi-phase line charge array on Quarz. Figs.8 and 9, respectively, show the electrical potential- and the mechanical displacement- response.

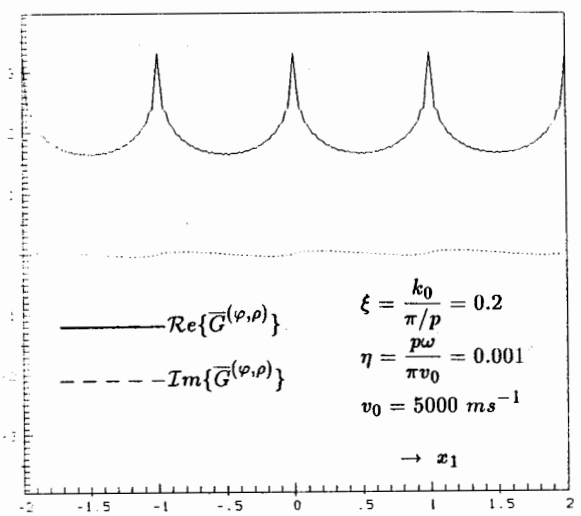


Fig.8 Electrical potential response to a line charge array

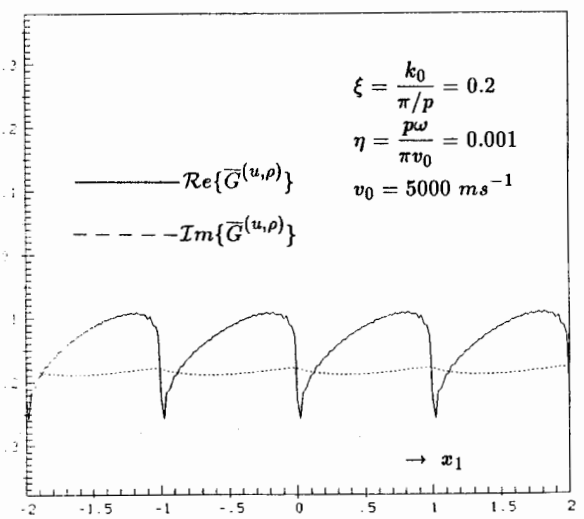


Fig.9 Mechanical displacement response to a line charge array

Fig.10 shows the same field quantities multiplied by the propagation exponential, $e^{-jk_0x_1}$.

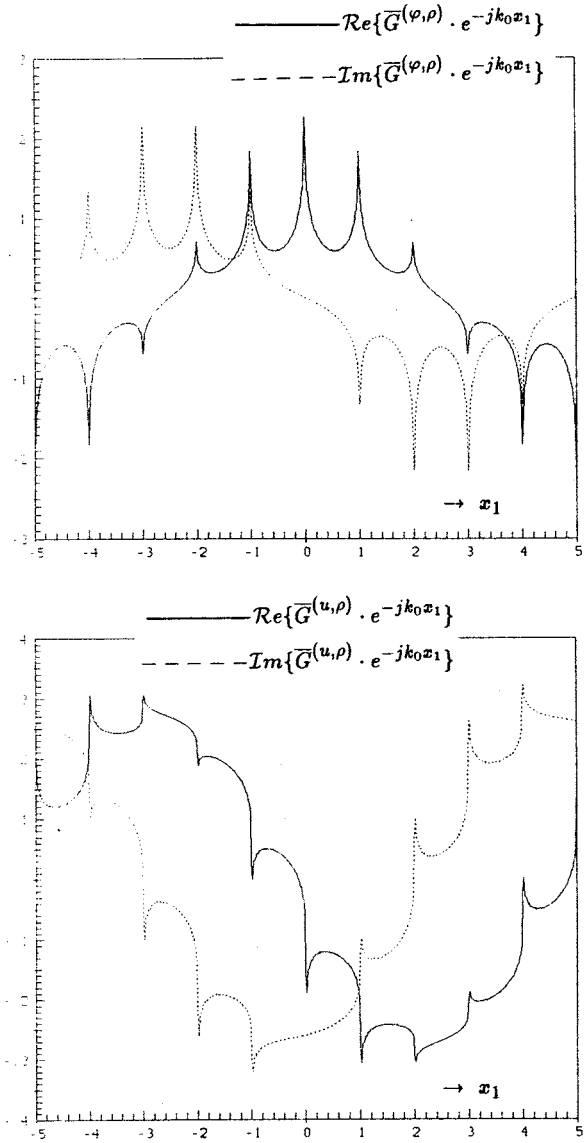


Fig.10 Potential and displacement response to a line charge array, $e^{-jk_0x_1}$ included.

Example 3: Our last example concerns the response of a Quarz plate to a three-phase electrode array. Fig.11 shows the corresponding charge- and potential- distribution.

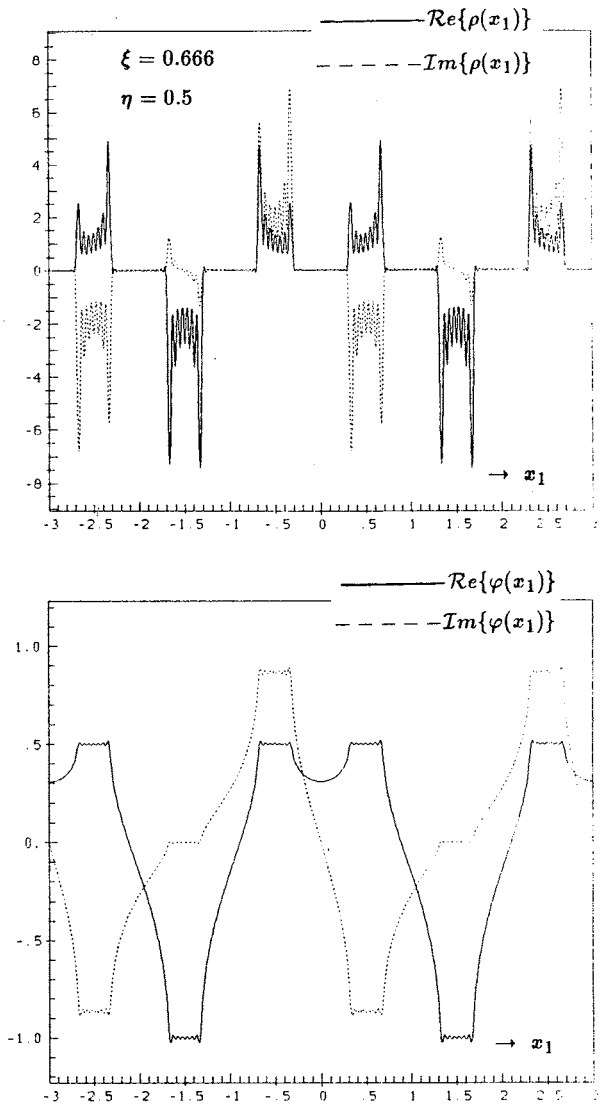


Fig.11 Charge- and potential-distribution in a three-phase transducer.

Conclusion

A general theory for the singly- and doubly-periodic structures is presented. It is shown that the concept of the phased line arrays leads to the definition of frequency- dependent dyadic periodic Green's functions. One fundamental charge- and three fundamental stress-distributions are suggested (element factors).

Acknowledgements

The work of A.R.B.W. was carried out partially under a Kurt Gödel grant, Ministry of Science and Research, Austria, whose support is gratefully acknowledged. The authors wish to thank Prof. F. Seifert and Dr. C. Ruppel for continuous encouragement and support.

References

[1] Kelvin, "Popular lectures", vol.1, pp.185, 1881.

[2] Vincent, "Phil. Mag., vol. 46, pp.537, 1898.

[3] R.L. Mooney, "Theory of an efficient interference filter", J. Opt. Soc. Amer., vol. 36, pp.256-260, 1946.

[4] R.M. White and F.W. Voltmer, "Direct Piezoelectric Coupling to Surface Elastic Waves", Appl. Phys. Lett. 7, pp.314-316, 1965.

[5] Q. Xue and Y. Shui , "Analysis of Leaky-Surface-Wave Propagating under Periodic Metal Grating, IEEE Trans. on UFFC, vol. 37, no.1, pp.13-25, 1990.

[6] A.R. Baghai-Wadji and A.A. Maradudin, "Rigorous Analysis of Surface Transverse Waves in Periodic Arrays with Arbitrary Electrode Profiles", to published in Proc. Ultrason. Conference, Hawaii, Honolulu, 1990.

[7] E.C. Titchmarch, "Eigenfunction Expansions Associated with second Order Differential Equations", Oxford University Press, 1958.

[8] F. Bloch, "Ueber die Quantenmechanik der Elektronen in Kristallgittern", Z. f. Phys., vol. 52, pp.555-600, 1928.

[9] A.R. Baghai-Wadji, O. Männer, and R. Ganß - Puchstein, "Analysis and Measurement of Transducer End Radiation in SAW in SAW Filters on Strongly Coupling Substrates", IEEE Trans. MTT, vol. 37, no. 1, pp.150-158, 1989.

[10] H. Engan, "Excitation of Elastic Surface Waves by Spatial Harmonics of Interdigital Transducers", IEEE Trans. Electron Devices, vol. ED-16, pp.1014-1017, 1969.

[11] A.R. Baghai-Wadji, S. Selberherr and F. Seifert, "On the Calculation of Charge, Electrostatic Potential and Capacitance in Generalized Finite SAW Structures", in Proc. IEEE Ultrason. Symp., pp.44-48, 1984.

[12] H. Bachl and Baghai-Wadji, "3D Electrostatic Field Analysis of Periodic Two-Dimensional SAW Transducers with Closed-Form Formulae", in Proc. IEEE Ultrason. Symp., pp.359-362, 1989.

[13] A.R. Baghai-Wadji, F. Seifert and K. Anemogiannis, "Rigorous Analysis of STWs in Nonperiodic Arrays Including Mechanical and Electrical Interactions", in Proc. IEEE Ultrason. Symp., pp.303-306, 1988.

[14] A.R. Baghai-Wadji, "Analysis of Surface Transverse Waves in Nonperiodic Arrays", in Proc. ISSWAS'89 Symp., pp.340-349, 1989.

[15] A.R. Baghai-Wadji, "Scattering of Piezoeletric Surface Transverse Waves from Electrodes with Arbitrary $\frac{h}{\lambda}$ ", in Proc. IEEE Ultrason. Symp., pp.377-380, 1989.

[16] S. Datta and B. Hunsinger, "Element Factor for Periodic Transducers", IEEE Trans. vol. SU-27, no. 1, pp.42-44, 1980.

[17] E. Langer, "Anregung und Ausbreitung elektroakustischer Wellen in piezoelektrischen Kristallen", PhD. Dissertation, Technische Universität Wien, Feb. 1986.

[18] B.A. Auld, "Acoustic Fields and Waves in Solids", A Wiley-Interscience Publication, John Wiley & Sons, New York, London, Sydney, Toronto, vol. I, 1973.

SESSION 2B

TIME AND FREQUENCY I

Chairman : M. BRUNET

Session 2B Number 14

TIME TRANSFER BY PASSIVE USE OF DIRECT BROADCASTING SATELLITE SIGNALS.

J A Davis, P R Pearce, P Menary*, M Hutchins**, M Sagin***.

National Physical Laboratory, Queens Road, Teddington, Middlesex, UK, TW11OLW.

* British Telecom International, Rugby Radio Station, Hillmorton, Rugby, UK, CV230AS.

** Hewlett Packard Ltd., Winnersh Triangle, Wokingham, UK, RG115DZ.

*** Radiocode Clocks Ltd, Radiocode House, Jennings Road, Kernick Road, Penryn, Cornwall, UK, TR109LY.

ABSTRACT

Time transfer experiments using the satellite TV method have been performed in order to compare atomic clocks at remote locations. Previous work has been extended using both the high powered Britsat satellite (D-MAC transmitting system), and the medium powered Astra satellite (PAL transmitting system). The satellites were viewed from four locations, with baselines varying from 40 to 400 km.

The earth stations consisted of domestic satellite receivers and small antennas. The antennas varied from 65cm diameter dishes to 40cm square arrays. Side by side observations of two independent sets of receiving equipment were undertaken.

Each location undertook common-view observations of both the Astra and Britsat satellites, enabling the errors in the two independent time transfer links to be assessed.

A pseudo ranging technique was applied to detect both the periodic motion of the satellites and any residual drift, using information from three widely spaced ground stations, each observing both satellites.

INTRODUCTION

The commissioning of several direct broadcasting satellite TV services within Europe has renewed interest in the use of geostationary satellites for common view time transfer experiments. Time transfer using common view of TV signals from geostationary satellites has been studied previously [1],[2],[3], including collaborative European experiments [4],[5]. Recent work at the National Physical Laboratory [6] used TV signals from the Astra satellite. This work demonstrated that time transfer at the sub-microsecond level was possible using low cost domestic TV satellite receiving equipment.

This paper reports the extension of this work using both the medium powered Astra (50 dBW) and high powered Britsat (Marco Polo) (60 dBW) satellites. The aim of the experiment is to determine the uncertainties in the time transfer method so that atomic clocks situated within the United Kingdom may be compared with absolute uncertainties better than 100ns irrespective of their distance from the National Physical Laboratory.

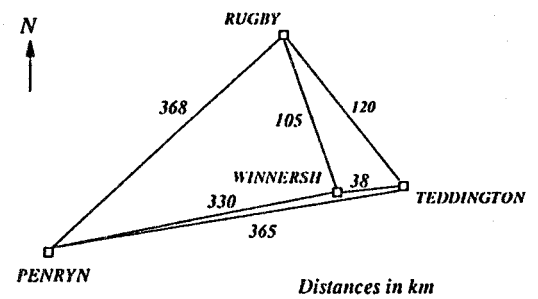
Several experiments are reported in this paper. Atomic clocks at two distant locations are compared via both the Astra and Britsat satellites. This permits direct comparisons of time transfers via two independent satellite links. The limiting uncertainties in a satellite common view time transfer experiment are determined by the motion of the satellite. Pseudo ranging techniques are employed, using only information gathered from three groundstations, to determine both the orbital components and residual drifts of the two satellites' motion. This information is used to reduce the uncertainties of the time transfer links, without the need of any form of external calibration eg by using GPS receivers or portable clock trips.

METHOD

Four laboratories participated in the experiment: The National Physical Laboratory, (Teddington), Middlesex, Rugby Radio Station, Hillmorton, (Rugby), Hewlett Packard Ltd, (Winnersh), Wokingham, Berkshire and Radiocode Clocks Ltd, (Penryn), Cornwall. Their relative locations are shown in Figure 1.

FIGURE 1

LOCATION AND DISTANCES BETWEEN GROUND STATIONS



The original experiment between Teddington and Rugby via the Astra satellite is described in [6]. The present system is similar except the television sets have been omitted (Figure 2). VHF switch boxes were used at Winnersh and Rugby in order to monitor the Astra and Britsat satellites using a single counter timer. At Penryn difference measurements were made between the line synchronization pulses from the Astra and Britsat satellite receivers. This enabled useful pseudo ranging information to be obtained without the use of additional equipment, but did not allow for simultaneous clock comparisons and pseudo ranging.

THE COMPARISON OF SATELLITE TV RECEIVING EQUIPMENT.

Astra is a medium-powered satellite transmitting a PAL signal. Britsat by contrast is a high-powered satellite transmitting a time-division packet signal. The Britsat receiver then generates locally it's own PAL signal. A pulse extractor based on a design developed at the Van Swinden Laboratories extracted a line synchronization pulse from the PAL signal. Amstrad and Sakura domestic Astra receivers, Philips LHB 6677/90 and Philips domestic Britsat receivers were used in the experiment. 65cm diameter dish antenna were used with Astra receivers, 60cm dish antenna and 40cm square array antennas were used with Britsat receivers.

The counter timer in each receiving apparatus measured a time interval between a 1 pps signal originating from an atomic clock, and a line synchronization pulse extracted from the television signal. The variations of these 'single receiver' time interval readings are of value in determining the orbital part of the satellite motion and are discussed later.

Direct comparisons were made between independent, side by side, sets of receiving equipment (Figure 3). Typical differential delays of 30ns and 70ns were measured for the Astra Sakura and Britsat Philips LHB 6677/90 receivers. The maximum variations of these delays were 10ns for the Astra Sakura and 20ns for the other receivers used. Each set of receiving equipment used at a remote location was calibrated at NPL against receivers used as references.

DETERMINATION OF THE ABSOLUTE TIME DIFFERENCES BETWEEN ATOMIC CLOCKS AT REMOTE LOCATIONS .

The WG84 co-ordinate system is used for calculating the positions of the earth stations (Table 1). An estimate of the satellite positions was obtained using the quoted mid-point of their orbit position.

D_{21Sat} is the delay difference between satellite signals reaching the ground stations G_2 and G_1 from satellite I. Corrections are added to this delay to account for different signal propagation delays occurring at each location (Table 2). The time interval TI_{21Sat} is measured between satellite signals reaching the ground stations G_2 and G_1 from satellite I was calculated with respect to the clocks at each location. $T_{c2} - T_{c1}$ is the time difference between the clocks at the two ground stations. Hence

$$TI_{21Astra} = D_{21Astra} + T_{c2} - T_{c1} \quad (1a)$$

$$TI_{21Britsat} = D_{21Britsat} + T_{c2} - T_{c1} \quad (1b)$$

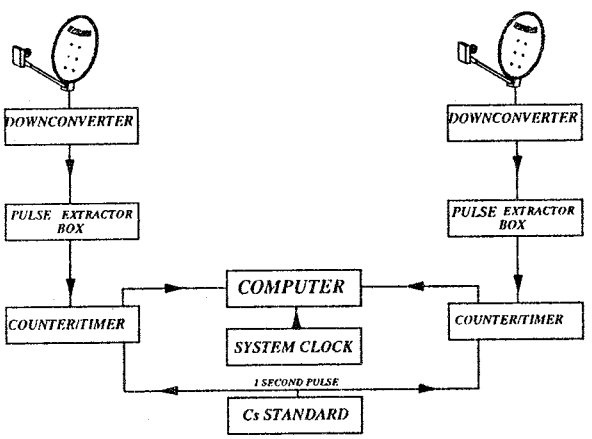
$$TI_{21Astra} - TI_{21Britsat} = D_{21Astra} - D_{21Britsat} = TI_{21(Astra-Britsat)} \quad (1c)$$

The variation of the measured value $TI_{21(Astra-Britsat)}$ depends only on instabilities in the time transfer links. The mean value of $TI_{21(Astra-Britsat)}$ was determined from several weeks measurements in order to average out the effect of satellite motion. The difference between the measured and calculated values of $TI_{21(Astra-Britsat)}$ is a measure of the discrepancy in the absolute time transfers undertaken via the Astra and Britsat satellites (Table 2). Absolute time transfers via the two satellites agree to better than 100ns.

Common view experiment have been undertaken between Teddington and Rugby using a ground based TV transmitter located at Sandy Heath [6]. This link has been calibrated via a caesium clock time transfer. The mean values of the satellite based TV time transfers and ground based link averaged over several weeks differed by 23ns and 2ns for the Astra and Britsat satellites respectively.

FIGURE 2

SIDE BY SIDE SYSTEMS



CHARACTERISTICS OF THE TIME TRANSFER LINKS STUDIED IN THIS EXPERIMENT.

Simultaneous observations of Astra and Britsat satellites allow for two independent time comparisons to be undertaken between each location. The difference between the time comparisons (Equation 1c) is independent of the performance of the clocks at each location. Table 3 gives the variation of these differences calculated on a daily basis.

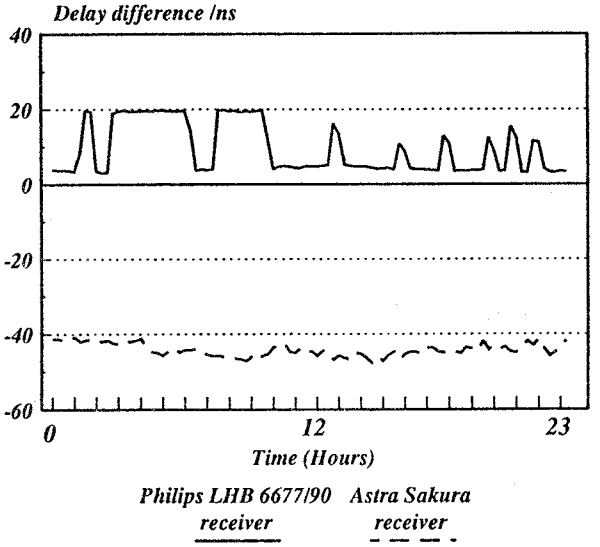
Examples of the time comparisons and Astra - Britsat difference comparisons are shown (Figure 4, Figure 5). Sidereal day periodic variations were observed on the comparisons. Example plots of $\mathrm{Log}_{10}\sigma_y$ against $\mathrm{Log}_{10}\tau$ is shown (Figure 14, Figure 15, Figure 16, Figure 17), for the Winnersh - Teddington and Rugby - Teddington comparisons.

TABLE 1
Co-ordinates of Ground Stations used in Astra and Britsat Common View Experiments

Laboratory Ground Station	Geographic Latitude (Degrees North)	Geographic Longitude (Degrees West)	Height Above Elipsoid (metres)
Teddington			
Astra	51.42509	0.34226	16
Britsat	51.42042	0.33722	12
Rugby			
Astra	52.36587	1.18644	105
Britsat	52.36587	1.18644	105
Winnersh			
Astra	51.43592	0.88726	50
Britsat	51.43592	0.88726	50
Penryn			
Astra	50.16311	5.12333	58
Britsat	50.16311	5.12333	58

FIGURE 3

Differential delay variations for two side by side receiver systems



The Winnersh - Teddington link was the shortest in the experiment. Hewlett Packard 5061A Option 004 caesium clocks were used at both locations. The drift rate between the two clocks showed a slow change of a few ns a day. This is attributed to the slow change in the rate of longitudinal drift of each satellite. The low level of σ_y recorded when the averaging time is a multiple of a sidereal day, is a result both of the high quality of the caesium clocks at each location, and of the quality of the time comparisons which may be obtained over short distances. The caesium clock at Rugby Radio Station was a Hewlett Packard model 5061A. The clock used at Penryn was a Rubidium frequency standard.

TABLE 2
Satellite to Ground Station Delay Differences

Ground Station Link	Delay Differences		Astra-Britsat Delay Differences	
	Calculated ns	Corrected ns	Calculated ns	Measured ns
Rugby - Astra	352185	357599	206796	206758
Teddington Britsat	150907	150803		
Winnersh - Astra	49783	55196	120756	120786
Teddington Britsat	-65490	-65560		
Penryn - Astra	91071	96484	1017274	1017302
Teddington Britsat	-920720	-920790		

FIGURE 4

Clock comparisons (Winnersh - Teddington) via both the Astra and Britsat satellites

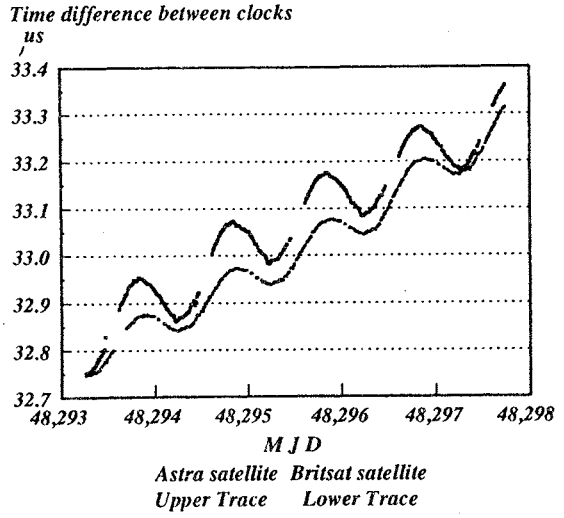


TABLE 3 Astra-Britsat Delay Difference Comparisons		
Link	Delay Difference Measured - Calculated ns	Standard Deviation of Measured Value ns
Winnersh-Teddington	30	32
Rugby-Winnersh	-68	57
Rugby-Teddington	-38	65
Penryn-Rugby	76	238
Penryn-Winnersh	-78	228
Penryn-Teddington	38	257

DETERMINATION OF THE SIDEREAL DAY PERIODIC MOTION OF THE ASTRA AND BRITSAT SATELLITES.

The orbital planes of the Astra and Britsat satellites will be inclined slightly out of the plane defined by the earth's equator. These orbits will also possess small eccentricities. When viewed from a frame of reference stationary with respect to the surface of the earth, small-amplitude sidereal-day periodic motions of the satellites are observed.

THEORY

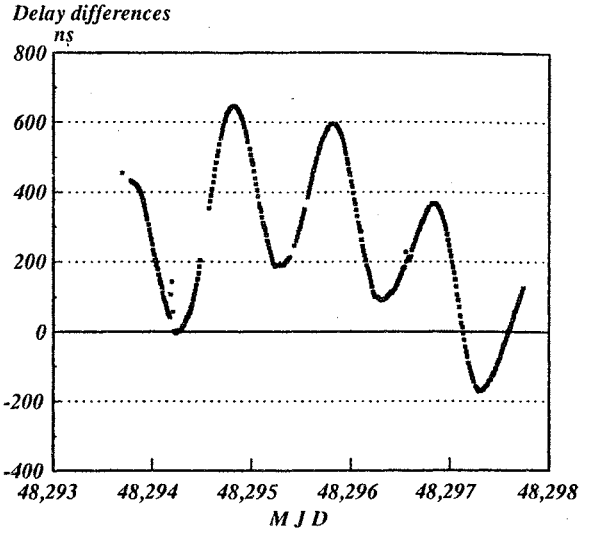
An earth centred co-ordinate system is used, fixed with respect to the earth's surface. The satellite orbit is described by its radius r, measured in meters from the earth's centre, longitude ϕ, measured in radians eastwards from the Greenwich meridian, and its latitude θ, measured in radians northwards from the plane of the equator.

The inclination of the plane of orbit of the satellite out of the plane defined by the equator will produce a sinusoidal motion of the satellite latitude of period one sidereal day [7]. The small eccentricity of the satellite orbit will produces similar variations of both the satellite radius and longitude. The variations of the satellite radius and longitude will have amplitudes er, and $2e$ respectively where e is the eccentricity [7]. The phase of the longitudinal variation is $\pi/2$ ahead of the phase of the radius variation.

The satellite position is given by $R(r,\theta,\phi)$. $t_{\alpha\beta}$ represent the differential delay between satellite TV signals reaching the ground stations α and β. $t_{T\alpha}$ is the time taken for signals to pass from the transmitting station T up to the satellite and down to the receiver at α. The variation of this delay may be obtained from the 'single receiver' readings discussed earlier, limited by the performance of the local oscillator used at the TV transmitting centre.

FIGURE 5

Astra - Britsat difference comparisons for the Penryn - Teddington link



Consider each of the components or the satellite motion and the clock comparison readings to be perturbed by a sinusoidal motion of period one sidereal day. Then

$$r = r_0 + dr\sin(\omega_0 t + \delta_r) \qquad (2a)$$

$$\theta = \theta_0 + d\theta\sin(\omega_0 t + \delta_\theta) \qquad (2b)$$

$$\phi = \phi_0 + d\phi\sin(\omega_0 t + \delta_\phi) \qquad (2c)$$

$$t_{\alpha\beta} = t_{0\alpha\beta} + dt_{\alpha\beta}\sin(\omega_0 t + \delta_{\alpha\beta}) \qquad (2d)$$

$$t_{T\alpha} = t_{0T\alpha} + dt_{T\alpha}\sin(\omega_0 t + \delta_{T\alpha}) \qquad (2e)$$

$dr, d\theta, d\phi$, $dt_{\alpha\beta}$, $dt_{T\alpha}$ are small perturbations in the satellite orbit parameters and measured delay differences, r_0, θ_0, ϕ_0, $t_{0T\alpha}, t_{0\alpha\beta}$, are the mean radius, latitude, and longitude of the satellite and mean delay differences between the two locations , ω_0 is the angular frequency equivalent to a period of one sidereal day, $\delta_r, \delta_\theta, \delta_\phi, \delta_{\alpha\beta}, \delta_{T\alpha}$, are the phases of the perturbations.

Each perturbing element in the satellite motion dr, $d\theta$, and $d\phi$ will produce a contribution to each of the delay perturbations $dt_{\alpha\beta}$ and $dt_{T\alpha}$. An alternative expression may now be used: $dr\sin(\omega_0 t + \delta_r) = dr_s\sin(\omega_0 t) + dr_c\cos(\omega_0 t)$, with similar expressions for $d\theta$, $d\phi$ and $dt_{t\alpha}$ and $dt_{\alpha\beta}$. Let the satellite radius perturbation be represented by $dr\sin(\omega_0 + \delta_r)$, where $dr = er_0$. Then the longitudinal perturbation will be given by $2edr\cos(\omega_0 t + \delta_r)$. $K_{r\alpha\beta}$ is the delay change introduced in the time comparison α β per unit change of the satellite radius r, with similar expressions for K_θ and K_ϕ.

The perturbations induced in the delay measurements may be obtained by summation of the effect of the sinusoidal perturbations of the component of the satellite co-ordinates. Consider two time comparison links $\alpha_1\beta_1$ and $\alpha_2\beta_2$. By equating co-efficients of $\sin(\omega_0 t)$ and $\cos(\omega_0 t)$ and incorporating the above, the following equations may be obtained.

$$dt_{s\alpha1\beta1} = K_{r\alpha1\beta1}dr_s - K_{\phi\alpha1\beta1}2edr_c + K_{\theta\alpha1\beta1}d\theta_s \qquad (3a)$$

$$dt_{c\alpha1\beta1} = K_{r\alpha1\beta1}dr_c + K_{\phi\alpha1\beta1}2edr_s + K_{\theta\alpha1\beta1}d\theta_c \qquad (3b)$$

$$dt_{s\alpha2\beta2} = K_{r\alpha2\beta2}dr_s - K_{\phi\alpha2\beta2}2edr_c + K_{\theta\alpha2\beta2}d\theta_s \qquad (3c)$$

$$dt_{c\alpha2\beta2} = K_{r\alpha2\beta2}dr_c + K_{\phi\alpha2\beta2}2edr_s + K_{\theta\alpha2\beta2}d\theta_c \qquad (3d)$$

The co-efficients $K_{r\alpha\beta}$, $K_{\phi\alpha\beta}$, $K_{\theta\alpha\beta}$ are easily calculated from the satellite and ground station co-ordinates. The four unknowns dr_s, dr_c, $d\theta_s$ and $d\theta_c$ were determined using a matrix method.

COMPARISON WITH EXPERIMENTAL RESULTS

The time comparisons measured for each link were averaged over a sidereal day. The differences between the original readings and the mean values are then calculated. This gives the "periodic" part of the delay measurements. The sidereal day periodic variations of the time comparisons for a three way link between ground stations G_0, G_1, and G_2 observing the same satellite are related as follows: $dt_{G0G2} = dt_{G0G1} + dt_{G1G2}$. The difference comparisons between Astra and Britsat for a single link being related as follows: $dt_{Astra-Britsat} = dt_{Astra} - dt_{Britsat}$

A curve fitting technique is used to fit the parameters of a sine wave to the periodic part of the clock comparisons via both the Astra and Britsat satellites and the difference comparisons. The above relationships are incorporated as constraints. A single estimate of the satellite orbit parameters is made for a triangle of ground stations. The use of one of the single receiver time interval comparisons paired with each of the three links allowed additional estimates of the orbit parameters to be determined. In practice only one of the single receiver comparisons may be used because of the relatively close spacing of the ground-stations in comparison with the distance to the satellite.

The sidereal-day periodic motions of both the Astra and Britsat satellites were determined from observations made at Teddington, Rugby, and Winnersh (Table 4). The amplitude and phase values given in Table 4 are as defined in equations 2. The term $\omega_0 t$ has value zero at midnight on the day under examination.

In order to check the validity of the calculated satellite orbit parameters the sidereal day variations of time comparison links not used in the orbit parameter calculation were determined. The orbital components of the satellite's motion were used to calculate values for the sidereal-day variations of one of the single-receiver comparisons, and the Astra - Britsat comparisons between Penryn and Teddington. These calculated values were compared with values obtained experimentally by the curve fitting technique (Table 5). One discrepancy occurred for the single-receiver readings obtained from the Astra satellite. The instability of the TV local oscillator used for the Astra satellite broadcasts made these readings unreliable.

Changes in the eccentricity and inclination of the satellite orbits are given over several days (Figure 6, Figure 7). The rate of change of the inclination of the orbital plane depends on the positions of the moon and the sun [7]. Experimental results shown a varying drift rate.

The satellite eccentricity will generally remain constant between longitude correcting station keeping manoeuvres. There is a small 14 day periodic variation because of the effect of the moon and a longer term slow drift due to the solar radiation pressure. Experimentally the effects was not resolved and the eccentricities of both satellites were found to remain approximately constant.

TABLE 4
CALCULATED ORBIT PARAMETERS OF SATELLITES
8th FEBRUARY 1991

	AMPLITUDE x10^{-4}	PHASE radians
ASTRA Eccentricity	2.65	1.28
Inclination	4.85 radians	-0.22
BRITSAT Eccentricity	1.69	1.24
Inclination	1.35 radians	3.84
PERTURBATIONS OF LINKS CALCULATED FROM ABOVE VALUES		
	AMPLITUDE s	PHASE radians
Penryn-Teddington	2.20×10^{-7}	2.28
Single Channel	5.17×10^{-5}	1.10
MEASURED PERTURBATIONS OF LINKS		
Penryn-Teddington	2.23×10^{-7}	2.89
Single Channel	5.51×10^{-5}	1.19

FIGURE 6

Eccentricity of the Astra and Britsat satellite orbits

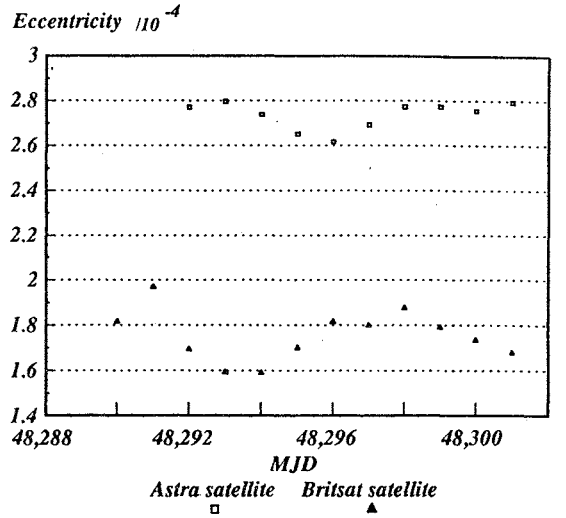


FIGURE 7

Inclination of the orbital plane of Astra and Britsat satellites

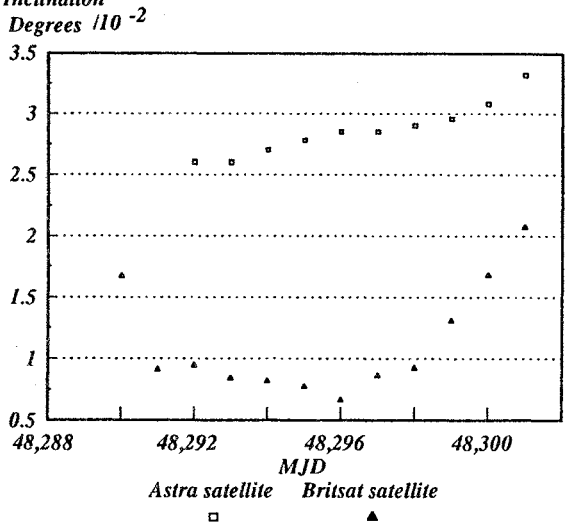
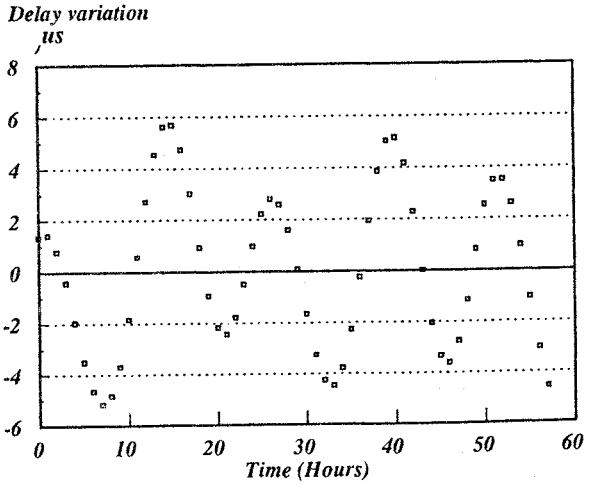


FIGURE 8

Single receiver time interval readings for Britsat satellite



Time recorded in hours from start of MJD 48251

Drift rate of local oscillator removed

There is a phase relationship between the sidereal day period sinusoidal motion observed on comparisons perturbed most by variations in the satellite range (single receiver readings) and comparisons perturbed most by variations in the satellite longitude. Winnersh - Teddington comparisons are approximately $\pi/2$ ahead of the single receiver comparisons. This is directly related to the phase relationship between the sidereal-day period longitude and radius perturbations produced by the eccentricity of the satellite.

PERTURBING EFFECT OF THE MOON AND THE SUN

Tidal effects of the moon and the sun are the main sources of non sidereal-day periodic motion of the satellite's position [7]. Sinusoidal variations to both the satellite radius and longitude of period 12.5 hours are produced. Typical resulting variations of the time transfer links have an amplitude of 40ns for the relatively long links to Penryn, and 5ns for the shorter links. In the example shown (Figure 8), the sidereal day part of the periodic motion was small, allowing the tidal perturbations to be clearly observed.

DETERMINATION OF THE LONGITUDINAL DRIFTS OF THE ASTRA AND BRITSAT SATELLITES

Consider two geostationary satellites, with widely differing longitudinal positions in their geostationary orbits. $S1(r,\theta,\phi)$, $S2(r,\theta,\phi)$ are the positions of the satellite, r, θ and ϕ being defined earlier. The satellite radius and latitude when averaged over a sidereal day will both to a first approximation be constant. The satellite positions $S1(\phi_1)$, $S2(\phi_2)$ are considered to be functions of their longitudinal values only.

The differential delay values D_{21Sat1}, D_{31Sat1}, D_{21Sat2} and D_{31Sat2} for satellite signals reaching the ground stations G_1, G_2 and G_3 have been defined earlier. $T_{c2}-T_{c1}$, and $T_{c3}-T_{c1}$ are the time differences between the clocks at the two remote ground stations and the calibrating laboratory. TI_{absat1} is the time interval measured between satellite signals reaching the ground stations G_a and G_b from satellite 1, relative to the clocks at each location. From equations (1)

$$TI_{21Sat1} = D_{21Sat1}(\phi_{Sat1}) + T_{c2} - T_{c1} \qquad (4a)$$

$$TI_{31Sat1} = D_{31Sat1}(\phi_{Sat1}) + T_{c3} - T_{c1} \qquad (4b)$$

$$TI_{21Sat2} = D_{21Sat2}(\phi_{Sat2}) + T_{c2} - T_{c1} \qquad (4c)$$

$$TI_{31Sat2} = D_{31Sat2}(\phi_{Sat2}) + T_{c3} - T_{c1} \qquad (4d)$$

The longitude of the two satellites, ϕ_{Sat1}, ϕ_{Sat2}, and the time differences of the two atomic clocks $T_{c2}-T_{c1}$ and $T_{c3}-T_{c1}$ may be determined. The dependence of the four values D_{21Sat1}, D_{31Sat1}, D_{21Sat2} and D_{31Sat2} on the satellite longitude is given by the parameters $K_{\phi 21Sat1}$, $K_{\phi 21Sat2}$, $K_{\phi 31Sat1}$, $K_{\phi 31Sat2}$. Subtracting equations (4c) and (4d) from equations (4a) and (4b) gives

$$TI_{21(Sat1-Sat2)} = TI_{21Sat1} - TI_{21Sat2} = D_{21Sat1}(\phi_{Sat1}) - D_{21Sat2}(\phi_{Sat2}) \qquad (5a)$$

$$TI_{31(Sat1-Sat2)} = TI_{31Sat1} - TI_{31Sat2} = D_{31Sat1}(\phi_{Sat1}) - D_{31Sat2}(\phi_{Sat2}) \qquad (5b)$$

The values of $TI_{21(Sat1-Sat2)}$ and $TI_{31(Sat1-Sat2)}$ may be determined for the two links. ϕ_{0sat2} and ϕ_{0sat1} represent the centre values of the satellite longitudes, $\delta TI_{21(Sat1-Sat2)}$ and $\delta TI_{31(Sat1-Sat2)}$ represent small changes in the delay differences, $\delta\phi_{Sat1}$ and $\delta\phi_{Sat2}$ be small variations of these values. Then

$$\delta TI_{21(Sat1-Sat2)} = K_{\phi\alpha 2\beta 1Sat1} d\phi_{Sat1} - K_{\phi\alpha 2\beta 1Sat2} d\phi_{Sat2} \qquad (6a)$$

$$\delta TI_{31(Sat1-Sat3)} = K_{\phi\alpha 3\beta 1Sat1} d\phi_{Sat1} - K_{\phi\alpha 3\beta 1Sat2} d\phi_{Sat2} \qquad (6b)$$

and $d\phi_{Sat1}$ and $d\phi_{Sat2}$ may be determined.

The earth's spherical asymmetry results in a longitudinal component of the gravitational potential. A series of station keeping manoeuvres are used to limit the resulting variation of a geostationary satellite. Between station keeping manoeuvres the change of the satellite longitude will take the form $\phi = \phi_0 + \alpha(t-t_0)^2$ [7] where ϕ_0 be the furthest longitude reached by the satellite in the direction of increasing gravitational potential, this occurring at time t_0. The constant α has a value of 1.53×10^{-3} degrees/day^2 and -9.34×10^{-4} degrees/day^2 for the Astra and Britsat satellites respectively. Plots of the expected free motion of the Astra and Britsat satellites are shown (Figure 9).

RESULTS

The Astra and Britsat satellites positioned at 19.2 E and 31.0 W are spaced a sufficient distance apart for the above method to be used. Four locations participated in the experiment. This permits a degree of redundancy to be introduced.

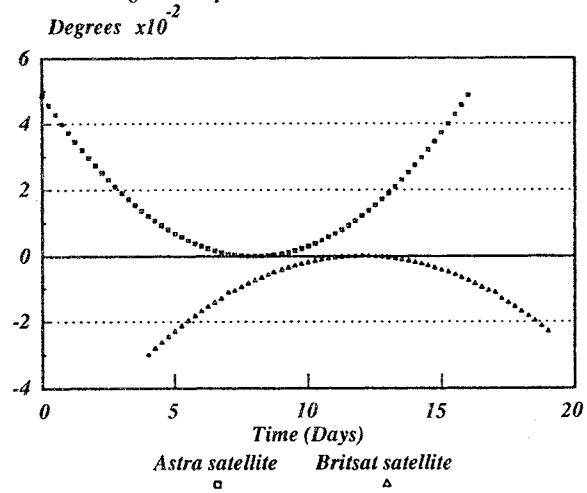
The time comparisons between each pair of locations were averaged over a sidereal day. One estimate of the absolute satellite longitudes of both the Astra and Britsat satellites was calculated for each 'triangle' of ground stations. Plots are shown for the mean daily satellite longitudes (Figure 10, Figure 11), calculated for the ground station triangles Teddington-Winnersh-Rugby, Teddington-Penryn-Rugby and Winnersh-Penryn-Rugby. Results are not plotted for the triangle Teddington-Winnersh-Penryn. This triangle possessed a very short perpendicular bisector (14 km) which made these results unreliable.

Plots of measured satellite longitude against time (Figure 10, Figure 11) display curves which agree well with the expected parabolic form. Sharp changes in the gradients of the curves were observed and attributed to longitudinal station keeping manoeuvres.

Astra - Britsat difference comparisons averaged over a sidereal day for the links Winnersh - Teddington, Rugby - Teddington, and Penryn - Teddington possess a slowly increasing gradient to their curves. These gradient changes may be calculated from the K_ϕ and α constants. Sharp changes in gradient occur on

FIGURE 9

Calculated variation of mean satellite longitude for Astra and Britsat satellites



Zero longitude represents position of highest gravitational potential.

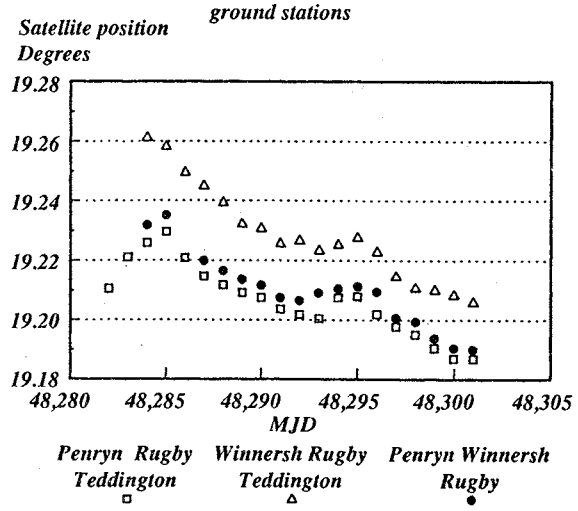
MJD 48295 at approximately 01:00 hrs and on MJD 48296 at approximately 01:00 hrs (Figure 12, Figure 13). These correspond to longitudinal orbit corrections of the Astra and Britsat satellites respectively, and agree with the sharp gradient changes shown in Figure 10 and Figure 11.

CORRECTION OF TIME COMPARISONS

The longitudinal position of the Astra and Britsat satellites are used for the correction of time comparisons. A ground stations triangle (eg Teddington-Rugby-Penryn) was used to determine the position of the satellites, a separate link (not one of the three forming the triangle) eg Teddington-Winnersh was then corrected. It should be noted that the clock independent Astra-Britsat difference measurements are used to calculate the satellite longitude positions and hence the link corrections. Even if the link being corrected forms part of the calibrating ground station triangle this will not invalidate the results. The orbital part of the satellite motion was corrected using orbit parameters determined from two separate links so as not to use any clock dependent information in the correction of the links.

FIGURE 10

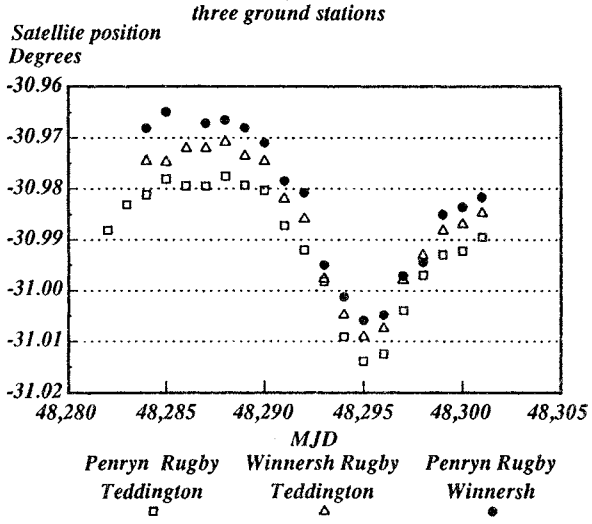
Astra satellite longitude calculated from three ground stations



Data averaged over a sidereal day

FIGURE 11

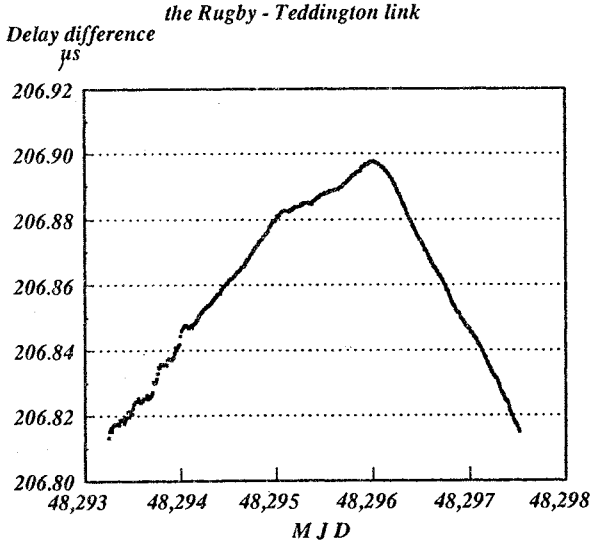
Britsat satellite longitude calculated from three ground stations



Data averaged over a sidereal day

FIGURE 12

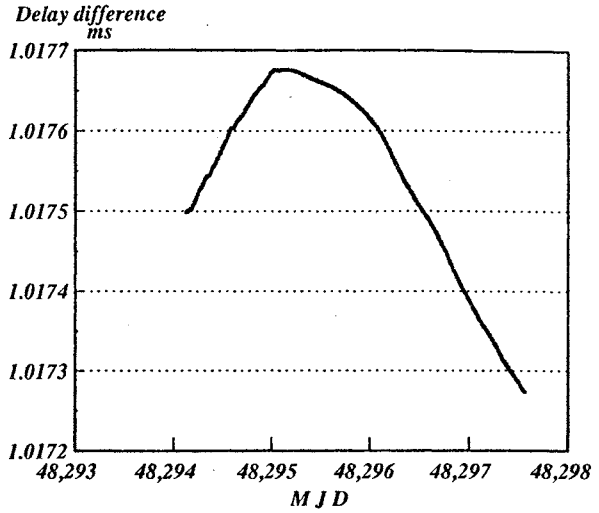
Astra - Britsat difference comparisons for the Rugby - Teddington link



Data averaged over a sidereal day

FIGURE 13

Astra - Britsat difference comparisons for the Penryn - Teddington link



Data averaged over a sidereal day

FIGURE 14

Plots of $Log_{10}(\sigma_y)$ against $Log_{10}(\tau)$ for clock comparisons via the Astra satellite

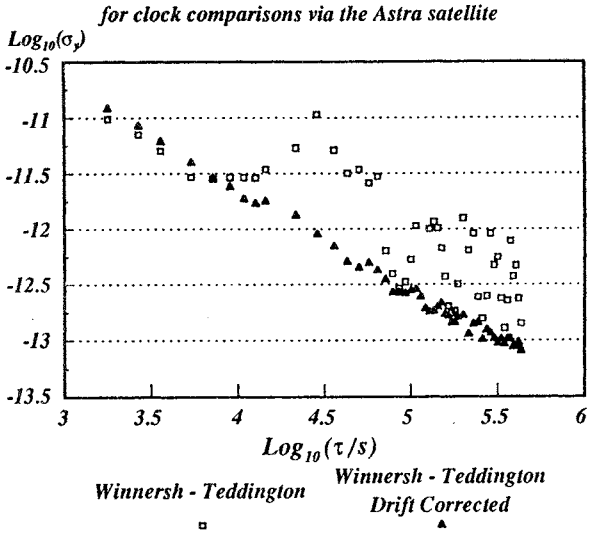
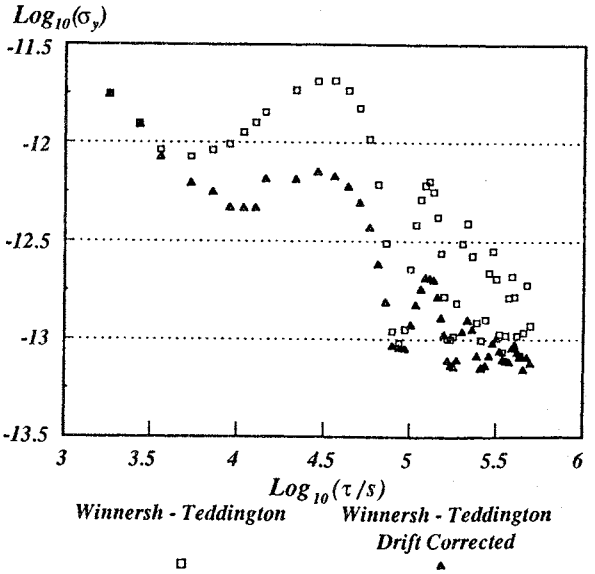


FIGURE 15

Plots of $Log_{10}(\sigma_y)$ against $Log_{10}(\tau)$ for clock comparisons via the Britsat satellite



Plots of $Log_{10}\sigma_y$ against $Log_{10}\tau$ are given for averaging times of up to five days (Figure 14, Figure 15, Figure 16, Figure 17). These have been calculated from the corrected and the uncorrected data collected over the same period. Plots for the Winnersh - Teddington link show a reduction in the values of σ_y for the corrected data for averaging times greater than a sidereal day. This reduction in σ_y was larger than that obtained by simply correcting the data for the sidereal day periodic part of its motion, thus demonstrating that the effect of the longitudinal drifts in the satellite motion have been in part corrected. There was a discrepancy in the corrected value of σ_y obtained, between comparisons made via the Astra and Britsat satellites. This was due to noise from the Astra receivers. The correction of the orbital elements of the satellite motion was not complete. However the amplitude of any residual periodic variations has been reduced by an order of magnitude for comparisons via the Britsat satellite. The reduction in the values of σ_y after the longitudinal drift correction was smaller, and the residual sidereal day variations greater for the Rugby - Teddington link compared with the Winnersh - Teddington link. The sidereal day periodic variations of the uncorrected data are both greater and more variable in the case of the Rugby - Teddington link then with the Winnersh -Teddington link. These variations are more dependent on the inclination of the satellite orbit which is more variable than the eccentricity. The Rugby - Teddington and Winnersh - Teddington links are approximately north - south and east - west orientated respectively (Figure 1).

FIGURE 16

Plots of $Log_{10}(\sigma_y)$ against $Log_{10}(\tau)$ for clock comparisons via the Astra satellite

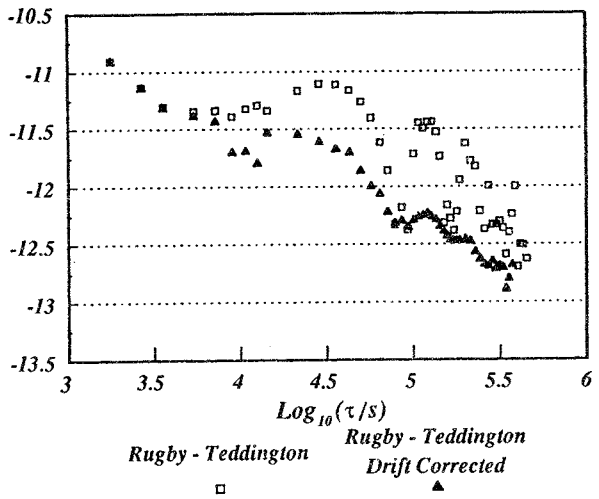
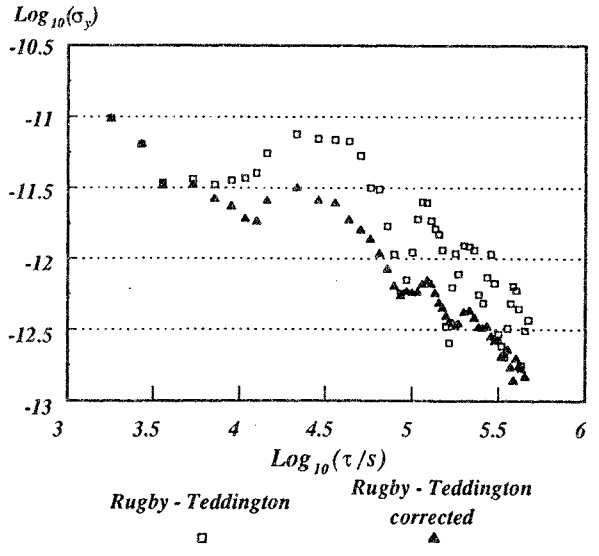


FIGURE 17

Plots of $Log_{10}(\sigma_y)$ against $Log_{10}(\tau)$ for clock comparisons via the Britsat satellite



ESTIMATION OF THE UNCERTAINTIES IN THE METHOD

The main sources of error in the method are uncertainties in the satellite radius, the effect of non sidereal day periodic motions within the satellite orbit, and uncertainties in the delay measurements made by the satellite receivers.

For a typical satellite longitude station keeping manoeuvre the satellite radius will vary by a few km [7]. A value of ± 5km is in excess of the typical maximum orbit variation. Such a variation would introduce uncertainties into the time transfer comparisons of the order of a few ns.

Periodic variations of period 12.0 hours and 12.5 hours due to the tidal effects of the sun and moon respectively will be present. These effects can be reduced by averaging over a sidereal-day by a factor of 0.04 for periodic motion of period 12.5 hours. Estimates for the maximum uncertainties given by the lunar periodic motion are a few ns.

The limiting error in the method is due to the variation of the differential delays within the ground stations. A typical maximum value of ± 20ns is used here, for both the Astra and Britsat satellites. Consider the widest spaced triangle of ground stations (Teddington - Rugby - Penryn). Errors of ± 20ns in the Astra-Britsat difference comparisons will produce errors in the determination of the satellite longitudes of $5x10^{-5}$ radians and $7x10^{-5}$ radians for the Astra and Britsat satellites respectively. These satellite longitude errors will produce errors in the corrected clock comparisons as given in Table 5.

The direct method of measuring the time difference between an Astra and Britsat line synchronization pulse using a single counter timer may introduce a further error. The problem is one of slippage between line synchronization pulses depending on which pulse arrives at the remote location first. With 50 frame synchronization pulses a second the resulting timing error is of the order of 2ns for each mismatch of 1 frame synchronization pulse.

CONCLUSION

The use of a direct broadcasting satellite transmitting a D-MAC signal has been shown to be suitable for common view time transfer experiments.

A method for the correction of the satellite motion has been demonstrated which allows comparisons of clock frequencies to a part in 10^{13} or better over short baselines using only information collected from three groundstations each observing two satellites.

The accuracy of the pseudo ranging method is limited by the differential delay variations of the satellite receivers and the spacing between the ground stations.

With calibration of the delays occurring at each ground station, accurate co-ordinates of the satellite dishes, and correction for the satellite motion, absolute time transfers with an accuracy of 100ns or better are possible using the Astra and Britsat satellites. Such accuracy has been demonstrated for the link between Rugby and Teddington and is consistent with the results obtained over the other links.

ACKNOWLEDGEMENTS

The authors thank B.R.Swabey, D.J.E.Knight, R.Jefferies, P.B.Coates and D.R.Allen for valuable advice given during the course of the work, C.Fleming, K.Roke and R.Cookman for the monitoring of equipment and collection of data.

TABLE 5
ERRORS IN CORRECTED CLOCK COMPARISONS

Ground Station Link	Error / 10^{-8}s	
	Astra	Britsat
Winnersh-Teddington	0.6	0.8
Rugby-Teddington	0.4	2.3
Rugby-Winnersh	0.2	1.5
Penryn-Teddington	6.3	5.9
Penryn-Winnersh	5.7	5.0
Penryn-Rugby	5.9	3.5

REFERENCES

[1] Tolman J., Ptacek V., Stecher R., : Microsecond clock comparison by means of TV synchronized pulses, IEEE Trans. Instr. and Meas vol. IM - 16, Sept 1969, pp 247 -254.

[2] Kirchner D, Iedler W R, : Timing by Satellite, recent developments and future experiments, Proc Int. Symp. on Satellite Transmissions, Graz, Austria, Sept 1985, pp 175 - 185.

[3] Sutcliffe D S , Swabey B R and Dare J H : International synchronous time measurement, British Electromagnetic Measurements Conference, National Physical Laboratory, Teddington, 1st - 3rd Oct 1985.

[4] Buzek O, Cermak J, Vondrak J, Cordara F, Galliano P G, Pettiti V, Tavella P, : Synchronization of time scales by television method using ECS satellites, preliminary results, Proc. Third EFTF, Besancon 1989, pp 204-214.

[5] Cordara F, Pettiti V, Cenci A, Fermi M, Sciarretta C : Time and Frequency comparisons in Europe by means of ECS 5 geostationary satellite, 22nd PTTI Vienna, Virginia 1990.

[6] Pearce P R, Davis J A, Allen D R, Swabey B R, Jefferies R, Knight D J E, Menary P, : Time transfer in the United Kingdom using a direct broadcasting satellite, Fourth EFTF, Neuchatel, 13th - 15th Mar 1990, pp 643-650.

[7] Soop M E, : Introduction to Geostationary Orbits, ESA SP-1053, European Space Agency 8-10 rue Mario-Nikis, 75738 PARIS CEDEX 15, France Nov. 1983.

Session 2B Number 15

Time transfer using a geostationary direct TV satellite.

F. Meyer*, M. Brunet**, M. Granveaud***, F. Vernotte*, M. Vincent*.

*Observatoire de Besançon, **CNES Toulouse, ***LPTF Observatoire de Paris.

Abstract

This paper reports the first results of a time transfer experiment between three stations located at CNES Toulouse (CT), LPTF in Paris (OP) and Besançon (OB), by use of the TV signals (D2MAC TV standard) transmitted by the TDF1 satellite.

In a first step, the aim of this experiment is to show that time transfers may be realized at the microsecond level all over France with no other information on the satellite position than the position of its parking box. Measurements concerning the three links OP-OB, CT-OB and CT-OP are presented in order to demonstrate this.

In order to reduce the errors introduced by the satellite motion, the three stations calibrated by GPS receivers, may be used to precise the satellite position before realizing the time transfer. This method and a first attempt of restitution of the satellite position are presented here.

The main sources of error on the determination of the time intervals are itemized and their influence on the accuracy of the time transfers are discused.

Introduction

Hertzian TV signals have been used as means of time transfer for about twenty years by common-view of a terrestrial TV transmitter. The development of direct TV using geostationary satellites provides common-view transmitters for more distant stations.

In order to exploit this possibility of time transfer as a complementary system to GPS, a specific receiver using the digital TV standard D2-MAC has been developped at the Observatoire de Besançon (OB) for a time synchronization experiment in cooperation with the CNES in Toulouse (CT) and the LPTF in Paris (OP). These three stations use GPS to compare their clocks and they will serve as reference stations.

The first method of time transfer uses only one reference station, and the accuracy of such comparizons was expected to be of the order of 1 microsecond.

The second method presented here, uses three reference stations in order to specify the satellite position, that is the main source of error in the first method. The aim is to obtain an accuracy of a few ten nanoseconds. The different sources of error that must be taken into account in this case, are enumerated and discussed.

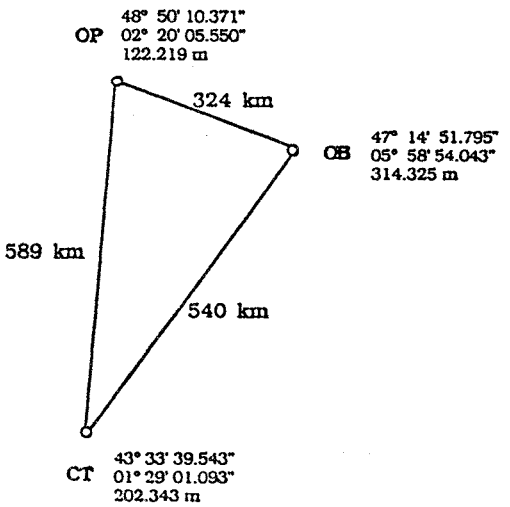


Fig 1 : Length of the different links.

Apparatus - Data acquisition

The receiver developped at OB directly uses the D2-MAC standard signal transmitted by the satellite. Fig. 2 shows the frame of the line 624 ; this line contains analogic and digital reference signals used for alignment, and especially reference levels for "Grey" and "White". That is this transition between the "Grey" and the "White" reference levels that has been choosen to generate the reference pulse issuing from the receiver.

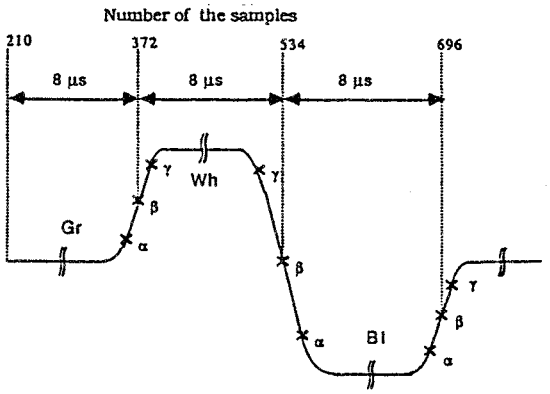


Fig. 2 The frame of the line 624

The principle of data recording is illustrated by fig. 3 : the 1 pps of the clock starts a 1 ns resolution time counter ; the TV signals are received and transmitted to the receiver through a downconverter and a demodulator. The receiver extracts the line 624 and generates the pulse that stops the counter. Data are then sent for processing to a PC computer via a parallel-IEEE interface, and to other stations via a modem.

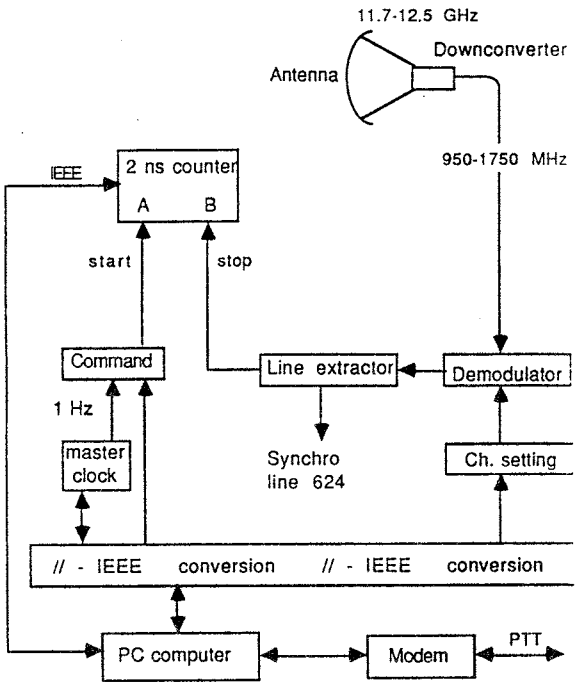


Fig 3 : Data acquisition system

Fig. 4 shows the variations of the time intervals measured at OB during a period of about four days. These variations are a consequence of the satellite motion and especially of the variations of the distance satellite-OB (the time amplitude of about 80 microseconds corresponds to a distance amplitude of about 25 km). The periodicity of 24 hours is clearly appearing ; a sharp change, due to orbital manoeuvres can also be observed.

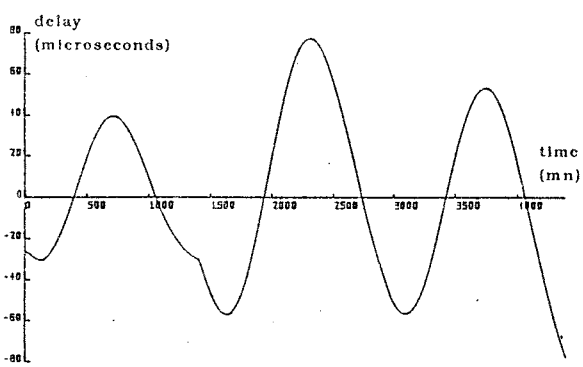


Fig 4 : Examples of time intervals measured at OB

In the operational configuration, measurements consist in one period of fifteen minutes with one measure per minute (16 measures), six times a day (02:00 UT, 06:00 UT, 10:00 UT...). Each group of 16 data is processed by a less square method to give, finally, six points a day.

Since the propagation delays of the TV signals differ according to the position of the reception site, it can happen that the time intervals measured at two different locations do not concern the same TV frame. These data are meaningless and were eliminated. This happens in about 4% of the time for the links CT-OP and CT-OB, and in 0.1% for the link OB-OP.

In order to analyze the noise in the reception system, some sessions were made during which data were recorded continuously every minute or every second by two side by side systems at OB. The time intervals deviations between the two systems are plotted for two different sampling rates : 1 second (fig. 5a) and 1 minute (fig. 5b).

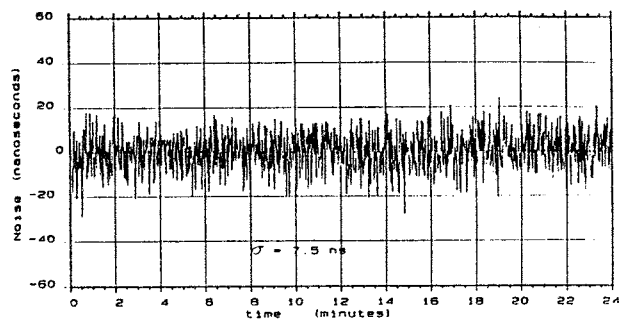


Fig. 5a : Time errors between 2 side by side systems Sampling rate = 1 s

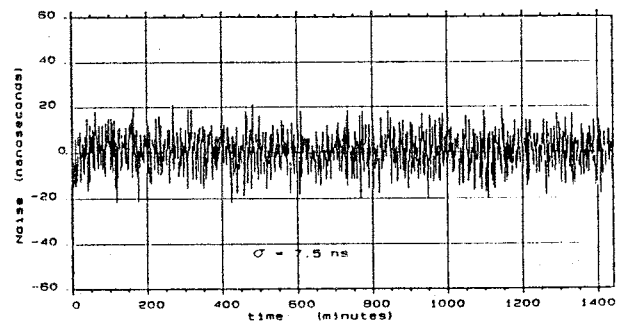


Fig. 5b : Time errors between 2 side by side systems Sampling rate = 1 mn

The standard deviation is the same in the two cases (about 8 ns), what seems to mean that the values are decorrelated and correspond to a white phase noise. This is confirmed by the log-log plots of the spectral density of instantaneous normalized frequency deviations (fig. 6).

Considering the nature of this noise, averaging 16 values means that the contribution of the receivers in the total uncertainty on the determination of the time intervals is reduced to $8 \cdot \frac{1}{\sqrt{16}} = 2$ ns

It has been observed that the signal to noise rate decreased when we reduced the diameter of the dish, indicating that the noise is not only due to the electronic of the receiver. Tests have been made with a larger dish (1.80 m

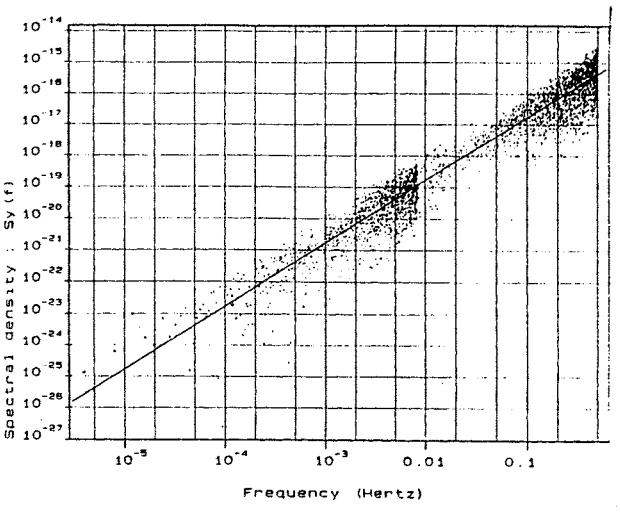


Fig 6 : Spectral density of instantaneous normalized frequency deviations

against 0.70 m) but no significant improvement was noticed.

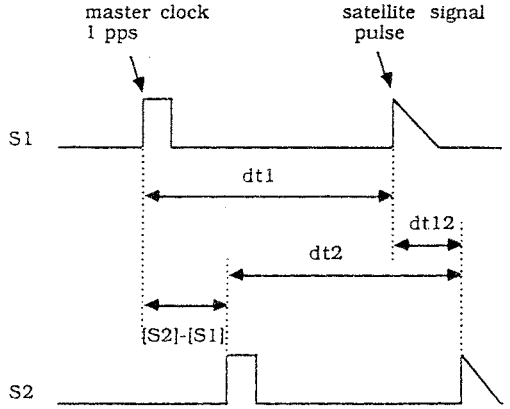
In fact, this noise is far more important than expected (about 200 picoseconds) ; it seems that it is issuing from the demodulator, that is a commercial one, designed for TV and not for clock synchronizations. Further investigations are going on on this subject.

Principle of the time transfer

Considering two stations, the time intervals measured (dt1 and dt2) depend on :

1. the differential propagation delay dt12 between the satellite and each station ;

2. the time difference [S1] - [S2] between the 1 Hz pulse of each clock.



So, the relationship between the time intervals measured, the time difference of the two concerned clocks and the differential propagation delay, can be written as follows :

$$[S1] - [S2] = dt1 + dt12 - dt2$$

This relationship can be considered from two different points of view :

-between 1 reference station and the user station, in order to determine [S1] - [S2] : considering that dt12 is known, the measurement of dt1 and dt2 gives an estimation of dt and realizes, strictly speaking, the time transfer ;

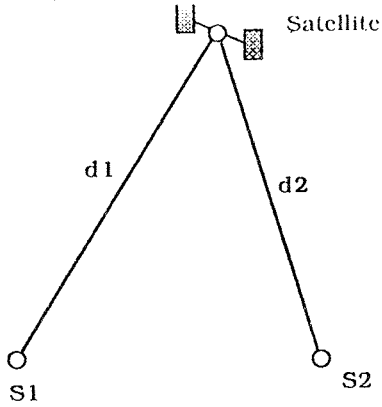
-between 2 reference stations, [S1] - [S2] is known (by GPS for example) : the measurement of dt1 and dt2 determines dt12 and then, gives an indication on the position of the satellite.

Thus, the time transfers can be made either immediately, i. e. with no more information on the satellite than the position of its "parking box", or after having first specified the position of the satellite, with the support of some reference stations, in order to reduce the uncertainty on dt12.

"Immediate" time transfers

If we neglect at the present the effects of different sources of error, the accuracy of the time transfers exclusively depends on the accuracy of the determination of the differential propagation delay dt12, that requires the position of each station and the satellite position.

If the positions of the stations are well known, the main source of error on dt12 is the satellite position since the size of the "parking box" it is maintained in (+- 0.1 degree in latitude and longitude, +-10km in range) generates significant uncertainties :



$$\Delta d = d_2 - d_1$$

The error on Δd can be written as follows :

$$|\delta \Delta d| = |\overrightarrow{\Delta f} \cdot \overrightarrow{\delta P}|$$

where :

$\overrightarrow{\Delta f}$ is the difference of the unit vectors station-satellite and

$\overrightarrow{\delta P}$ is the difference between the center of the parking box and the real position of the satellite.

The size of the parking box (±.1° in longitude and latitude) gives the following values :

$$|\Delta P_x| \leq 10 \text{ km}$$

$$|\Delta P_y| \leq 73.6 \text{ km}$$

$$|\Delta P_z| \leq 73.6 \text{ km}$$

Finally, we obtain the following uncertainties for the three links :

OB - OP : 2.70 µs

CT - OP : 2.95 µs

CT - OB : 2.77 µs

These values are confirmed by the measurements that have been made concerning the links CT-OP, OB-OP and CT-OB (fig 7a, 7b, 7c). In the three cases, the amplitude of the variations is well short of the theoretical value.

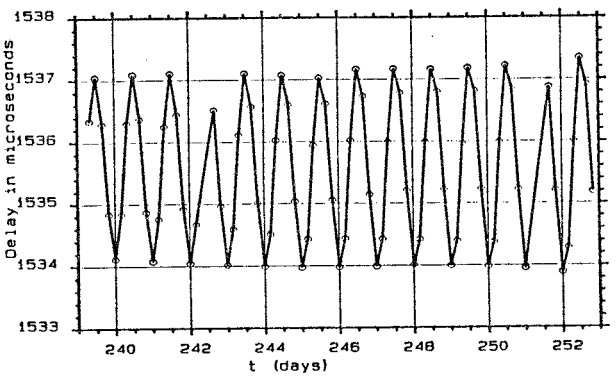


Fig. 7a : Differential propagation delays CT-OP

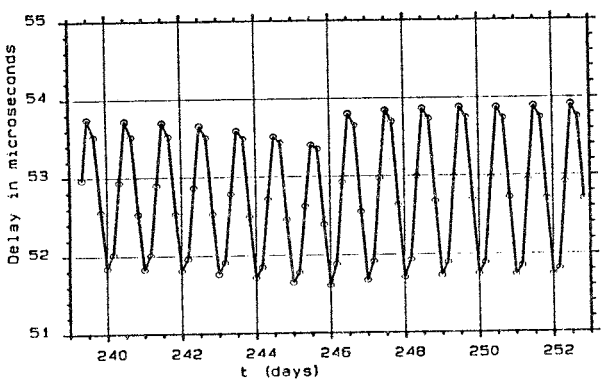


Fig. 7b : Differential propagation delays OB-OP

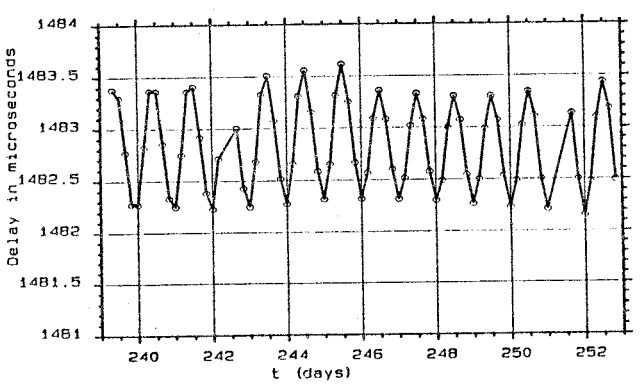


Fig. 7c : Differential propagation delays CT-OB

Fig. 8a, 8b, 8c show an example of the variations of the differential propagation delays after data have been averaged over one day.

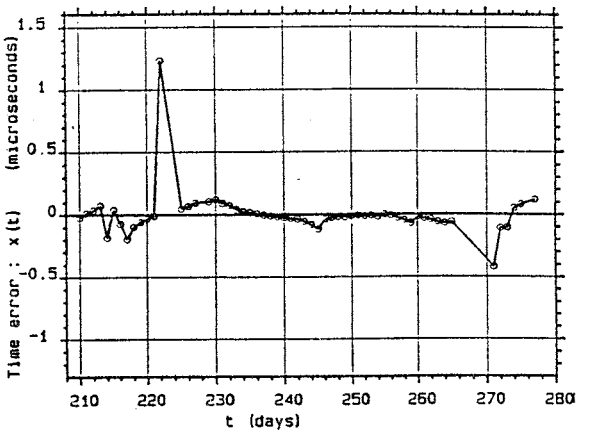


Fig 8a : Averaged data, link CT-OP

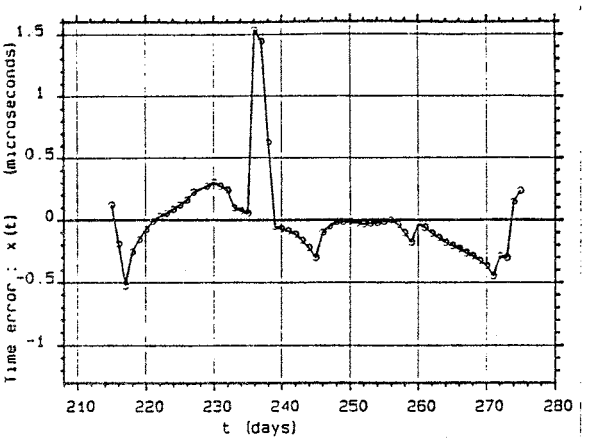


Fig 8b : Averaged data, link OB-OP

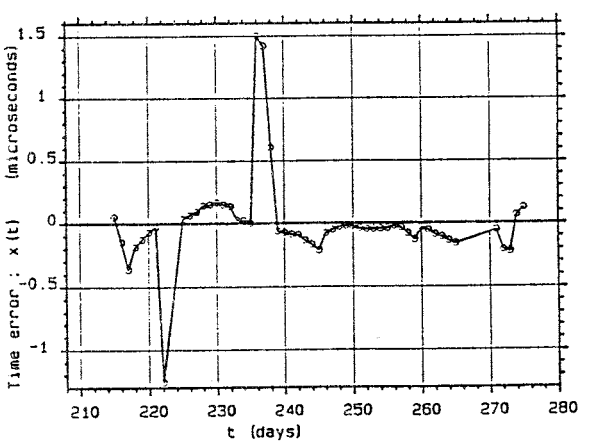


Fig 8c : Averaged data, link CT-OB

Several aberrant points can be seen on these plots and correlated with incidents occuring at one or the other station :

the peak appearing on fig. 8a and 8c at MJD 48222 is due to a fail of the CT receiver. On fig. 8b and 8c the values for MJD 48236 to MJD 48238 are due to the presence of snow on the antenna of OB. (Some other fails were due to a computer failure, a change in the reference channel treated by the receivers...)

After having removed aberrant points each time that the cause of the fail was brought on the fore, we obtain the following plots (fig. 9a, 9b, 9c) :

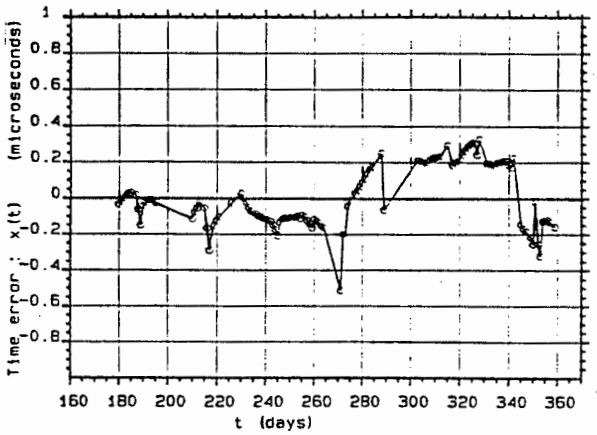


Fig 9a : Averaged data, link CT-OP

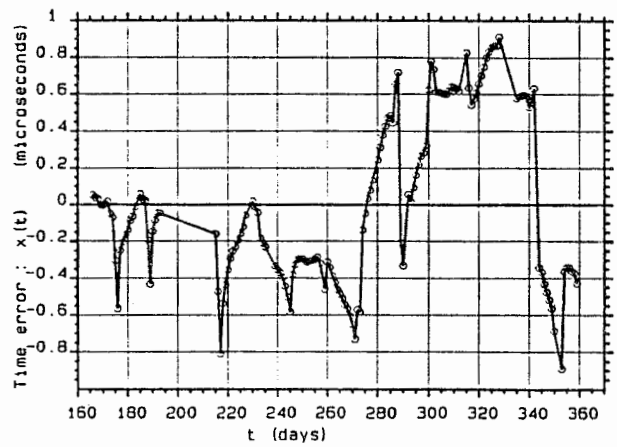


Fig 9b : Averaged data, link OB-OP

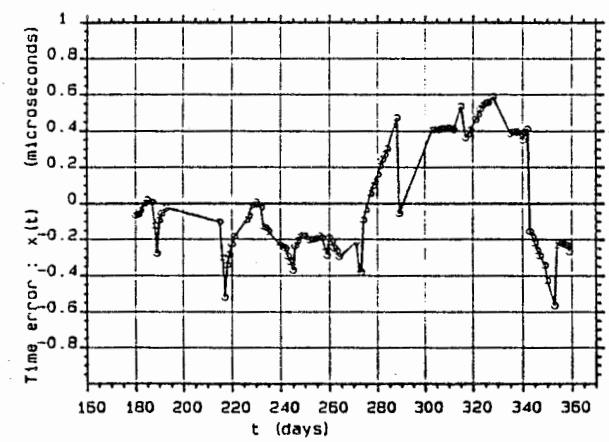


Fig 9c : Averaged data, link CT-OB

According to these plots, we can conclude that each of the three links has a long term stability better than 1 microsecond (0.5 microsecond for the CT-OP link).

Concerning the daily variations of the links, (fig. 10a, 10b, 10c show the absolute values of these variations), if we except some isolated points (further work is made to see whether or not these points are actually due to the satellite motion), we can see they are less than 120 ns (CT-OP), 200 ns (CT-OB) and 350 ns (OB-OP), most of the values being respectively short of 40 ns, 60 ns, 100 ns.

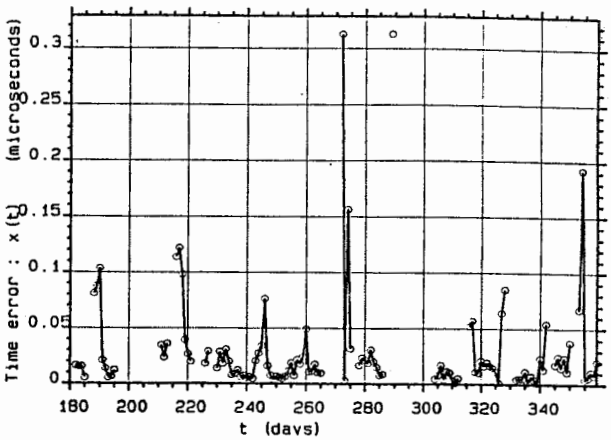


Fig. 10a : Daily variations (absolute values) CT-OP

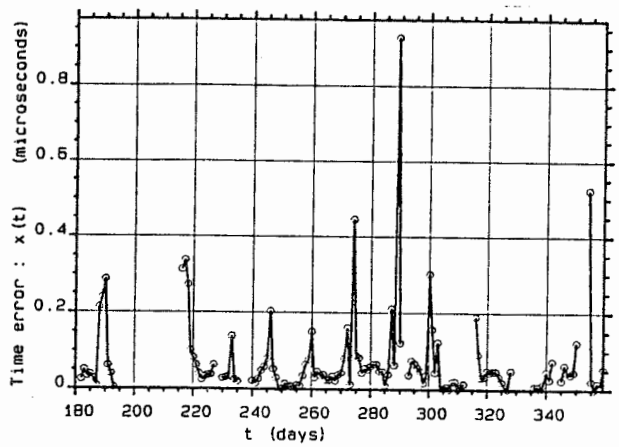


Fig. 10b : Daily variations (absolute values) OB-OP

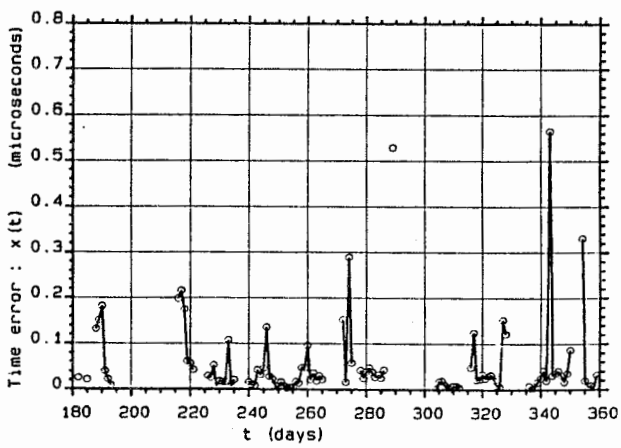


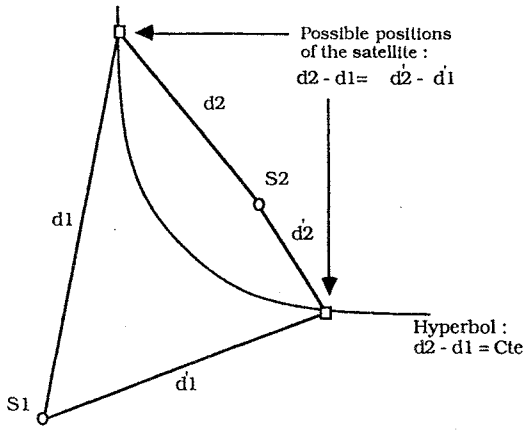
Fig. 10c : Daily variations (absolute values) CT-OB

Three reference stations method

The data collected by the three stations can be used to precise the satellite position if the three master clocks are already compared by an independant way as GPS.

In this case the value of the differential propagation delay measured between 2 stations S1 and S2 correspond

to the difference of the distances |TDF1-S1| and |TDF1-S2|. that means that the satellite must be found on an hyperbol if restrected in the plane (fig. 11), and, in the space, on the hyperboloïd generated by the revolution of this hyperbol around the [S1 S2] axis.



Every hyperboloïd gives a relation that can be written as follows :

$$\frac{X^2}{\frac{d_{ij}^2 - dl_{ij}^2}{4}} + \frac{Y^2}{\frac{d_{ij}^2 - dl_{ij}^2}{4}} - \frac{Z^2}{\frac{dl_{ij}^2}{4}} = -1$$

where :

d_{ij} *is the distance between the two stations* S_i *and* S_j

dl_{ij} *is the length of the differential path.*

X, Y, Z are the coordinates of the satellite.

Fig. 11 : Definition of the hyperboloïds.

The use of a third station provides a second hyperboloïd and situes the satellite on the curve of intersection of the two hyperboloïds.

As the data used to determine that curve are differential delays, they do not provide any information on the satellite altitude, i. e. the curve, that is quasi linear at the level of the satellite parking box, is oriented in direction of the reference stations.

The maximum uncertainty generated among the curve (for an altitude variation of +- 10km) is below 10 ns for every link within France.

In this method, as the desired level of precision is a few ten nanoseconds, the uncertainty on the differential propagation delays (GPS : σ=3ns, dispersion in the reception system : σ=2ns after averaging 16 values, uncertainty on the tropospheric corrections) must be taken into account. Concerning the tropospheric corrections, the errors on the models worked out by the CNES do not exceed 1 ns. This gives a global uncertainty of about 4 ns and generates the following errors on the determination of the satellite position :

longitude : +- 0.002 degrees (1.5 km)
latitude : +- 0.004 degrees (3.9 km)

Assuming these errors, we may expect the accuracy of the time transfers to be of the order of 10 ns in the most of the cases (stations located inside the reference triangle or near one of the reference stations), up to 30 ns in the worst case (farest stations).

As an illustration of this method, we present here a first attempt of restitution of the satellite position : on the two plots below, each rectangle represents the calculated satellite position, assuming a 5 nanosecond uncertainty on the time intervals measured. Each of the six groups of rectangles correspond to one of the six data acquisition periods (02:00 UT, 06:00 UT, 10:00 UT...). So, inside each group, the daily drift of the satellite may be observed.

Fig. 12a shows the satellite motion between MJD 48239 and MJD 48245, (just before the orbital manoeuvres that are also observable on fig. 8, 9 and 10 at MJD 48245, the daily drift is clearly increasing) while fig. 12b between MJD 48245 and MJD 48252 shows this motion after the manoeuvres (the drift is quite reduced).

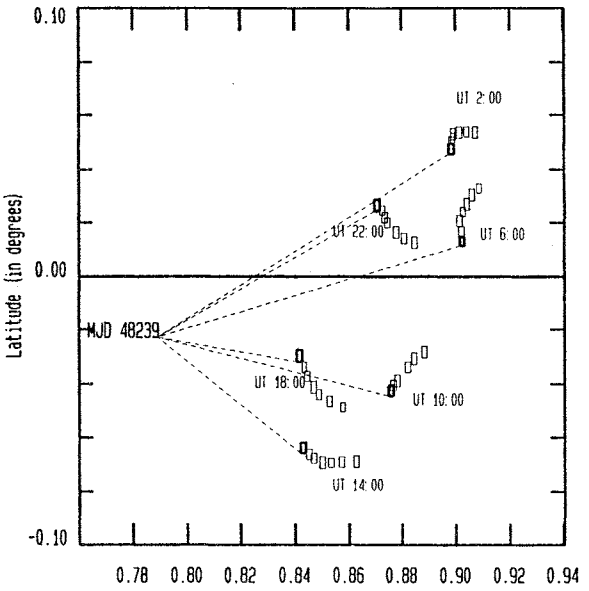


Fig. 12a : before the orbital manoeuvres.

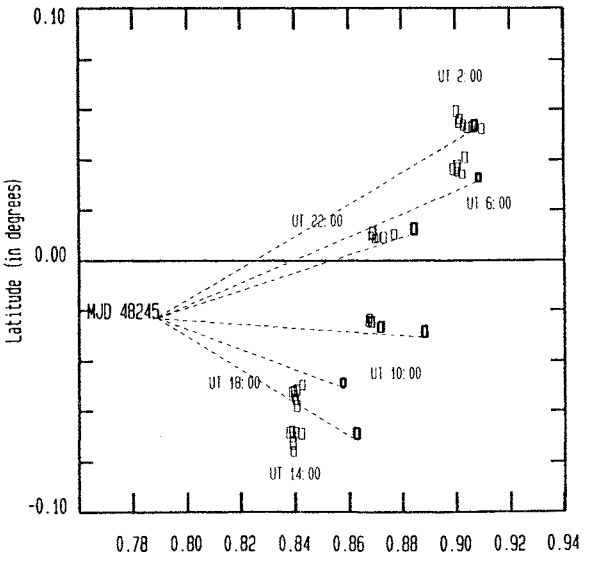


Fig. 12b : after the orbital manoeuvres.

Conclusions

Concerning time transfers from only one reference station, it has been shown that their long term stability is better than 1 microsecond for each of the three links that have been tested, while the daily variations can be estimated to be short of 100 ns.

As most of the stations within France can be linked by the use of shorter paths (by choosing the nearest of the three stations OP, OB, CT as a reference station), we can expect the uncertainty of the clock-comparisons provided by this method to be below these values for every transfer over France. This will be prooved by a precise estimation of the influence of the position of the station to be linked relative to the reference station position.

In order to verify the results obtained by the use of three reference stations, a fourth station is about to be put into service at the Observatoire de la Côte d'Azur in Grasse.

Also, this will provide a possibility for calculating the satellite position from 4 reference stations, i. e. three hyperboloïds, what theoretically gives the precise position of the satellite. In fact, since the hyperboloïds are very close together relative to the altitude of the satellite, a small error in the time intervals produces large errors on the calculated position, particularly on the altitude. Nevertheless, simulations have shown that, by this way, the uncertainty on the time transfers could be divided by 2.

The calculation of the parameters of the satellite orbit from three, then four stations will also be studied. If needed, the possibility of installing a two-way station at the satellite control station (in order to obtain a accurate information on the altitude of the satellite) will be envisaged. In this case, the precision of the calculated satellite position could allow the uncertainty on the time transfers to be less than 10 ns.

Acknowledgements

The authors wish to thank B. Mougin for his work of collecting and filing data.

References

O. Buzek, j. Cermak, j. Vondrak, F. Cordara, P.G. Galliano, V. Pettiti, P. Tavella : Synchronization of time scales by television method using ECS satellites. Proceedings of the Third EFTF, Besançon 1989 pp 204, 214.

S. Starker, H. Nau, J. Wilner, G.V. Kinal : Geostationary navigation satellites as a means of time transfer. Proceedings of the fourth EFTF, Neuchâtel 1990 pp 313, 319.

P.R. Pearce, J.A. Davis, D.R. Allen, B.R. Swabey, R. Jefferies, D.J.E. Knight, P. Menary : Time transfer in the United Kingdom via a direct broadcasting satellite. Proceedings of the fourth EFTF, Neuchâtel, 1990 pp 643,650.

Session 2B Number 16

A Direct Sequence Spread-Spectrum Modem Design for Time Transfers via Geostationary Satellites

D. D. Davis and D. A. Howe
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

Abstract

The National Institute of Standards and Technology (NIST) is developing a direct sequence spread spectrum (DSSS) modem. This development is part of a program to do various time transfer experiments using geostationary satellites. The purpose of the program is to exploit low-cost present and future commercial satellite resources in order to achieve timing accuracy of a few nanoseconds or better in a two-way time transfer scheme.

The modem's technical specifications will be presented. The modem has the capability for variable PN sequence chip rate from 102.3 kchips/s to 2.5 Mchips/s. This is because many aspects of time transfer experiments are affected by the DSSS chip rate, for example, occupied bandwidth and power (hence, satellite transponder cost), resolution, accuracy, processing gain, and time to acquire lock. Two types of codes can be selected: (1)Gold code of length 1023 bits, and (2)truncated maximal length sequence (MLS) of length 10,000 bits.

Functionally the modem operates with an AT-compatible computer. The computer is used to (1)select the modem's operating configuration, (2)send and receive data (using the modem's communications capability) to another earth station, and (3)provide a convenient resource for scheduling time transfers, computing the involved stations' clock-difference in real-time, and storing results.

Introduction

The Time and Frequency Division of the National Institute of Standards and Technology continues to do research in satellite-based time synchronization. Such systems provide higher accuracy than HF and VLF methods since signal path delay is better known and delay instability is lower. In previous years, time transfer experiments were designed around a fairly limited number of space-segment resources. However, today a multitude of satellite communications opportunities exist. This writing will specify modulation-demodulation (or "modem") signal parameters and describe development of equipment at NIST which is suited to present-day satellite resources for two-way time transfers. A large number of references exist on the subject of the two-way technique, however the direct-sequence spread-spectrum modem called Mitrex* has been used successfully in experiments since 1983 (1). The Mitrex experiences together with results from the military's Global Positioning System will serve as a starting point for specifying aspects of two-way modem characteristics.

The Direct Sequence Spread-Spectrum Technique

Most of what follows will show that a spread-spectrum signal modulation represents good use of the resources in geostationary communications satellites, the type used in most 2-way satellite time transfers (2, 3). Figure 1 shows where a spread-spectrum modem is used in a typical satellite earth station which is set up to do two-way time transfers. There are three general categories of spread-spectrum techniques: (1) direct-sequence, (2) frequency-hop, and (3) chirp. There are also hybrids or combinations of these three techniques which can be used to take advantage of specific properties.

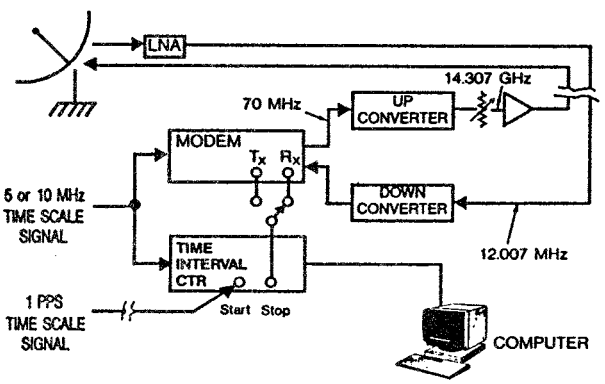


Fig. 1. Generalized equipment scheme used in two-way time transfers which shows the use of the DSSS modem.

Spread-spectrum signals using direct-sequence modulation provide the highest resolution range measurements, hence the highest resolution in timing. This property is due to the high speed of the PN sequence representing "1" and its inverse "0". Synchronization to the sequence at a receiver must be to within one chip (one bit in the sequence) in order to decode the overlying data. Typically the synchronization is good to at least one percent of one chip. Therefore, the inherent timing resolution between a transmitter and receiver is better than the time period corresponding to one-percent of a chip (4). The relatively high speed of the PN sequence translates into a short period for one chip, hence better inherent timing resolution. The ability to resolve time at a receiver is key to establishing improved levels of precision and accuracy. Experience with the Mitrex modem operating at the chipping rate of 2.5 Mchips/s indicates long-term reproducibility to ± 1 ns (5). This suggests that the Mitrex scheme has a timing uncertainty of 0.5% of the

*Identification of a commercial company does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any identified entity is the only or the best available for the purpose.

This is a contribution of the U. S. Government and is not subject to copyright.

chip length. Furthermore, a NIST-designed GPS receiver using a DSSS format at 1.023 Mchips/s (the C/A code) has also shown reproducibility to ± 2 ns which is 0.4% of the chip length (6).

Type of PN Code

One of the most important aspects of the two-way timing system described here concerns the specification of the PN code. The type of code, its length, and its chip rate determine limitations of timing performance. Many aspects of these timing systems are affected by chip rate, for example, occupied bandwidth and power (hence transponder cost), resolution, accuracy, acquisition of lock time, and processing gain. For this reason, the modem will have a variable chip rate with a range which is compatible with commercial satellite resources and which uses these resources effectively. The positive experience gained with GPS points to specifying the same or similar type of code (Gold) and length (1023 bits). Important properties of good codes of length L are: (1) codes should have a balance of "ones" and "zeros," (2) autocorrelation function is two-valued with difference of 1/L, (3) cross correlation function from dissimilar codes is below $1/\sqrt{L}$ relative to the peak in the autocorrelation function, and (4) there are enough different codes to satisfy present and future network capacity.

There appears to be no overriding reason not to use Gold codes, the other prevalent type being maximal length sequence (MLS). Truncated MLS codes such as used in the Mitrex modem have the small advantage of hardware simplicity if an even power-of-ten number of chips occur over a convenient time interval, in this case 10,000 bits every 4 ms. Alternatively, a Gold code of length 1023 requires that a signal be generated which can deliver 1023 bits over some convenient (and variable as mentioned above) interval of time. The Mitrex modem has achieved widespread use by time standards laboratories which are experimenting with two-way time transfers via satellite. For reasons of compatibility, it would be desirable to include an option to do two-way time exchanges with the Mitrex modem.

There is a limitation of six MLS codes which meet the requirements described above (7). This may be adequate for present needs but is restrictive in terms of future needs. On the other hand, Gold codes are sets of linear nonmaximal codes generated by a combination of linear maximal codes of equal length. Where a large number of codes is required, Gold codes are preferred. In the case of code length 1023 bits, there are 1024 separate possible codes and in all likelihood more than six with particularly good properties. GPS uses thirty-six Gold codes generated from two MLS codes.

One should note that the GPS Gold code of length L = 1023 is almost ten times shorter than the truncated MLS code of length L = 10,000. This will result in poorer cross-correlation performance and lower processing gain for the GPS Gold compared to the truncated MLS codes. Accuracy limits of both types should be similar at similar chip rates if precautions are followed which reduce the possibility of an accuracy error due to code interference. The use of different codes to carry different information on the same frequency is called code division multiple access (CDMA). Two-way time transfer requires that two different sets of data be exchanged. Operation at the same frequency is desirable so CDMA is used (2, 3). If there is no phase coherence among the DSSS signals which combine in the CDMA scheme, then there will be no accuracy error in a DSSS timing signal (8). This is the main precaution to be observed.

Equipment Used in NIST Design

The NIST spread-spectrum modem works in conjunction with an AT-compatible computer. Precise timing, RF and modulation, and signal-structure aspects are in electronics outside the AT-compatible computer. Although the modem cannot work without a PC-AT interconnection, we have commonly referred to these external electronics as the "the modem." As such, the modem consists of three main printed-circuit boards: (1) a transmit board (called Tx) consisting of RF, analog, PN generator, and digital control circuitry, (2) an analog receive board (called Rx-analog) having RF and other analog functions, and (3) a digital receive board (called Rx-digital) that contains the PN generator and other digital functions.

Figure 2 shows the scheme of the equipment. A 70 MHz intermediate frequency is used as the input to the Rx portion of the modem; 70 MHz is also used as the output from the Tx portion. This frequency is commonly used for most upconverters and downconverters incorporated in satellite earth stations. Some frequency agility around 70 MHz is required to take advantage of low cost earth terminals called very small aperture terminals (VSAT) which are becoming more readily available from various manufacturers. NIST has done a number of experiments using VSAT apparatus, and results show that they can be used effectively in 2-way satellite timing (9).

The NIST designed modem automatically makes the difference between the external reference 1 pulse/s and Tx (1 pulse/s as transmitted) equal to zero ± 0.1 ns. This is done by internally adjusting the phase of the 5 or 10 MHz reference feeding the PN generator on the Tx board. The phase is adjusted using a slewable divider and is made to coincide with the external 1 pulse/s reference signal. (In the hardware implementation, the PN sequences are contained in EPROM so there are no actual MLS registers.)

Tx Board

Figure 3 shows a block diagram of the modem Tx board. The electronic functions are indicated along with pertinent control functions (or interface lines).

The reference frequency signal usually originates from a frequency standard's 5 or 10 MHz output. The input of the modem is selectable between 5 and 10 MHz. A 10 MHz reference for remaining functions of the Tx board originates from a phase microstepped VCXO (voltage controlled crystal oscillator) which is locked to the external reference frequency. The internal 10 MHz signal can be advanced or retarded in 100 ps increments, and its purpose is to put the 1 pulse/s inverted sequence "on time" (see previous section).

A 70 MHz synthesizer is used to generate the Tx carrier. This synthesizer has a range of ± 500 kHz in 10 kHz increments. The variable frequency function is necessary to accommodate low-cost VSAT RF earth stations. These terminals often synthesize a microwave transmit signal which is close enough to the desired transmit frequency to meet regulatory requirements but which is not the exact assigned frequency. Tx frequency agility in the modem will compensate for any frequency offset which is present using low-cost VSAT's. A second use for this frequency agility is to compensate for any error in the usual 2.3 GHz frequency offset between receiving and retransmitting through the satellite transponder. This function would be necessary if the receiving earth station had no receive frequency agility of its own.

The PN generator has its clock rate (or chip rate) determined from a variable modulo divider with 10 MHz input

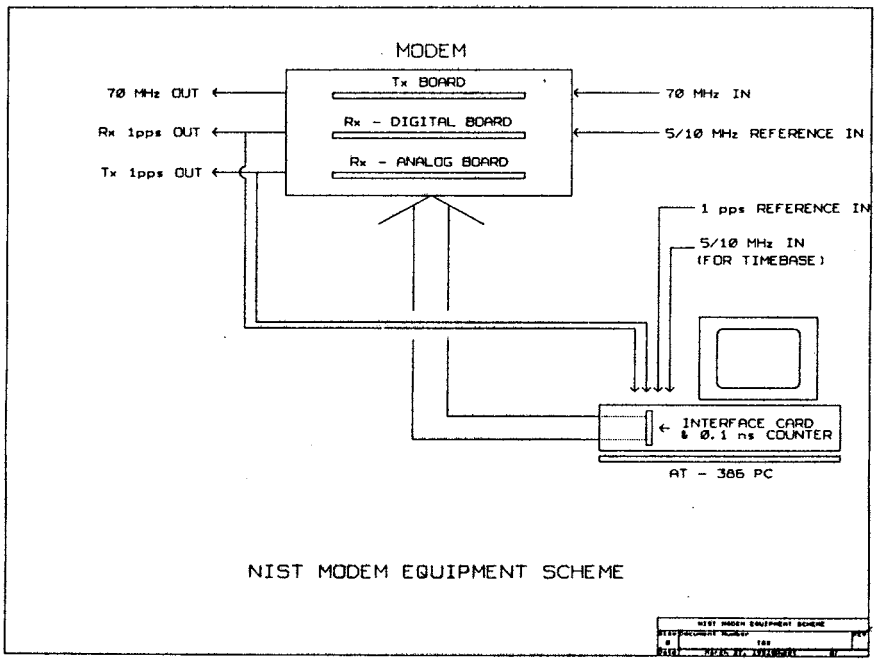


Fig. 2. The modem consists of three printed-circuit boards and a time-interval counter and interface card which is used in a 80386 computer.

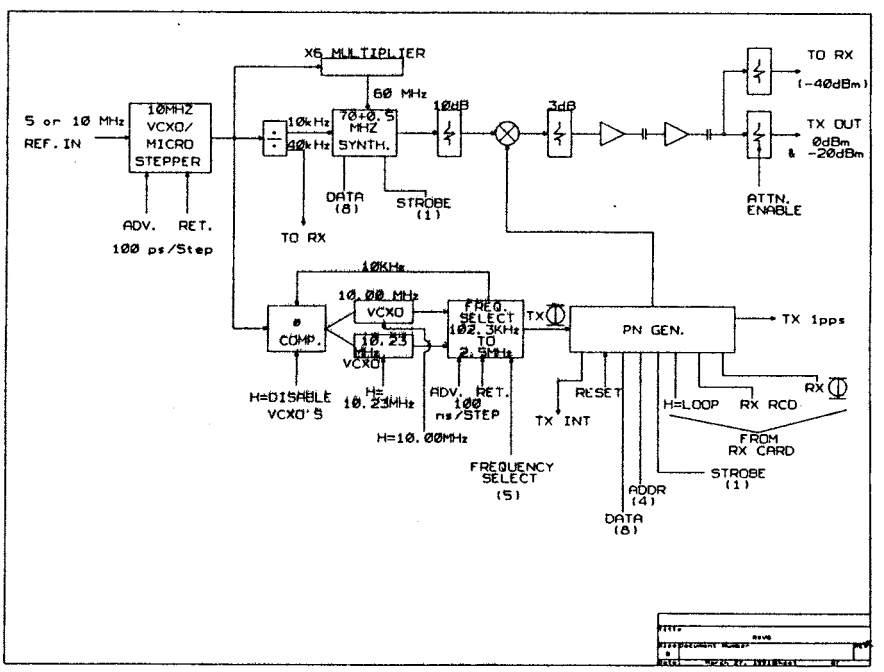


Fig. 3. Block diagram of the Tx portion of the modem. It consists of a 70 MHz RF generator, PN sequence generator, and bi-phase modulator.

and generating an output over the range of 102.3 kHz to 2.5 MHz (corresponding to 102.3 kchips/s to 2.5 Mchips/s). One of two VCXO's can be enabled as the source for the variable divider, either 10.23 MHz (for Gold sequences) or 10.00 MHz (for truncated MLS sequences). The phase of the PN clock can be incremented in 100 ps steps.

There are 36 Gold code sequences and eight truncated MLS sequences which can be selected. The Gold codes are identical to those used in the GPS C/A code and the chip rate is variable between 102.3 kchips/s to 2.046 Mchips/s. The eight truncated MLS sequences are compatible with the Mitrex modem and have a fixed rate of 2.5 Mchips/s. All sequences are programmed into a 64K × 8 EPROM IC.

Since the PN sequence chip rate is variable from 102.3 kchips/s to 2.5 Mchips/s, the frame rate is also varied (frame rate is chip rate divided by number of chips in the PN sequence). The 1 pulse/s modulation is simply a 1 bit/s data transmission. The rest of the 1 s period is normally occupied by the unaltered PN sequence. With no significant degradation, one can transmit communications data between 1 pulse/s bits as long as the PN sequence is still tracked in the demodulation and the 1 pulse/s bit is properly identified with a unique time frame marker. Data transmission capability has been included in this modem, and the rate of the communications data capability varies from 10 bits/s at 102.3 kchips/s to 50 bits/s at maximum chip rate.

The PN sequence (code) can be changed without perturbing the phase of the sequence. Two or more codes can be transmitted sequentially to construct additional different codes whose length is a multiple of 1023 chips for Gold codes and 10,000 chips for truncated MLS codes.

Table 1 summarizes the modulator specifications. Reference input and output timing pulses are TTL-compatible.

Rx Boards

The demodulator portion of the modem consists of two printed circuit boards--the Rx-analog and Rx-digital boards. As the names imply, the Rx-analog board contains analog circuitry such as the RF amplifiers, mixers, and filters. The Rx-digital board contains digital circuitry such as the PN generator, synthesizers for L.O. (local oscillator) frequencies, and several digital control functions.

Figure 4 shows a block diagram of the demodulator. From the 70 MHz source (usually the output of a microwave downconverter), the signal is first fed through a reed relay which can select the output of the Tx board for an internal modem loop test. After going through a switchable 20 dB attenuator, the signal is split and amplified into two signal channels, one is designated the "punctual" (P) channel, the other is designated the "early-minus-late" (E-L) channel. the demodulator's PN generator has two outputs, one which is the P (punctual), or on-time, PN sequence and the other which is the E-L PN sequence.

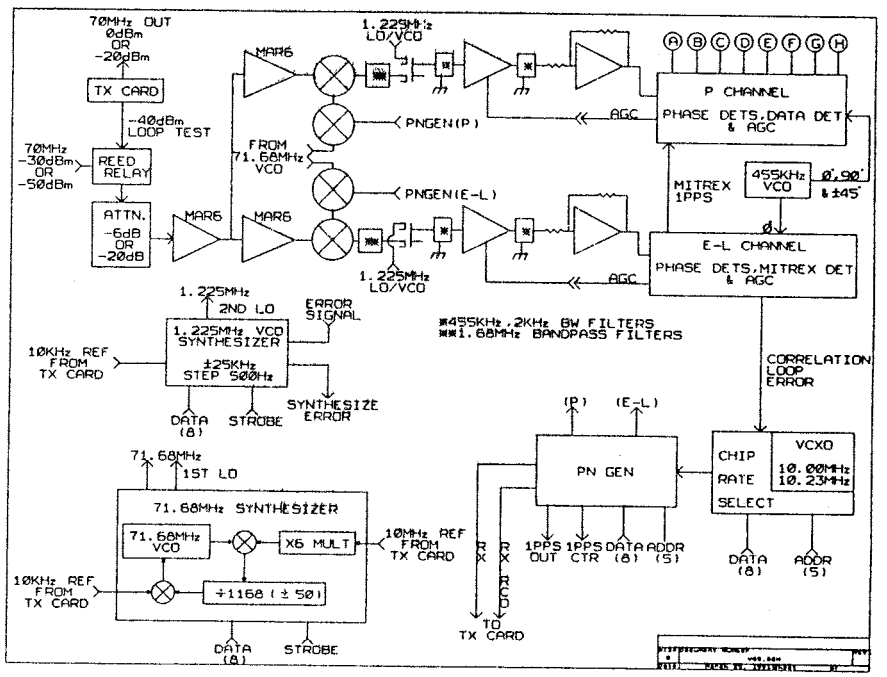


Fig. 4. Block diagram of the Rx portion of the modem. It consists of two separate triple conversion receiver sections, one for punctual (P) correlation and one for early-minus-late (E-L) correlation. The purpose and action of each are described in the text.

Two isolated 71.68 MHz signals are generated from a synthesizer. The "P" PN sequence is mixed with one 71.68 MHz signal and then mixed with the "P" RF channel's signal. The "E-L" sequence is mixed with the other 71.68 MHz signal and then mixed with the "E-L" RF channel signal in a similar way.

A 1.225 MHz signal from another synthesizer is mixed with the newly created 1.68 MHz difference frequency of the P and E-L channels. This second frequency conversion produces a difference frequency of 455 kHz in both channels which is filtered and amplified.

A 455 kHz VCO (referred to 40 kHz from the 5 or 10 MHz external signal) generates signals at four phases 0°, 90°, and ±45°. These signals are variously phase-compared to the 455 kHz IF signal of the P channel to track the RF carrier, demodulate the communications data, and indicate carrier lock (phase-coherent tracking with the feedback error signal applied to the 1.225 MHz L.O. synthesizer is possible over a frequency range of ±25 kHz). A 455 kHz VCO signal at 0° phase is used as the reference for the phase detector of the E-L channel. The output of this phase detector represents the error signal from cross-correlation of the received PN sequence with the PN sequence as generated by the Rx (demodulator's) PN generator. This error signal advances the phase of the Rx PN generator until it is "on time." That is, the cross-correlation to the received PN sequence is unity.

Table 2 summarizes the demodulator specifications. An output 1 pulse/s timing signal is available which is TTL compatible. The time difference between the external 1 pulse/s reference and the 1 pulse/s as received (demodulated by the modem) can be measured using a time-interval counter which is integrated with the computer interface card.

AT-Compatible Computer and Interface and Counter Board

The modem is controlled by an AT-compatible personal computer with 80386 CPU. A 386 AT-compatible machine was chosen for several reasons:

1. Open architecture is flexible and adequate to handle control data, data saving, calculations, and interfacing to other equipment.

2. AT-compatible electronics are readily available in a variety of physical sizes and arrangements.

3. Consumer demand and competition in the PC industry have brought prices down to where it is inexpensive.

4. A large base of software and programmers make software development more efficient.

In addition, the 80386-based computers have the advantage of a 32-bit architecture with an associated large addressable memory space. The 80386 CPU also has a larger, more useful instruction set compared to any of its predecessors.

The hardware consists of the computer and hard disk storage with keyboard, monitor, and a plug-in interface card to the modem. The card also includes the necessary circuitry to measure time intervals to a resolution of one-tenth nanoseconds using the same 5 or 10 MHz external reference as the modem itself. An interpolator measures the occurrence of the normal start and stop trigger pulses relative to the zero-crossings of the external reference signal. The counter readings are latched and sent to an interface portion of the card which allows the counts to be read by the AT-compatible buss.

The counter has a self-calibration feature for the interpolation circuit which establishes accuracy to better than ±1 ns. This calibration is done automatically (thru associated AT software) and can be done as often as once every few minutes. Reproducibility of the counter readings is ±0.1 ns over a period of 30 min or more under normal operating conditions.

Under software control, the modem is configurable in any mode as described in this writing. Time-interval counter data and communications data can be read, stored, and analyzed, and data can be transmitted via a telephone modem to another location.

Direction of Development

As of this writing, the Tx board, Rx boards, and computer interface card are complete. The software development is an on-going task and is a specification only. Software control adequate for testing the modem is expected to be completed by summer of 1991. DOS is the operating system and code will be written in C language with a limited number of assembly routines.

The versatility of modem control by means of a 80386 AT-compatible computer offers many future options as this modem project proceeds. As such, much of the work will move to development at the software level as the hardware is specified, completed, and debugged. We plan to integrate other existing computer programs as tools and extensions of the computer into other tasks such as earth station automation.

Acknowledgement

The authors gratefully acknowledge the work of John Lowe, Xiao Cuiying, and Victor Zhang in development and debugging much of the hardware.

References

1. P. Hartl et al., "High accuracy global time transfer via geosynchronous telecommunications satellites with Mitrex," Z. Flugwiss., Welttraumforschung, vol. 7, pp. 335-342, 1983. Mitrex is a brand name of the Institute for Navigation, Germany.
2. D. W. Hanson, "Fundamentals of two-way time transfers by satellite ,"Proc. of 43rd Annual Symposium on Frequency Control, Denver, CO, June, 1989, pp. 174-178.
3. D. Kirchner and W. Riedler, "Timing by satellite: Methods, recent developments, and future experiments," in Proc. Int. Symp. Satellite Transmissions, Graz, Austria, Sept. 1985, ESA SP-245.
4. J. K. Holmes, Coherent Spread-Spectrum Systems. New York: Wiley, 1982; see also R. C. Dixon, Spread Spectrum Systems. New York: Wiley, 1975, pp. 177-205.
5. D. A. Howe, "Progress toward one-nanosecond two-way time transfer accuracy using Ku-band geostationary satellites," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, No. 6, Nov., 1987.
6. D. W. Allan, D. D. Davis, M. Weiss, A. Clements, B. Guinot, M. Granvesud, K. Dorenwendt, B. Fischer, P. Hetzel, S. Aoki, M.-K. Fujimoto, L. Charron, and N. Ashby, "Accuracy of international time and frequency comparisons via global instrumentation and measurement, IM-34, no. 2, 118-125, 1985.
7. R. C. Dixon, P. O. Box 100, Palmer Lake, CO 80133, private communications.
8. D. A. Howe, "Time tracking error in direct-sequence spread-spectrum networks due to coherence among

signals," IEEE Transactions on Communications, Vol. 38, No. 12, December 1990, pp. 2103-2105.

9. D. A. Howe, "Ku-band satellite two-way timing using a very small aperture terminal (VSAT)," in Proc. 41st Annual Symposium on Frequency Control, Philadelphia, PA, May 1987, pp. 149-160.

Table 1

SPREAD SPECTRUM MODULATOR SPECIFICATIONS

Frequency (IF frequency) and source impedance of device:
70 MHz coherent with reference input frequency. 50 ohms source impedance.

Output power level:	Selectable -20 dBm and 0 dBm.	
Code modulation method:	BPSK	
Types of codes:	1. Gold	
	2. Truncated MLS	
Number of Gold codes:	No less than 36	
Number of Truncated MLS codes:	8*	
Gold code (chip) rate:	Selectable	2.046 Mchips/s
		1.023 Mchips/s
		409.2 kchips/s
		204.6 kchips/s
		102.3 kchips/s
Truncated MLS code (chip) rate:	2.5 Mchips/s*	
Gold code (bit length:	1023 chips	
Truncated MLS code (bit) length:	10,000 chips	
Gold code bit rate:	Variable	2000 bits/s at 2.046 Mchips/s
		1000 bits/s at 1.023 Mchips/s
		400 bits/s at 409.2 kchips/s
		200 bits/s at 204.6 kchips/s
Truncated MLS code bit rate:		100 bits/s at 2.5 Mchips/s
Ref. input frequency signal:	Selectable	5 MHz
		10 MHz

Accepts sine or square wave signal and terminates at 50 ohms with a sensitivity of -10 dBm.

Ref. input time signal (must be coherent with ref. input freq.):
1 pulse/s
Positive-going trigger. Threshold adjustable from +0.5 to +2.0V.

Ref. output time signal (as transmitted) : 1 pulse/s

Time on positive-going edge of at least +3.5 VDC pulse relative to system ground with a source impedance of 50 ohms and duration of one chip.

Communications date rate (Gold code):	10 bits/s at 102.3 kchips/s
	20 bits/s at 204.6 kchips/s
	40 bits/s at 409.2 kchips/s
	50 bits/s at 1.023 Mchips/s
	50 bits/s at 2.046 Mchips/s
Communications data rate (trunc. MLS) :	50 bits/s at 2.5 Mchips/s

An internal slewable divider will be incorporated which divides the external 5 or 10 MHz reference signal to a signal which represents 1 pulse/s as transmitted. The divider will be automatically slewed so that the 1 pulse/s as transmitted is within 100 ps of the external reference 1 pulse/s time signal.

---

*Duplicates format of Mitrex model 25000

Table 2

SPREAD SPECTRUM DEMODULATOR SPECIFICATIONS

Freq. of operation (IF freq.): 70 MHz automatically tunable over ±25 kHz. (Automatically acquires correct "center" of freq.) Terminated at 50 ohms.		
Input level sensitivity: Adjustable -60 dBm to 0 dBm	max. sensitivity min. sensitivity	
Types of codes (same as modulator):		1. Gold 2. Truncated MLS
Number of Gold codes:		No less than 36
Number of Truncated MLS codes:		Eight*
Gold code (chip) rate:	Selectable	2.046 Mchips/s 1.023 Mchips/s 409.2 kchips/s 204.6 kchips/s 102.3 kchips/s
Truncated MLS code (chip) rate:	2.5 Mchips/s*	
Gold code (bit) length:	1023 chips	
Truncated MLS code (bit) length:	10,000 chips	
Gold code bit rate:	Variable	2000 bits/s at 2.046 Mchips/s 1000 bits/s at 1.023 Mchips/s 400 bits/s at 409.2 kchips/s 200 bits/s at 204.6 kchips/s 100 bits/s at 102.3 kchips/s
Truncated MLS code bit rate:		250 bits/s at 2.5 Mchips/s
Ref. input frequency (for lock acquisition):		5 MHz or 10 MHz.
Accepts sine or square wave signal and terminates at 50 ohms with a sensitivity of -10 dBm.		
Ref. output time signal (as received):		1 pulse/s
Timed on positive-going edge of at least +3.5 VDC pulse relative to system ground with a source impedance of 50 ohms and duration of one chip.		
Communications data rate (Gold code):		10 bits/s at 102.3 kchips/s 20 bits/s at 204.6 kchips/s 40 bits/s at 409.2 kchips/s 50 bits/s at 1.023 Mchips/s 50 bits/s at 2.046 Mchips/s
Communications data rate (trunc. MLS):		50 bits/s at 2.5 Mchips/s

---

* Duplicates format of Mitrex model 25000.

Session 2B Number 17

PRELIMINARY COMPARISON OF TIME TRANSFERS VIA LASSO, GPS AND TWO-WAY SATELLITE.

P.Uhrich, R.Tourde, M.Granveaud.
Laboratoire Primaire du Temps et des Fréquences, Paris, France.

P.Grudler, F.Baumont, Ch.Veillet, D.Féraudy, J.M.Torre, J.F.Mangin, J.Gaignebet, J.L.Hatat.
Observatoire de la Côte d'Azur, Grasse, France.

W.Hanson, A.Clements, J.Jespersen, M.Lombardi.
National Institute of Standards and Technology, Boulder, USA.

D.Kirchner, U.Thyr.
Technische Universität, Graz, Austria.

G.Kirchner, H.Ressler.
Institut für Weltraumforschung, Graz, Austria.

W.J.Klepczynski, P.J.Wheeler, W.Powell, A.Davis.
U.S. Naval Observatory, Washington, USA.

W.Lewandowski.
Bureau International des Poids et Mesures, Paris, France.

Abstract.

Presently, the three primary techniques for achieving nanosecond time transfer are Common-View GPS, Two-Way Satellite Time Transfer and the Laser Ranging Technique of the LASSO Experiment. Common-View GPS offers an economical way to achieve near-nanosecond time transfer accuracy using signals from the satellites not affected by Selective Availability. Two-Way Satellite Time Transfers are capable of 100 ps precision and can be calibrated to the nanosecond level to remove systematic errors. The laser ranging technique promises to be the most intrinsically accurate technique because systematic delays can be accurately measured by means of terrestrial target and differential delays between stations can be measured by a transportable calibration station.

Trough the cooperative efforts of several international laboratories, all the three methods were implemented at two stations within Europe, CERGA and TUG. At each of these two stations, a common reference clock to all three systems allows the direct intercomparison of these three state-of-the-art techniques. This paper will discuss the operations at these two stations, present the available data and an intercomparison of it.

Introduction.

The loan of an US portable Two-Way Satellite Time Transfer station to a European laboratory was discussed for the first time during a meeting held in Delft (Holland) in June 1989. After the administrative problems had been solved, and an authorization to transmit a radio signal from France has been obtained, it was finally decided to set this station up at Observatoire de la Côte d'Azur (OCA, Grasse, France). There was a unique chance to compare the three state of the art techniques connected to the same remote atomic clocks: Common-view GPS, Two-Way Time Transfer (MITREX) and the Laser Ranging Technique of the LASSO experiment. The only other European station equiped for these three techniques was Observatory Lustbühel (Graz, Austria). The joint experiment started June 22 1990.

This paper is the following of the paper presented at the last PTTI meeting [1]. It emphasizes the MITREX technique and results between OCA and GRAZ. MITREX data are compared with data obtained from other techniques.

1.General remarks on Two-way Time Transfer.

1.1. Two-way time transfer technique.

Two-way satellite time transfer uses a telecommunication satellite for the comparison of the clocks. Ground stations able to transmit and receive radio signals are needed near the clocks to be compared. Preliminary experiments have showed it has nanosecond level capabilities.

The technique is described on **figure 1.** The two stations transmit simultaneously a coded signal which is referred to their clock 1 Hz signals. This 1 Hz signal starts a time interval counter (TIC). The satellite transponder receives the transmitted signals, and retransmits them to the opposite station. Each station receives the coded signal from the other one, from which the clock signal is extracted to stop the TIC.

The clock offset $H_A - H_B$ is given by [2]:

$$H_A - H_B = 1/2 [I(A) - I(B) + (D_{TA} - D_{RA}) - (D_{TB} - D_{RB}) + (D_{AS} - D_{SA}) - (D_{BS} - D_{SB}) + D_{ASB} - D_{BSA}] + C$$

where

$I(A)$ and $I(B)$ are the results of the TIC measurements,
D_{TA} and D_{RA} are Uplink and Downlink delays in ground station A,
D_{TB} and D_{RB} are Uplink and Downlink delays in ground station B,
D_{AS} and D_{SA} are propagation delays between ground station A and satellite,
D_{BS} and D_{SB} are propagation delays between ground station B and satellite,
D_{ASB} and D_{BSA} are delays in the satellite transponder,
and C is an additionnal delay due to the earth rotation (Sagnac effect).

TIC of picosecond precision (20 ps in the case of this experiment) are needed. The ionospheric delays are negligible for ns clock comparisons, if the frequencies of the signals carriers are high enough [3]. In the case of an experiment with Eutelsat, the carrier frequencies are in the Ku-band (11 - 14 GHz), so the propagation delays cancel (Intelsat uses Ku-band or C-band). The Sagnac effect can be calculated with a sufficient accuracy, even without knowing the exact position of the satellite nor the ground stations [2].

For short distance clocks comparisons, the same transponder is used by both signals, so the delays in the satellite

transponder cancel. For long distance, there could be a systematic offset if two different transponders, or two different carriers, are used on board the satellite.

Accuracy also depends on the measurements of internal delays in the ground stations. Two different techniques could be used for that. First one is global calibration of both stations with the use of a third one [2]. This portable station C is used first in parallel with the station A, with the same reference clock, and the delay (A - C) is obtained. Then the portable station C is used the same way in station B, providing (B - C) delay. It is then easy to get the global delay calibration (A - B), but only for a given date. The second technique is an absolute calibration of the delays in the complete ground station [4]. Both techniques are not very simple to use. Untill now, we don't know about the long term stability of the internal delays of ground stations.

1.2. MITREX modem.

The modems used on the link between Grasse and Graz are of the MITREX 25000 type [5], conceived by Prof. Hartl (University of Stuttgart). This device is the only one available for nanosecond Two-way Time Transfer. It needs the 1 pps (pulse per second) and 5 or 10 MHz clock frequencies as input. A 1 pps signal is generated in the modem, which is transformed by spread-spectrum techniques into a binary sequence (2.5 MHz chip rate), and coded as a pseudo-random noise signal up to 8 orthogonal codes. It is then put on a 70 MHz carrier, which is usual for links with telecommunication satellite terminals. On the receive channel, the pseudo-random noise coded signal from the remote station is correlated with the proper code to rebuild a 1 pps signal to compare with the signal coming from the clock.

2. The three techniques between OCA and GRAZ.

2.1. LASSO technique.

The technique has been described in a previous paper [1]. There has been no more LASSO results since the common sessions obtained in November 90.

2.2. GPS technique.

The Common-View GPS time transfer is a well known technique [6]. Therefore only the specifications of the OCA/GRAZ GPS time comparisons will be briefly discussed here.

The receivers are both of the classical NBS type (one channel, C/A code, L1 carrier, calculated ionospheric delays), except that the software used at OCA was partly of a different type than in GRAZ concerning the use of Bloc II satellites. This is no longer the case since last December. Both stations are following the BIPM Common-View schedule implemented every six months, but during the first part of the experiment only the Bloc I satellites have been used, to avoid any influence of the Selective Availability (SA). The Time section of the BIPM has calculated a specific schedule for the OCA/GRAZ link, in order to get a minimum of one track every half hour. In this way, the involved GPS receivers cover the Two-way and the LASSO sessions. The schedule was implemented on MJD 48244, December 19. Since then, all available satellites (including Bloc II without SA) have been used.

The coordinates of both antennas are very well known. They have been calculated by BIPM from laser geodetic points not far from receiver's antennas. This means that the coordinates' unaccuracy for the OCA/GRAZ link is less than 10 cm, and that its influence on the accuracy of the time comparisons is negligible.

2.3. Two-way MITREX link.

2.3.1. OCA earth station.

The transmit and receive terminal at OCA is of the VSAT type (Very Small Aperture Terminal). The diameter of the antenna is 1.8 m. Its approximate position is: 06° 55' N, 43° 45' E. The OCA earth station is described in **figure 2**, including the link between the atomic clock giving UTC to the GPS receiver. The main telecommunication characteristics of this station are G/T = 21 dB/K and EIRP = 49.2 dBW , when the power is set to achieve a C/N_0 (Carrier to Noise) of 55 dB/Hz (clear sky conditions).

2.3.2. GRAZ earth station.

The GRAZ earth station is described on **figure 3**. The steerable antenna of the earth station, dedicated to satellite telecommunication experiments, has a diameter of 3 m. The characteristics are G/T = 24 dB/K and, according to Eutelsat link budget calculations, EIRP = 52 dBW to obtain a C/N0 of 55 dBHz (maximun EIRP of this station is 72 dBW). The configuration of the different devices is slightly different from that of the OCA MITREX earth station. One can see that there is a synchronization device providing the 10 MHz frequency to the Modem.

The approximate position of the earth station is 15° 30' E, 47° 04' N.

On January 14 1991 (MJD 48270), a change of the reference clock was necessary at GRAZ station. The running clock was a HP option 04, the new one is a normal tube clock.

2.3.3. Eutelsat satellite.

The SMS (Satellite Multi-Service) transponder 14 of the satellite Eutelsat I-F2 was used during the first part of the experiment. A change of satellite (switch of SMS service between I-F2 and new satellite I-F4) occured October 16, because I-F2 was too high regarding the nominal position, and was too old for repositionning. The nominal position of both satellites was 7° East ± 0.1°. No operation on the ground stations has been needed. The common transmission (Tx) and reception (Rx) carrier frequencies of both stations were:

$$F_{Tx} = 14022.000 \text{ MHz}$$
$$F_{Rx} = 12522.000 \text{ MHz}$$

The choice of the channels on the satellite transponder was limited because of the inability of fine adjustments to the frequency of the VSAT station. The on-board oscillator of the satellite has proven to be very stable during our experiment, which was very important for the use of a VSAT station [7]. But this oscillator was steered by a Eutelsat ground station. This will no more be the case in the next future with the Eutelsat II generation which will have a free on-board oscillator [8].

2.3.4. Procedure.

The regular MITREX clocks comparisons started the 22nd of June 1990. They were interrupted from October 12 until November 9, because there was some water in the filter of the receiving part of the OCA antenna. It will probably stop at the end of March 1991.

We have used a regular tracking schedule of 3 sessions a week (Monday, Wednesday, and Friday). Each of these satellite sessions started at 12:00 UTC, and ended at 12:30 UTC. The MITREX measurement sessions were 2 min long [9], one point every second, except at the start of the experiment where 4 min sessions were scheduled. This has been organized to be sure to get a 2 min long usable session. After a while, the MITREX clocks comparison proved to be working well and only 2 min sessions have been scheduled. At the end of August, the three sessions of one week were scheduled to last 20 min. A look at the results (see next chapter) proves that the choice of 2 min long sessions was enough for the OCA/GRAZ MITREX link.

During the first part of the experiment, before interruption at OCA, two MITREX sessions were scheduled at 12:15 and 12:20 UTC, with the OCA 1 pps reference used to start the TIC. After the interruption at OCA, a third MITREX session has been added at 12:27 UTC, with a 1 pps reference issued by the Modem, called 1 pps Tx.

3. Results.

The whole campaign of MITREX and GPS measurements can be seen on **figure 4**. The long term drift of the clocks difference (118 ns/day) has been removed. When considering the events that occured during this period of time, it has been decided to split the results in three different parts:

Part I. Before OCA MITREX intermission (June 22 to October 10).

Part II. After OCA intermission (November 9 to January 14).

Part III. After change of the reference clock at GRAZ (until February 22).

No calibration has been made on the two-way MITREX link between OCA and GRAZ.

The distance between OCA and GRAZ is about 800 km. The GPS precision between these two stations is usually around 2 ns. The GPS accuracy is estimated to be more or less 2.5 ns [6].

The precision obtained for LASSO sessions in November 1990 is around 1 ns. There has been no calibration for the LASSO experiment. Because there has been only a few measured LASSO points attached together no LASSO result can be seen on the common results diagram (figure 4). See [1] for more details.

3.1. Two-way Precision.

For the precision plots, selection of raw data points within a window of ± 5 ns was done. Very few points have been dropped.

The plot of the precision values of the 12:15 UTC and 12:20 UTC sessions can be seen on **figure 5** for the Part I of the experiment. The mean values of the precisions obtained are:

$\langle sigma_{12:15} \rangle = \langle sigma_{12:20} \rangle = 0.83$ ns.

Nearly all points are below 1 ns, except for one noticeable point named A. Raw data of the MITREX session for point A are on **figure 6**. One can see that something occured right in the center of the session.

For Part II and III of the experiment, **figure 7** shows the precision of the values, for the 12:15 UTC, the 12:20 UTC and the 12:27 UTC sessions. One can see the influence of the presence of an unknown clean carrier (measured at 12523.667 MHz) in the MITREX bandwidth on December 10: precision data are degraded up until nearly 1.5 ns.

The pps_{Tx} remained very stable compared to the OCA pps_{Ref}, and its influence can be seen on this figure. For Part II of the experiment, the mean values of the precision were:

$\langle sigma_{12:15} \rangle = \langle sigma_{12:20} \rangle = 0.85$ ns, $\langle sigma_{12:27} \rangle = 0.78$ ns.

Some points of Part III of the precision data seem to be greatly degraded. Because of the strong relationship between the two MITREX stations in the measured data, it is not easy to separate effects coming from one location or the other. These points represent MITREX sessions that are globaly degraded, giving some sigma values over 1.4 ns. It could come from MITREX links that were for some reason bad during these sessions.

Mean values of the precision for Part III are:

$\langle sigma_{12:15} \rangle = 0.90$ ns, $\langle sigma_{12:20} \rangle = 1.01$ ns,

$\langle sigma_{12:27} \rangle = 0.89$ ns

What appears to be a small global degradation, could come from the change of the GRAZ clock.

3.2. Accuracy.

Figure 8 shows the plots of three 20 min sessions on MJD 48123 - 125 - 127. When looking carefully at the data, one can see how MITREX comparisons follow the 120 ns/day offset, giving an offset of nearly 1.7 ns/20mn, due to the long term drift of the clocks. For 2 min sessions, this offset is around 170 ps, and so remains negligible compared to the noise of the MITREX link.

The calibration Ref-Tx data are ploted on **figure 9**. This is the only easy available Modem calibration, and its use is necessary when Two-way is compared with other techniques: it has to be refered to the same clock signal. These data have been measured just before the 12:15 UTC sessions. They have been obtained by using the pps_{Tx} output of the Modem to stop the TIC. The standard deviations of the calibration data over the whole measurement period are:

$Sigma_{Ref\text{-}Tx\ OCA} = 1.9$ ns, $Sigma_{Ref\text{-}Tx\ GRAZ} = 0.4$ ns.

These calibration data have been used for the computation of the GPS-MITREX data.

Figure 10 shows the plot of GPS-MITREX data. A simple mean over the 2 min sessions has given the MITREX points. Vondrak smoothing has been applied to the GPS data to interpolate the GPS points to the middle of the appropriate Two-way sessions. The goal of the Vondrak smoothing is to minimize the noise coming from the method of comparison, in order to get only the noise of the clocks. The smoothed values are interpolated for the desired time using Lagrange's interpolation formula. Vondrak smoothing acts as a low-pass filter. For Part I and Part II of the experiment, a Vondrak cut-off period of about 1.5 days has been used. After the change of the GRAZ clock, a Vondrak cut-off period of 0.5 day has been applied. Three points have been excluded, because there were only few GPS traks during these days, and the results of the smoothing and interpolation were not reliable. The mean values of GPS-MITREX data and their standard deviations have been computed:

$\langle GPS\text{-}MITREX \rangle_{Part\ I} = -205.8$ ns, $Std\ Dev._{Part\ I} = 2.6$ ns,

$\langle GPS\text{-}MITREX \rangle_{Part\ II} = -210.5$ ns, $Std\ Dev._{Part\ II} = 3.1$ ns,

$\langle GPS\text{-}MITREX \rangle_{Part\ III} = -207.3$ ns, $Std\ Dev._{Part\ III} = 3.6$ ns.

The value of around -200 ns comes from the difference of all the delays at both stations, including the delays of MITREX ground stations which have not been measured.

There is an offset of about 5 ns between the mean values GPS - MITREX before and after the interruption at OCA. This can not be a consequence of the change of satellite at this time, because delays in the transponder of the satellite cancel. The transmit and receive electronics of the OCA antenna were dried. This operation can not explain all the 5 ns offset. It could also have arisen from a change in the delays at one (or both) station(s), either MITREX ground stations, or GPS receivers.

Another explanation could be the slow daily variation (4 min) of the sidereal period of the GPS satellites. At the start of the experiment, there was a group of satellites that was scheduled to be observed around the time of the MITREX measurements (**figure 11.1**). As time went by, the tracks slowly moved away (**figure 11.2**). After OCA intermission, there were no more satellites available for GPS measurements around the time of the MITREX sessions (**figure 11.3**), until the new schedule was implemented on December 19, for which BIPM has computed a specific schedule between OCA and GRAZ, including nearly one tracking every half of an hour (**figure 11.4**).

Conclusions.

The results obtained with MITREX on a 800 km basis with one of the stations being a VSAT are very promising. The precision achieved for normal sessions is generaly around 0.8 - 0.9 ns. Accuracy variations of the difference GPS-MITREX reach 10 ns for a 9 months experiment. If the 5 ns offset before and after the OCA interruption is considered as a change in the delays at one location, which could be some way measured, then the MITREX/GPS comparison on the OCA/GRAZ link gives an uncertainty of one sigma of around 3 ns, which scarcely exceeds the uncertainty of GPS accuracy alone.

The question of the delay calibration in ground stations, and their possible changes with time, remains an essential problem to be solved for nanosecond time transfer with Two-way technique. But other points are of particular importance: stability of the station frequency distribution, relationships with UTC(lab), all other delays involved, TIC reliability, trigger levels,...

One of the goals of future Two-way experiments should be an attempt to look at the influence of mixing several pseudo-random noise coded signals into one satellite transponder. To share a transponder with other users at the same time, could simplify administrative problems and reduce prices for operational regular Two-way links.

Acknowledgements.

All the participants express their gratitude to Prof. Hartl (University of Stuttgart) for having loaned the MITREX modem which was used at OCA.

The authors gratefully aknowledge Eutelsat for having given three months of free time at the start of the experiment. The participants are grateful to MM. Fiedler, Chabrol (Eutelsat), MM. Meunier, Bacot (France Telecom), and M. Schladofsky (OPTV) for having made this experiment easier.

This work is supported by Bureau National de Métrologie (France), Austrian Academy of Sciences and Jubilee Fund of the Austrian National Bank.

Bibliography.

[1] C.Veillet, D.Féraudy, J.M.Torre, J.F.Mangin, P.Grudler, F.Baumont, J.Gaignebet, J.L.Hatat (OCA), W.Hanson, A.Clements, J.Jespersen, M.Lombardi (NIST), D.Kirchner (TUG), G.Kirchner, H.Ressler, S.Fassl (Space Res. Inst. Graz), W.J.Klepczynski, P.J.Wheeler, W.Powell, A.Davis (USNO), P.Uhrich, R.Tourde, M.Granveaud (LPTF).
"LASSO, Two-Way and GPS time comparisons: a (very) preliminary status report."
23rd Annual PTTI meeting, Washington, Dec. 1990 (to be published).

[2] D.W.Hanson.
"Fundamentals of two-way time transfers by satellite."
43rd An. Symp. on Freq. Control, 1989.

[3] J.Jespersen.
"Impact of atmospheric non-reciprocity on satellite two-way time transfer."
43rd An. Symp. on Freq. Control, 1989.

[4] G.DeJong.
"Accurate delay calibration for two-way time transfer."
21st Annual PTTI Meeting, 1989.

[5] P.Hartl.
"MITREX 25000. A Modem for microwave time and ranging experiments via telecommunication satellites."
Institute for Navigation, Univ. Stuttgart, January 1985.

[6] W.Lewandowski, C.Thomas.
"GPS time transfer"

IEEE Special Issue on Time, August 1991 (in press).

[7] D.Kirchner.
"Design considerations for Ku-Band two-way satellite time transfer terminal."
4th EFTF, 1990.

[8] Eutelsat.
"Eutelsat SCPC Satellite Multiservice System specification. Revision 2."
March 1990.

[9] W.J.Klepczynski, P.J.Wheeler, W.Powell, J.Jeffries, A.Myers, R.T.Clarke (USNO), W.Hanson, J.Jespersen, D.Howe (NIST).
"Preliminary comparison between GPS and Two-way Satellite Time Transfer."
42nd An. Symp. on Freq. Control, 1988.

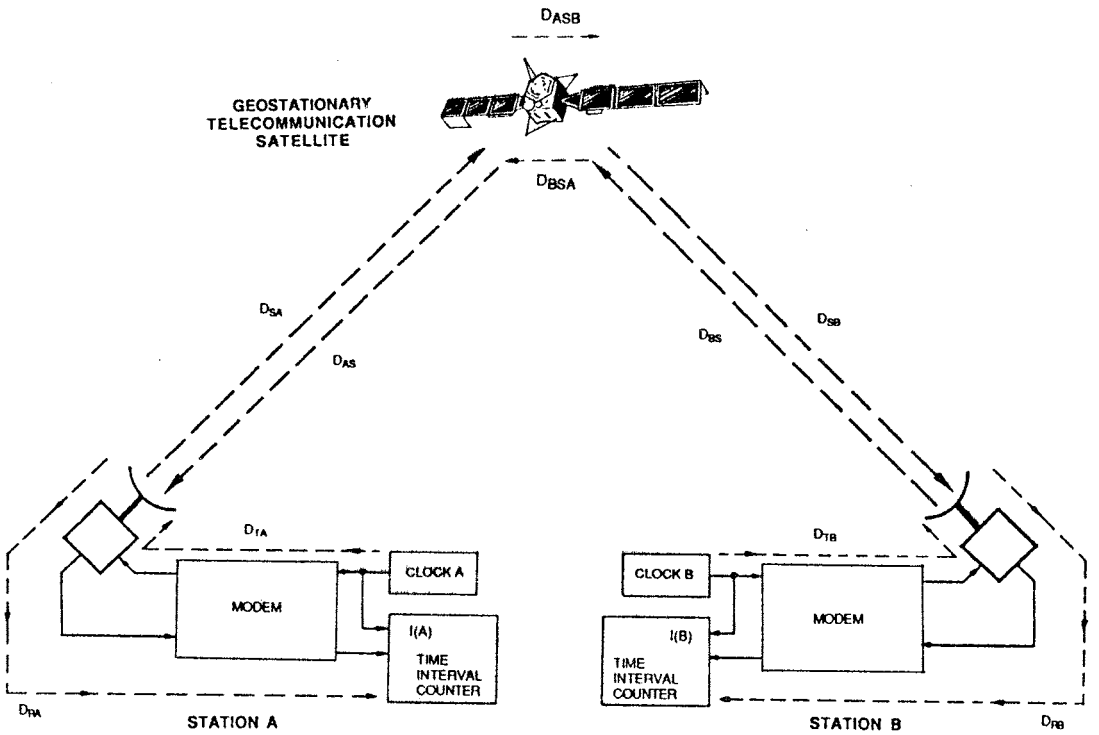


Figure 1. Two-way time transfer technique.

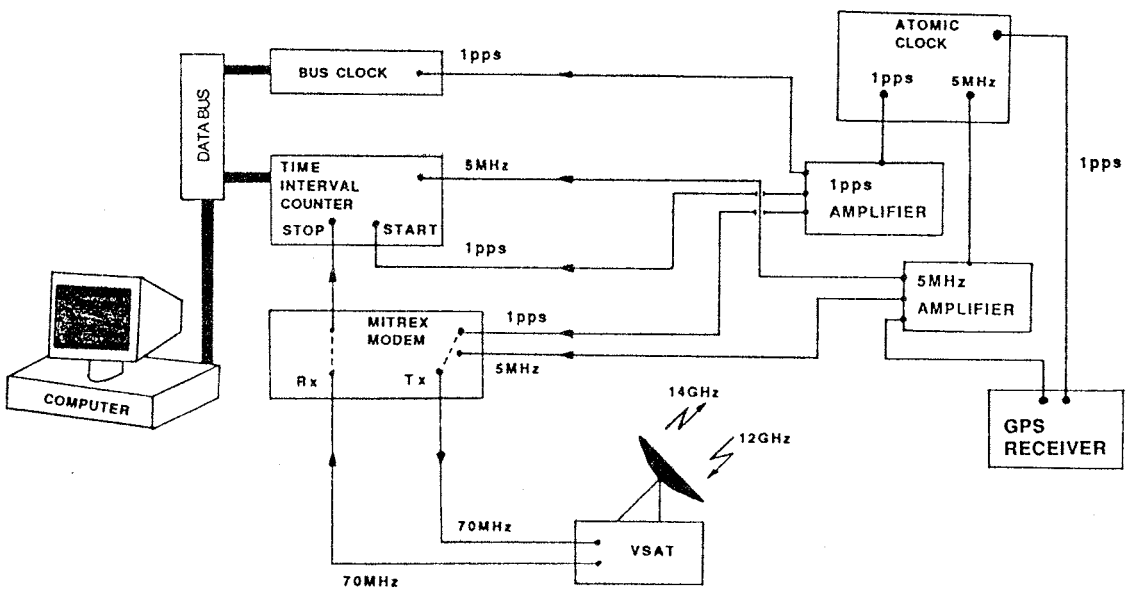


Figure 2. OCA earth station.

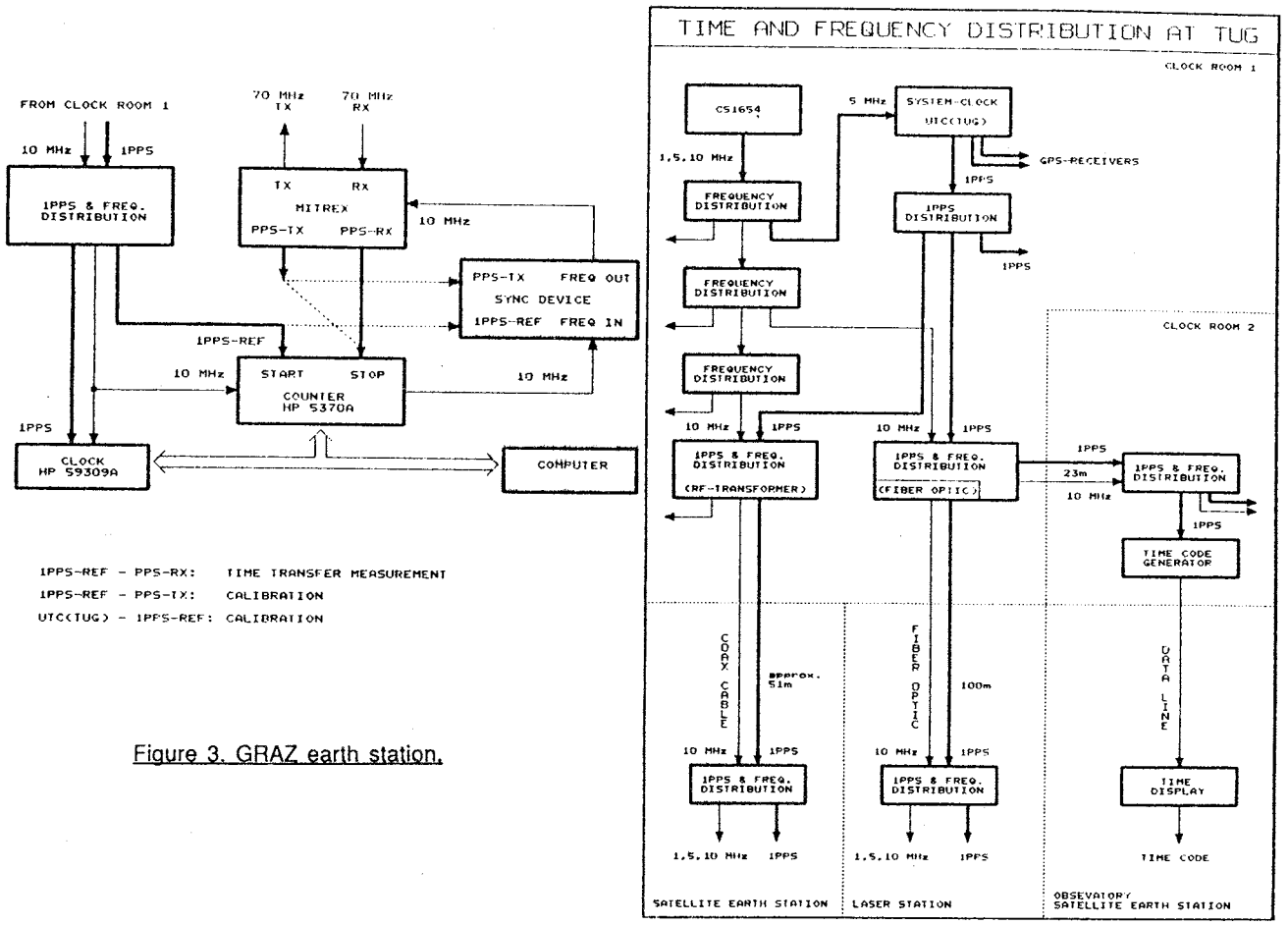


Figure 3. GRAZ earth station.

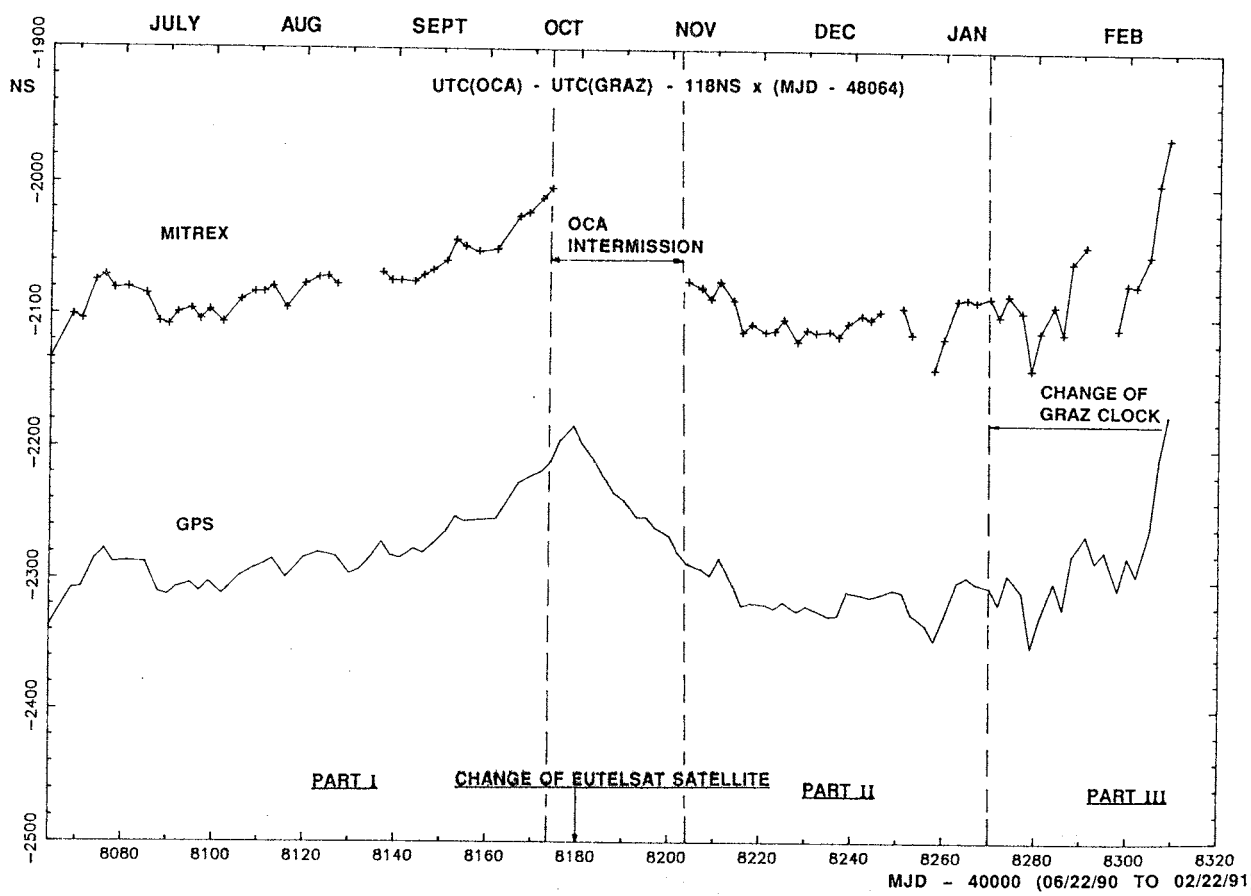


Figure 4. Experimental MITREX and GPS results.

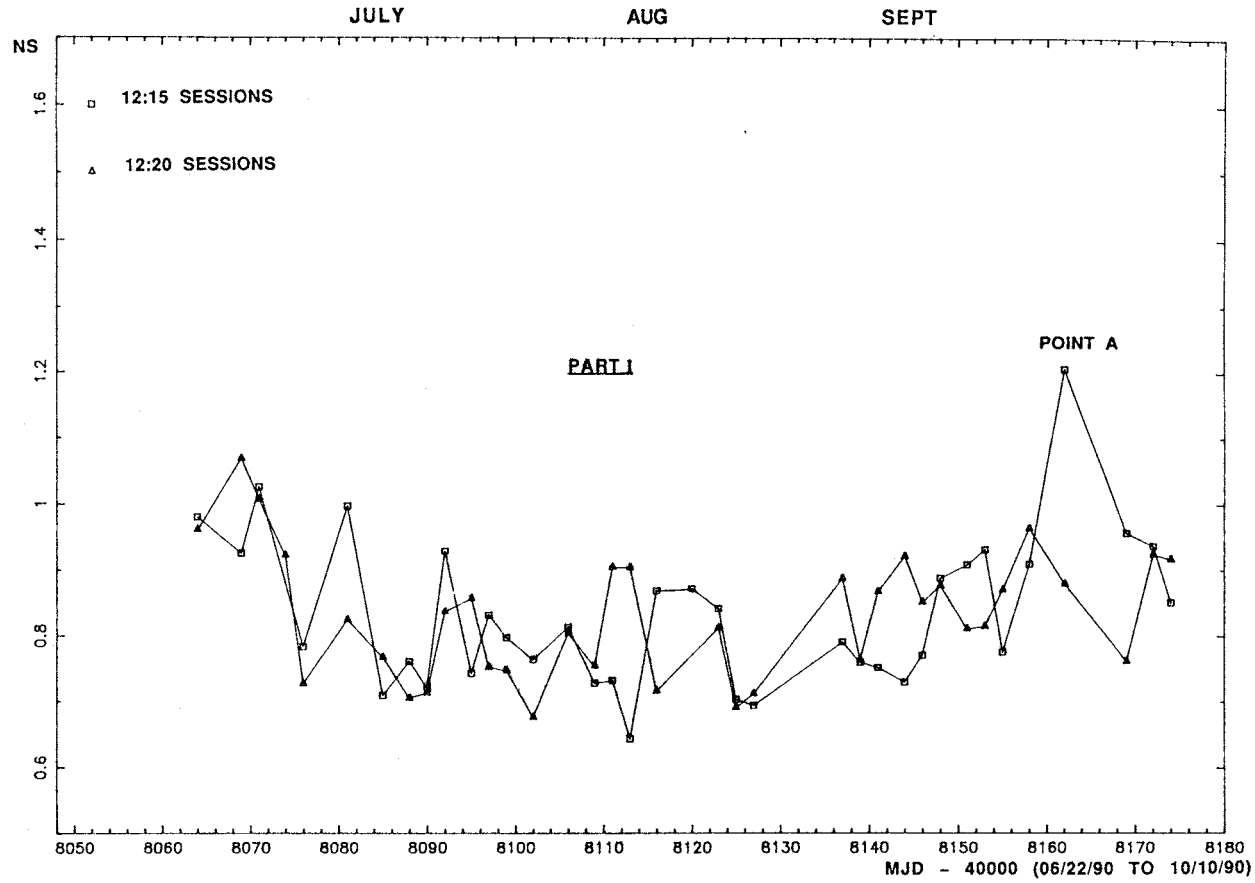


Figure 5. MITREX precision Part I.

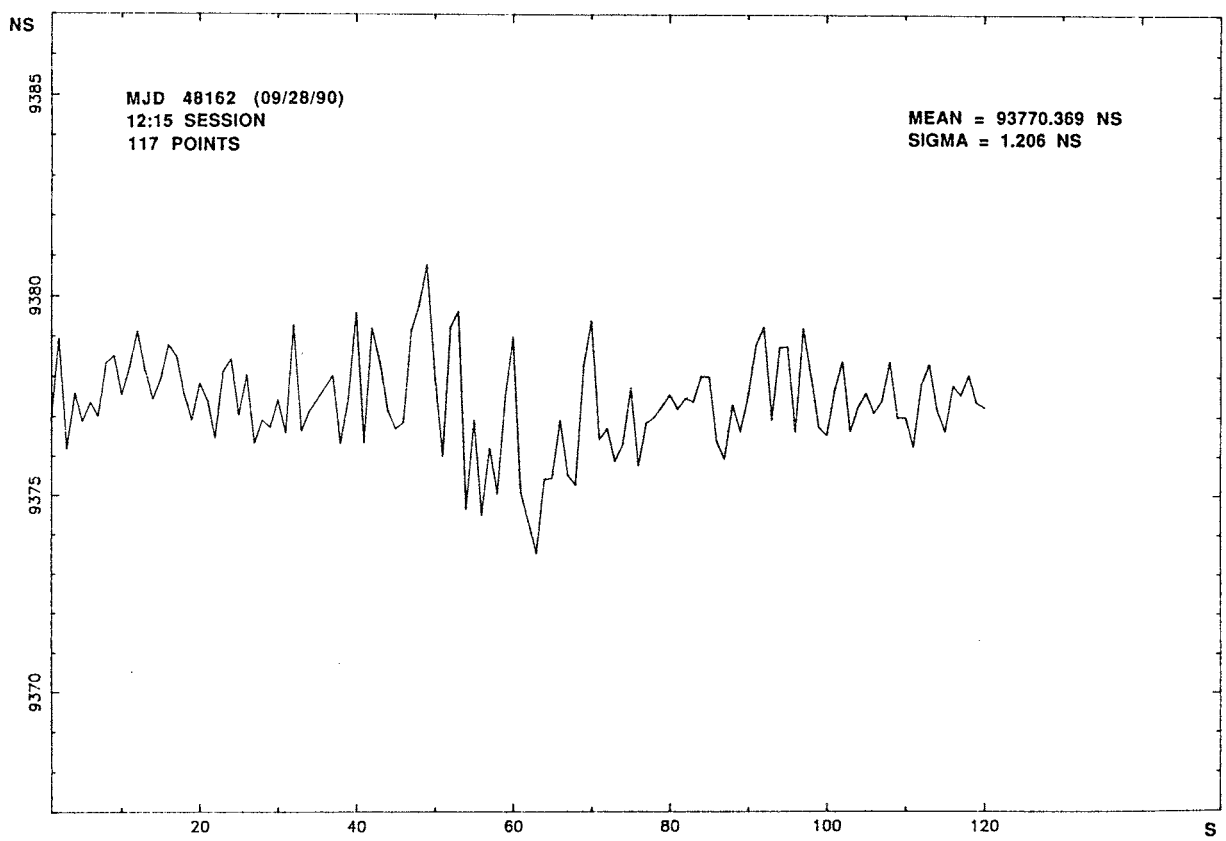


Figure 6. Raw MITREX data of point A of figure 5.

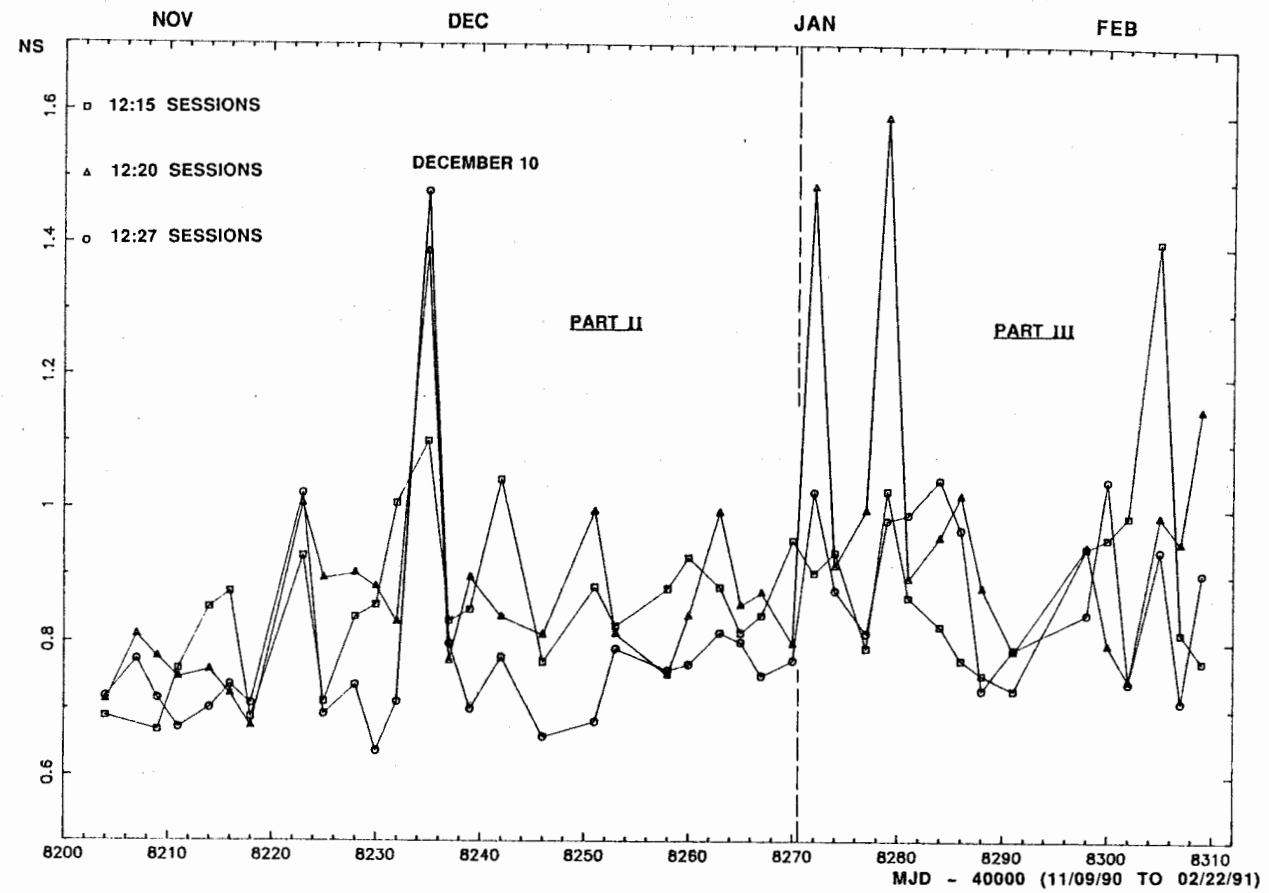


Figure 7. MITREX precision Part II and Part III.

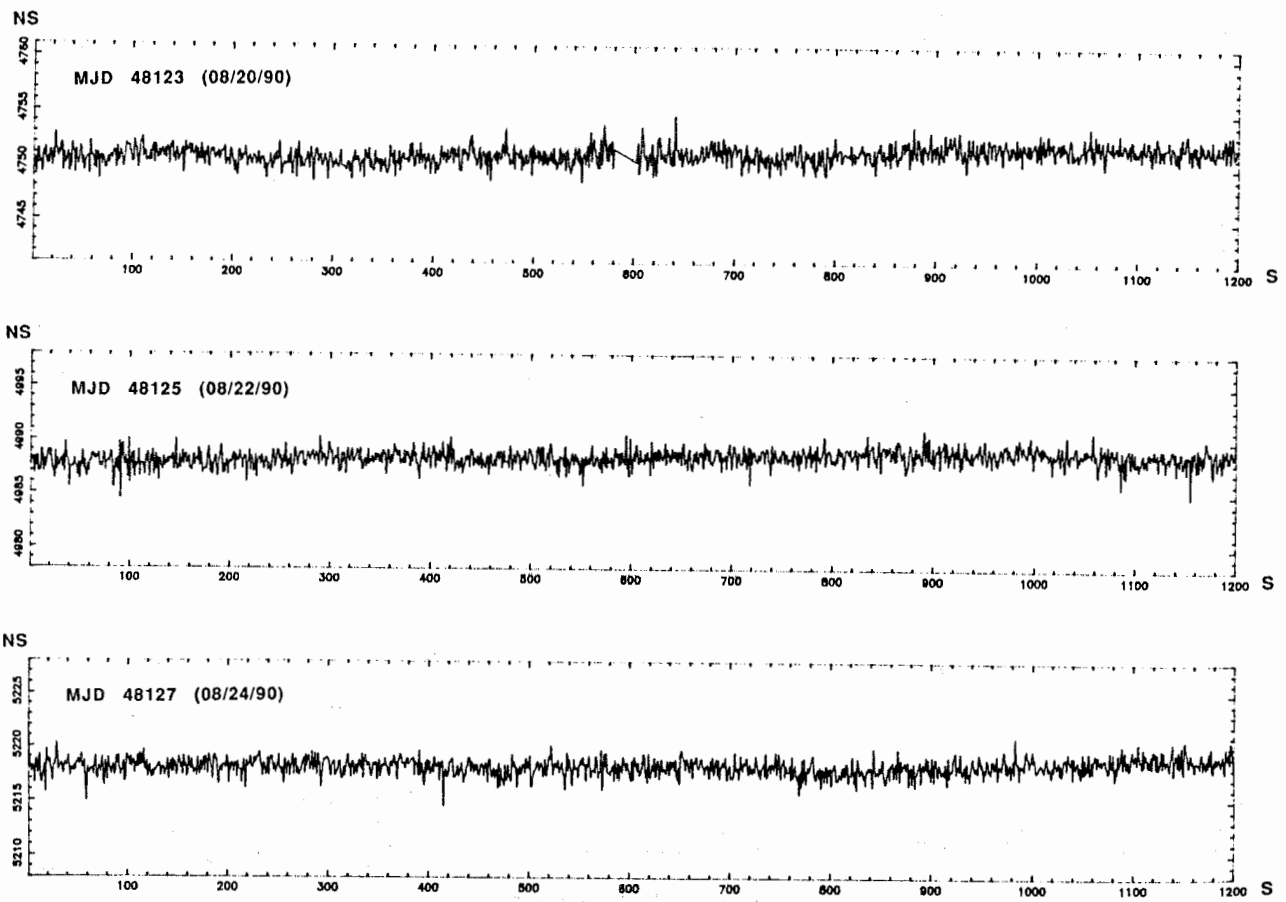


Figure 8. Plots of 20 mn sessions.

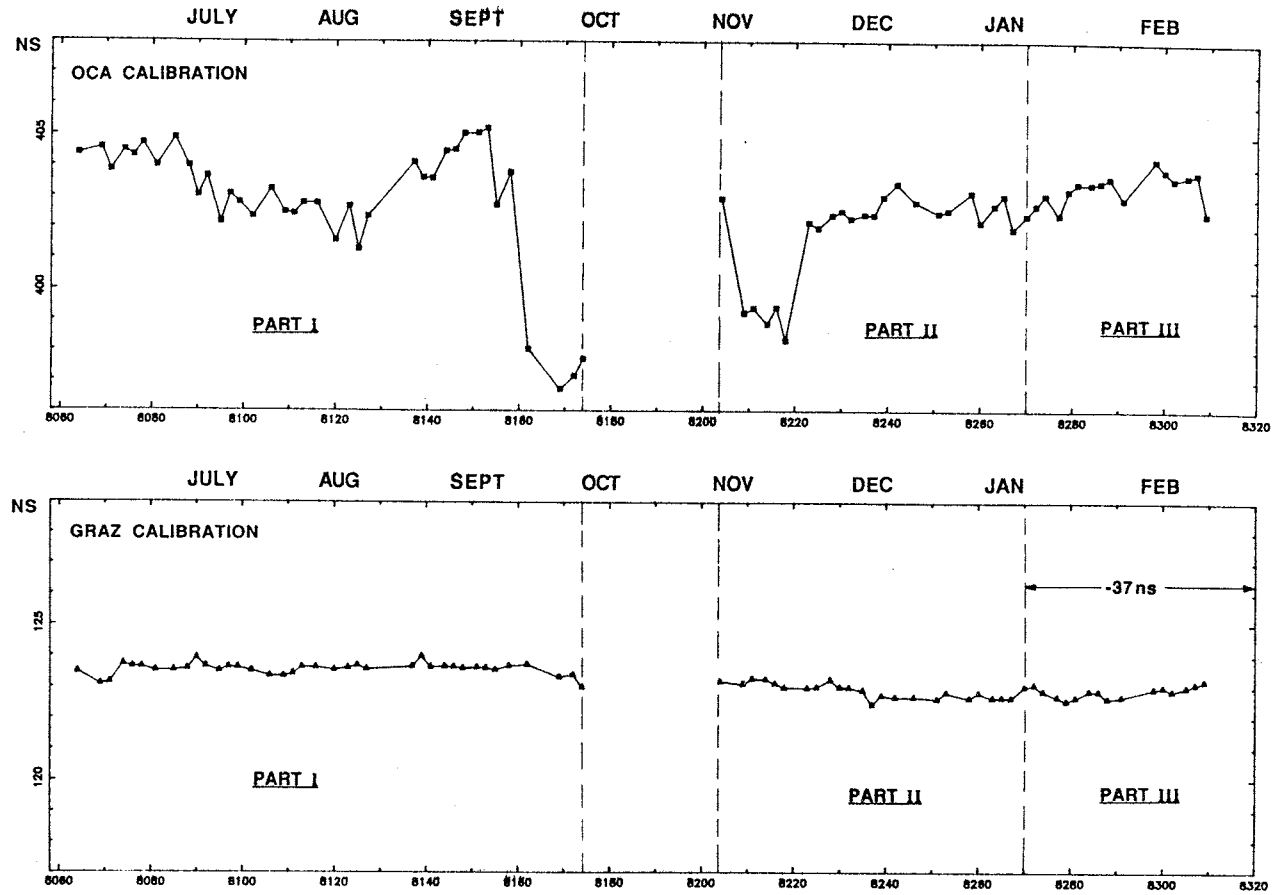


Figure 9. Ref-Tx calibrations.

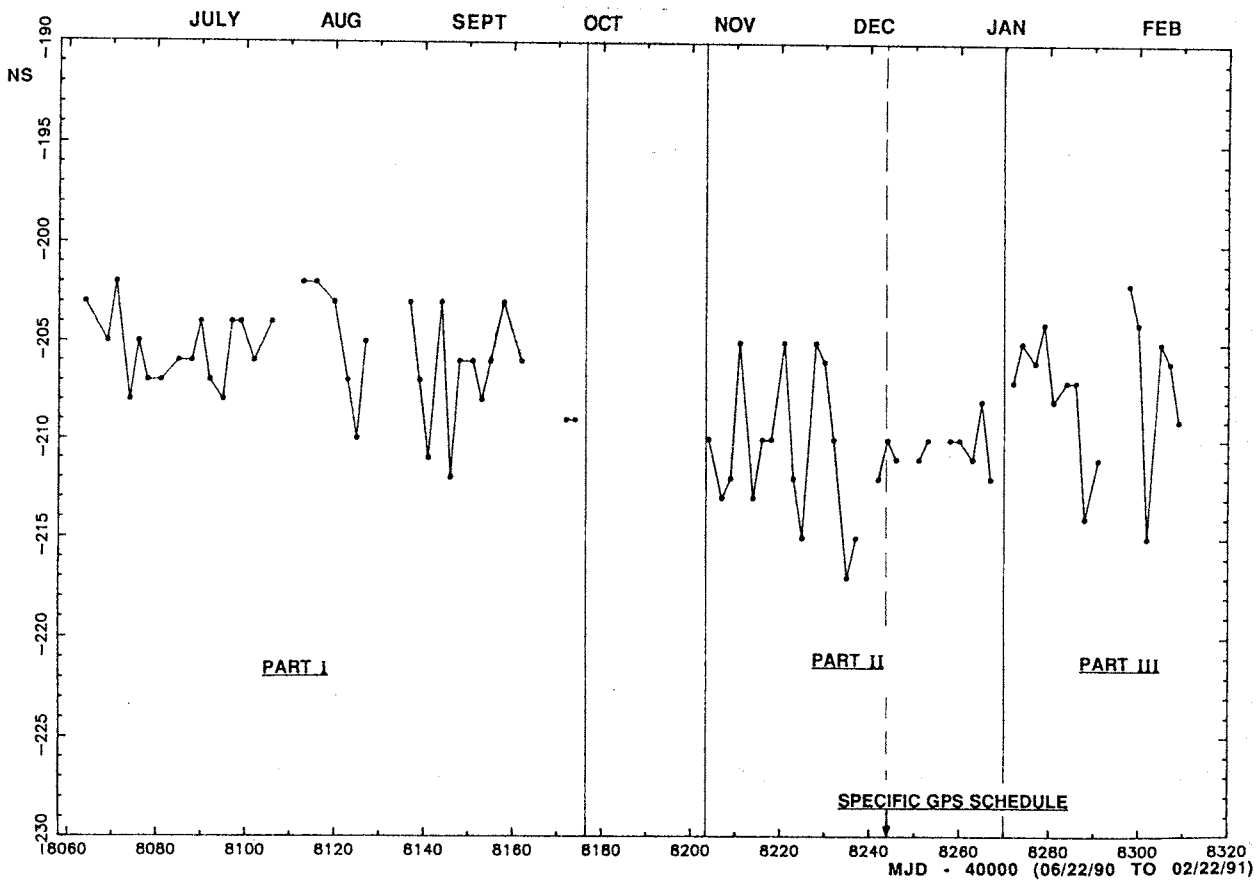
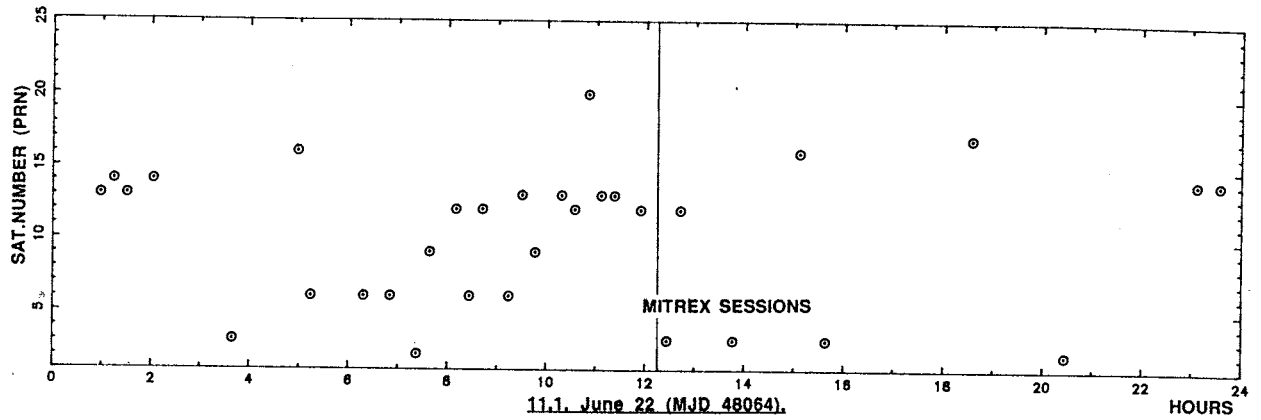
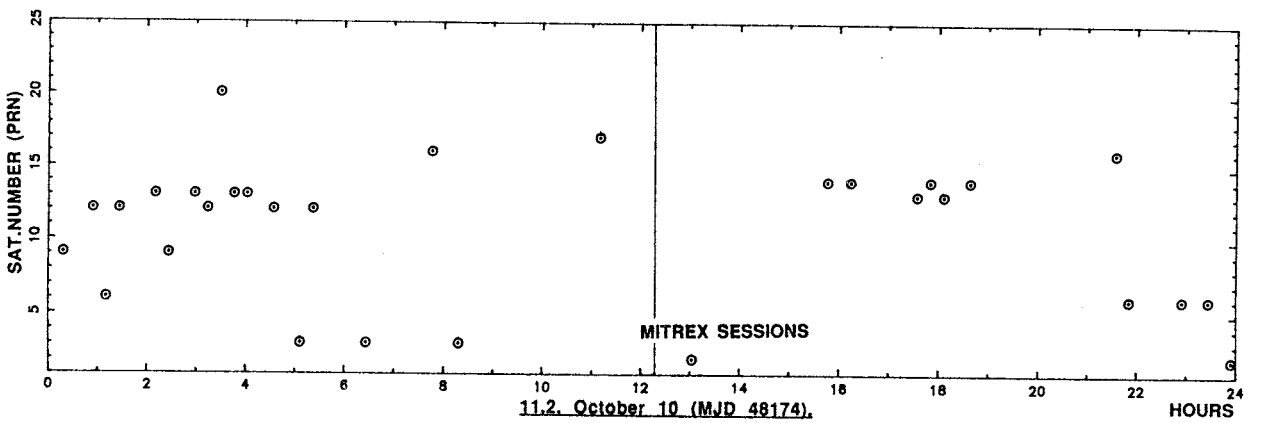


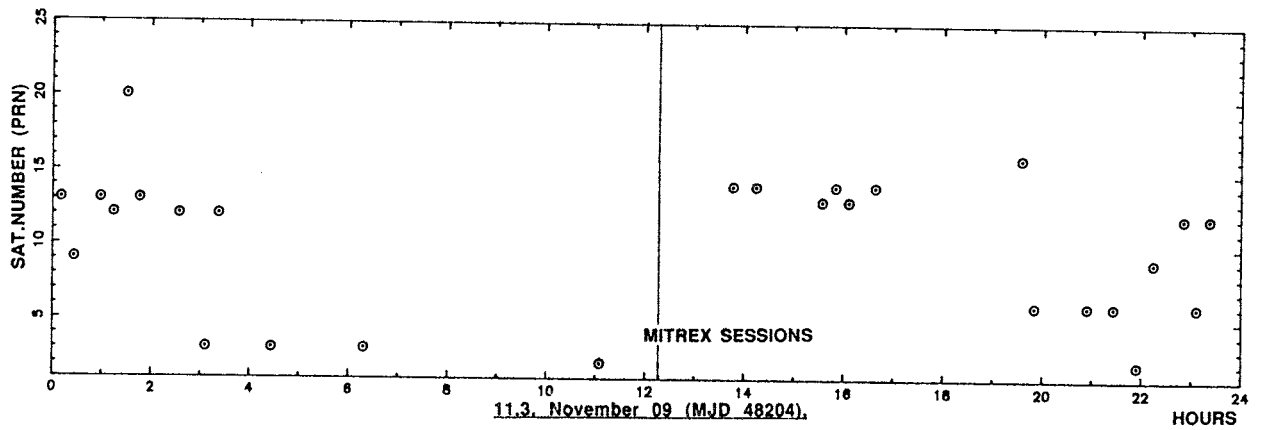
Figure 10. GPS - MITREX.



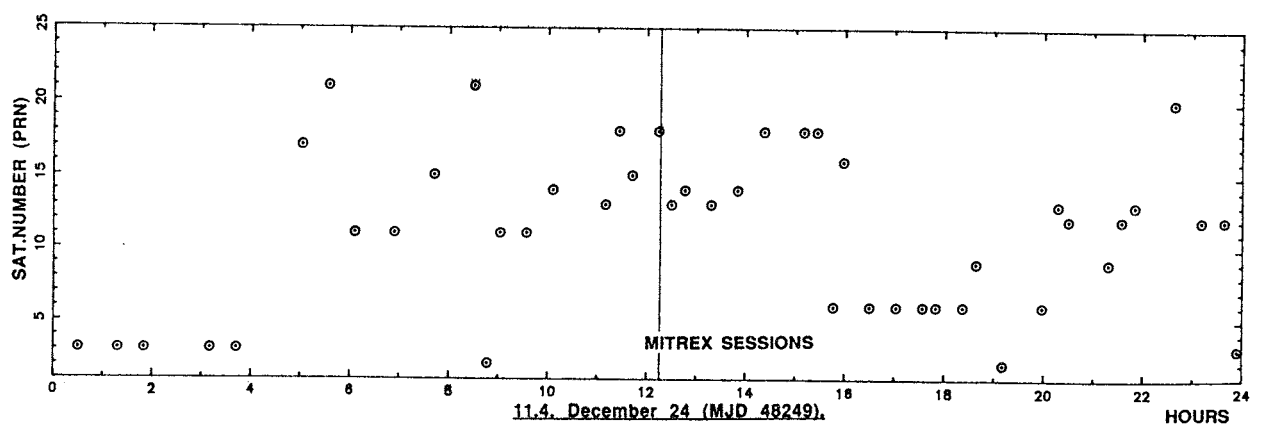
11.1. June 22 (MJD 48064).



11.2. October 10 (MJD 48174).



11.3. November 09 (MJD 48204).



11.4. December 24 (MJD 48249).

Figure 11. GPS traking schedules.

SESSION 3B

TIME AND FREQUENCY II

Chairman : M. GRANVEAUD

Session 3B Number 18

VERY FAST DIRECT DIGITAL SYNTHESIZER WITH DRO TIME BASE FOR INSTRUMENTATION APPLICATIONS

M.T. Diep*(+), J.F. Zürcher*, F. Gardiol*, P. Dénervaud**

* Laboratoire d'Electromagnétisme et d'Acoustique, EPFL, Lausanne, Switzerland
** Centre Suisse d'Electronique et de Microtechnique, Neuchâtel, Switzerland

Abstract

A direct digital synthesizer (DDS) for the frequency range 1 Hz to 100 MHz has been developped for metrological applications. After describing the principle used for DDS and its practical implementation in a module, we present the measurement results, with special emphasis on the phase noise characteristics.
We also present the first development stage of a clock reference source for DDS based on ceramic coaxial resonator oscillator

Introduction

The development of a dedicated instrument for fiber optic measurements (OFDR : Optical Frequency Domain Reflectometry) required the realisation of a direct digital synthesizer for the frequency range 1 Hz to 100 MHz, as well as a reference clock.
The basic structure of such an instrument is similar to that of a network analyser, but with electro-optic interfaces for fiber applications
The advantage of the DDS for this application is the speed of frequency change over the whole frequency range (including low frequencies). Also, there is phase continuity during the change.
This technique, initially applied for low frequencies, is becoming important in regard to classical synthesis techniques for the HF range. The rapid evolution of components allows us to reach frequency switching times that are inaccessible by other approches .
The principal problem concerns the spectral purity due to the reconstitution of a sampled signal. However, certain high speed D/A converters give satisfactory results in a large number of applications.

The synthesiser module

Principle of the direct digital synthesis

The essential part of a direct digital synthesizer is the numerically controlled oscillator (NCO). It is a "chip" that performs the logical operations synchronised with the reference clock.
The NCO is composed of :
- a Δphase register of N bits (actually between 24 and 32 bits dependent on the "chip").
- an adder, working with the same number N of bits as the Δphase register
- a phase accumulator, which is a register that stores the output value of the adder after each operation
- a logical circuit for the control signals

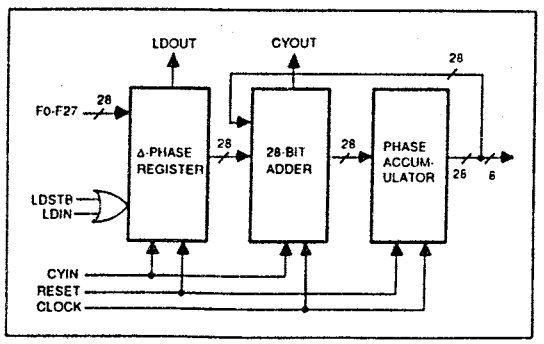


Fig 1 : Structure of the NCO

At each clock impuls the adder realises the sum of the content of the Δphase register and the phase accumulator. The new value is then transferred and memorised in a phase accumulator and will be taken into account during the next operation.
The content of the Δphase register (which has been loaded by a PC), controls the output frequency.
In fact one can imagine a vector that moves in a trigonometric circle and occupes 2^N positions that correspond to a phase value.
The binary value of the Δphase register defines the phase jump in the circle after each clock impuls, thus the phase variation and similarly, the frequency.

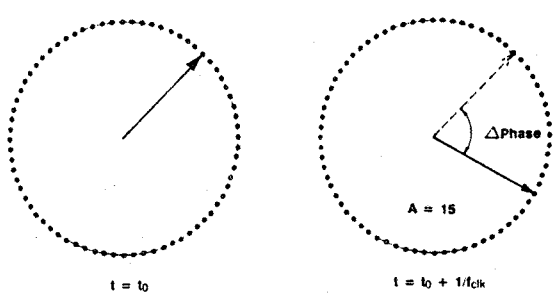


Fig 2 : Trigonometric representation of NCO operations

With the unity value in the Δphase register we have the minimum output frequency which is the resolution of the

(+) M.T. Diep presently at Asulab, Neuchâtel, Switzerland

synthesizer.

$$\text{Resolution : } f_R = f_{clock} / 2^N$$

N is the number of bits of the NCO

With the A value stored in * phase register we have an output frequency given by :

$$f_{out} = f_{clock} * A / 2^N$$

The output of the phase accumulator is a growing series of samples. The transformation in an analog signal would result in a saw-type signal.
In order to obtain a series of samples with a sinusoidal variation, the most significant bits (between 8 - 12) address a memory containing a sinus conversion table. It has to be noted that a signal with arbitrary shape could be obtained by an appropriate memory content.
The N_m data bit of memory, corresponding to a sinusoidal variation of amplitude, are sent to a D/A converter
At the output of the converter, a low-pass filter smoothes the signal at high frequencies and rejects the image frequency (alias) at $f_I = f_{clock} - f_{out}$.

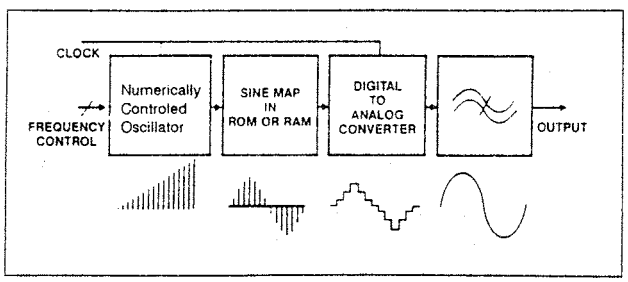


Fig 3 : Classical structure of a DDS

Spectrum of the output signal

The sampled nature of the signal induces a distorsion that leads to the appearance of non-harmonic spurious components.
The first sources are the phase and amplitude errors associated with the sampling process. The amplitude of this spurious signal is a function of the number of sampling bits. The spurious level is theoretically reduced to 6 dB under the carrier per bit of quantification. For example, if the quantification is performed on 8 bits, a spurious rejection of 6 * 8 = 48 dBc is expected.
The spectrum of the sampled signal contains also "image frequencies" at $M * f_{clock} + f_{out}$ (M = 1,2,3...). The first of this products ($f_{clock} - f_{out}$) is particularly perturbing (alias) and should be filtered out by the output filter. This imposes a limit for f_{out} at 0.4 * f_{clock},
($f_{out\ max}$ = 0.5 * f_{clock} limit : Niquist-Shannon theorem) so that the filter can reject the image appearing at 0.6 * f_{clock}.
The D/A converter also induces spurious signals that are attributed to :

- the time skew
- the integral and differential non-linearity
- the glitch energy

The main components of these impurities, due to DAc, are harmonics of f_{out} as well as their interferences with the clock frequency. The amplitude of the output signal of a DDS is not independent of the chosen frequency and it varies as a sinx/x law

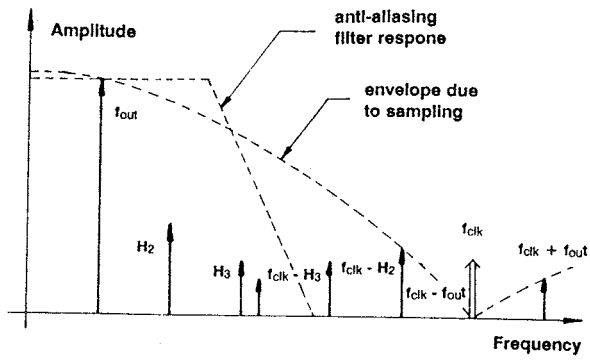


Fig 4 : Output spectrum with main spurious signals

Our realisation :

We realized a compact module containing the complete DDS with a data acquisition circuit (data generated by a PC). The power supply stabilisation with entry protection is included in the module.

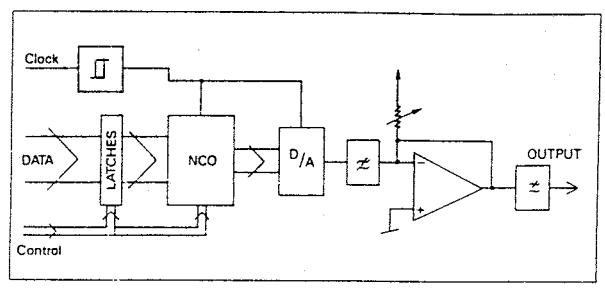


Fig 5 : Bloc diagram of the realized DDS

Apart from the classic bloc diagram of the DDS, we also added a high-speed comparator delivering appropriate clock signals to the D/A converter and the NCO via a delay line.
The "antialiasing" filtering at the output is performed in two stages. The insertion of an operational amplifier between the two filters allows us :
a) to adjust the DC offset at output,
b) to compensate the amplitude losses due to the the conversion operation (sin x / x response law)

The principal characteristics are :

- Frequency range : 1 Hz to 100 MHz
- Frequency resolution : 1 Hz
- Frequency swiching time : 150 ns
- Output level : 0 dB
- Flatness (1 Hz to 100 MHz) : $^{+}$ 1 dB
- Power supply requirements :
  + 8... + 10 V, 100 mA
  - 8... - 10 V, 2,2 A
- Dimensions : 150 * 150 * 52 mm
- Weight : 1 kg

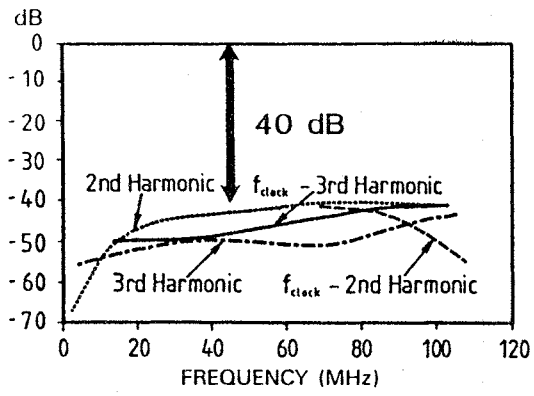


fig 6 : Spurious signal level

Phase noise

For narrow band applications, only the phase noise is of interest, with the condition that no spurious signals due to conversion are close to the output frequency.
The phase noise generated by two synthesizers has been measured with the following measurement setup :

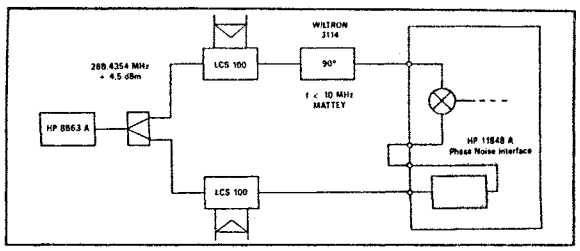
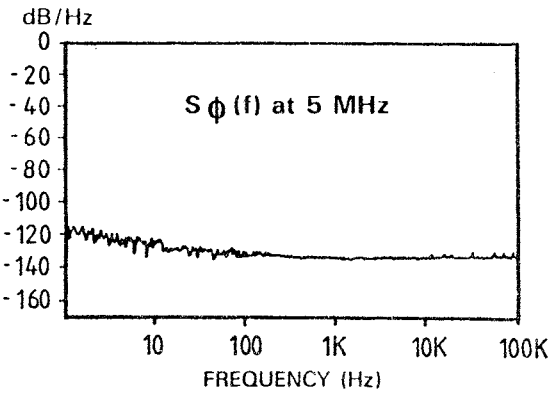


Fig 7 : Phase noise measurement setup

The following results have been obtained :



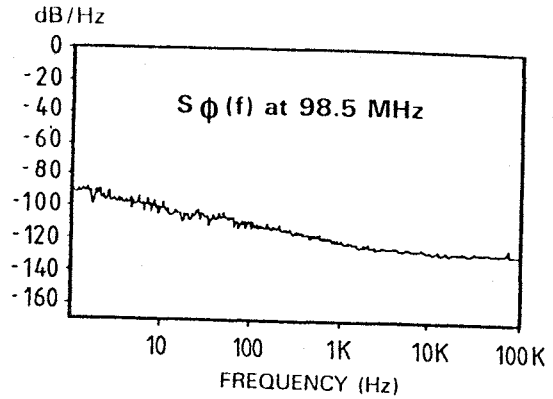


fig 8 : Phase noise for two synthesizers

It has to be noted that the measured noise is the one generated by the two synthesizers. The noise of the reference clock does not appear. In principle, the DDS allows one to obtain signals with better phase noise characteristics than the reference clock. The reason is that the operating frequency is lower than that of the reference clock, and in regard to the reference the DDS acts as a frequency divider.

Reference clock

Realisation

The time base has been designed for future generations of the DDS. Three output frequencies can be chosen :

$f_{out1} : 2^{31}$ Hz = 2.147483648 GHz
$f_{out2} : 2^{30}$ Hz = 1.073741824 GHz
$f_{out3} : 2^{28}$ Hz = 268.435456 MHz

The central part of the time base is the voltage controlled oscillator (VCO) using a ceramic coaxial resonator, selected for its excellent Q factor. The VCO frequency is controled after division by the traditional phase locked loop (PLL) method. A quartz oscillator with temperature compensation (TCXO) is used as reference for the PLL.

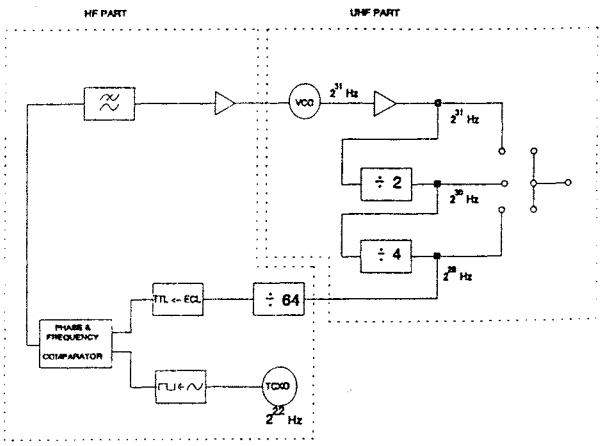


Fig 9 : Block diagram of the time base

The circuit is divided in two parts :
- the UHF part (VCO + dividers + amplifiers) realised in microstrip technology
- the HF part and PLL control (printed circuit board)

The UHF Part

The VCO, with its coaxial ceramic resonator, oscillates at a central frequency of about 2^{31} Hz. This frequency can be modified by + 2,5 MHz by using a tuning DC voltage between 0 - 5 V.
The Wilkinson divider, on microstrip, delivers two signals, one for the direct output (2^{31} Hz), the other for the input to divider 1 (1:2). An output of this divider delivers a reference at 2^{30} Hz and the former is used at the input of divider 2 (1:4). A 2^{28} Hz signal is then obtained. One output signal of divider 2 is sent to the HF circuit where the successive divisions are operated, up to the TCXO frequency.

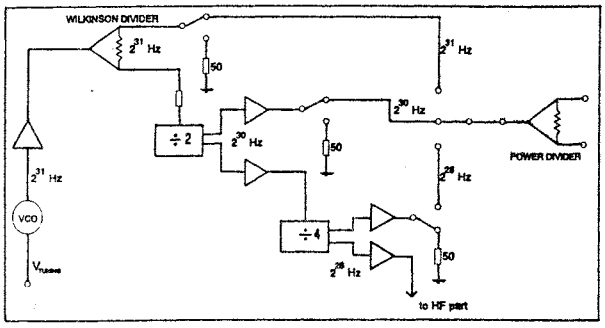


Fig 10 : Structure of the UHF part

Description of the VCO

The coaxial ceramic resonator has been chosen for its reduced dimensions, high Q factor and low cost. The oscillator is of the Colpitts type using a FET transistor. The feedback is performed by the capacitive divider made up of the internal gate-source capacitor and of the hybrid (internal + external) drain-source capacitor. The feedback elements in the FET source allow us to regulate the phase, and so guarantee the start of the oscillation.
A varactor diode is directly coupled to the resonator so that the frequency can be varied over the required range.
The resonator is extremely dependent on the environment. It is necessary to shield it in order to guarantee a correct operation

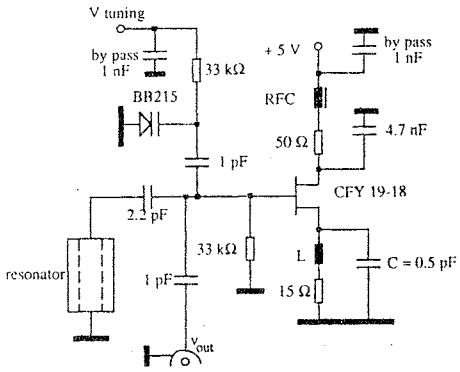


Fig 11 : Schematic of the VCO

The microstrip circuit of the UHF part is as follows :

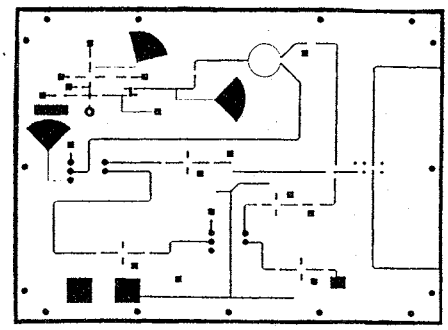


Fig 12 : Microstrip circuit for UHF part

HF part

This part has not yet been treated in the actual project. The technical characteristics of the time base have not yet been measured.

Conclusions

The direct digital frequency synthesis technique offers the following advantages compared to the more familiar PLL :
- extended synthesis range, particularly for low frequencies (here from 1 Hz to 100 MHz)
- high resolution (1 Hz)
- phase continuity and very low phase noise
- very fast frequency switching time (150 ns)

While the present DDS is limited to 100 MHz, the upper frequency range can be extended to the GHz domain by frequency mixing.

Acknowledgement

This project is financed by Swiss Foundation for Microtechnology Research (FSRM, Neuchâtel, Switzerland)

References

[1] A Digital Frequency Synthesizer
Joseph Tierney, Charles M. Rader, Bernard Gold
IEEE Trans. on Audio and Electroacoustics,
March 1971

[2] Phase Noise in Direct Digital Synthesizers
Edward M. Mattison and Laurence M. Coyle
42nd Annual Frequency Control Symposium - 1988

[3] An analysis of the Output Spectrum of Direct Digital Synthesizers in the presence of Phase-Accumulator Truncation
Henry T. Nicholas, III and Henry Samueli
41st Annual Frequency Control Symposium 1987

[4] Complex Waveform Using Direct Digital Synthesizer Techniques
Robert P. Gilmore
Qualcomm, Inc, San Diego, CA, USA

[5] Analog-Digital and Digital-Analog Conversion
Bernard Loriferne
Heyden & Son Ltd, 1982

[6] A User's Handbook of D/A and A/D Converters
Eugene R. Hnatek
Wiley Interscience 1976

[7] Microwave Transistor Oscillators
G. Hodowanec
Microwave Journal, June 1974

[8] High-Performance Wideband Cavity-Tuned Solid-State Oscillators
Edward G. Cristal, Arthur N. Woo,
Phillip G. Foster, and Ronald F. Stiglich
Hewlett-Packard Journal, July 1982

Session 3B Number 19

An Ultra Low Phase Noise Synthesizer for Radar Applications

Bruno Fognini

Siemens-Albis AG, Zurich, Switzerland

0. Abstract

This paper describes the topology of a synthesizer for advanced radar systems where ultra low phase noise and high spurious rejection combined with fast frequency switching capability are of primary importance.

The "Twin-Principle" meets these requirements: A fast SPDT-switch which combines two PLL-Synthesizers is optimised for fast frequency settling. The two PLL-synthesizers are optimised for the lowest possible phase noise. This has been achieved with a architecture that uses no dividers in the loop itself; therefore it is decreased to a simple offset-loop which acts as a tracking filter. The channel generation is done by two divider and multiplexer blocks outside of the loop.

1. Introduction

The synthesizer described below is used as the frequency generator and the local oscillator in a ground based surveillance radar. Therefore the synthesizer has to exhibit the lowest possible phase noise, as it influences directly the clutter rejection of the radar system. Long term drift is of marginal importance.

The frequency switching must be executed during the receiving deadtime, which is in the range of 5us to 10us. As this particular radar system operates in the burst mode this frequency switching occurs only once per 1ms.

A spurious rejection in the range of 80 dB is recommended for good jammer immunity.

The total number of channels is given by the available transmitting band (L-Band) and the emitted pulsewidth. A number between 40 and 50 channels is sufficent.

2. Design evaluation

There exist different methods to design a synthesizer, but whichone meets the above mentioned specifications best?

The three basic designs with their appropriate advantages and disadvantages are:

Direct Synthesis:	+ excellent phase noise + fast switching - critical spurious performance - low channel numbers + up to MW
Indirect Synthesis:	+ very good phase noise + very good spurious performance - slow switching + higher channel numbers + up to MW
Digital Synthesis:	+ very good phase noise + fast switching + very high channel numbers - bad spurious performance - low frequency band

New structures can be found by a combination of these basic designs.

One simple solution is to combine two identical synthesizers of the indirect type by a SPDT-switch. Figure 1 shows the block diagram.

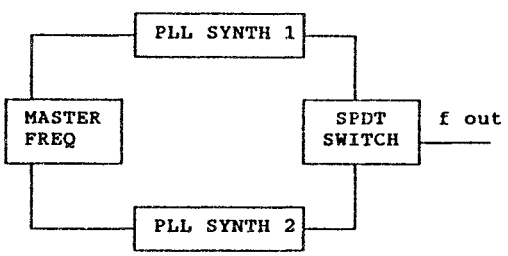


FIG.1 Blockdiagram "Twin-Synthesizer"

We have called this type of synthesizer the "TWIN-SYNTHESIZER". (Other people call it the "Ping-Pong" priciple.) It has the following attributes:

Twin Principle:	+ very good phase noise + fast switching - slow preset time + very good spurious performance + high channel numbers + up to MW - needs two identical synthesizers + gives a certain redundancy

3. Description of the "Twin-Synthesizer"

The "Twin-Synthesizer" (Fig.1) consists of two synthesizers of the indirect type (PLL). The output of both PLL-Synthesizers are fed to a SPDT switch. A following frequency multiplier converts the frequency to the desired output band.

The SPDT switch determines the fast switching time from one PLL-Synthesizer to the other. During the operation of one PLL-Synthesizer the other may change its frequency. This settling of the PLL-Synthesizer is called preset time. To achieve the fast switching time, the information for a new frequency must be given to the PLL-Synthesizer at least by this time before the frequency switching command.

To have random access to all channels the two PLL-Synthesizer must be identical. However one can divide the frequency band in an upper and a lower half and so each PLL-Synthesizer has to cover only half the bandwidth. Other groupings are possible too: i.e. PLL-Synthesizer 1 generates the even channels and PLL-Synthesizer 2 the odd channels.

3.1 The PLL-Synthesizer

The "Twin-Synthesizer" performs only as good as its PLL-Synthesizers do. Therefore special precautions must be taken.

As the switching time is not limited by the PLL-Synthesizer, it can be fully optimised to the best phase noise performance.

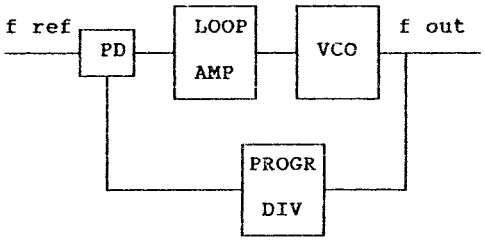


FIG.2 Blockdiagram "Indirect Synthesis"

The well known PLL arrangement (Fig.2): VCO, programmable divider, phase detector and loop amplifier is not recommended for this application as noise applied to the input of the phase detector and the noise of itself is multiplied by the ratio (:n) of the programmable divider which gives a degradation of the phase noise at the output of 20*log(n).

A better solution is the "Offset-Loop" where the divider ratio gets smaller but for ultra low noise application it is still inacceptable. So it is evident: provided the phase noise of f off is ideal the best solution is a design which needs no divider in the feedback loop!

Figure 3 shows such an arrangement. There are no dividers in the loop itself and therefore the loop acts as a tracking filter.

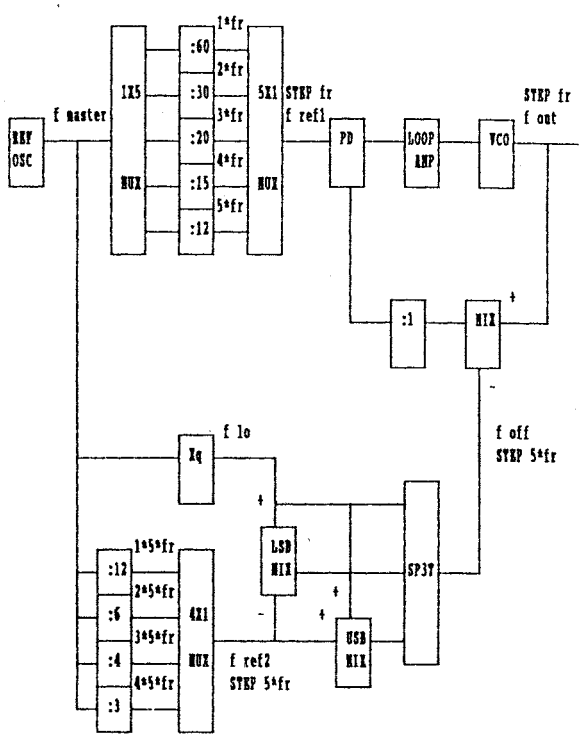


FIG.3 Blockdiagram PLL-Synthesizer

The output frequency f out of the offset loop is the sum of f ref and f off:

$$f_{out} = f_{ref} + f_{off}$$

In this architecture the reference frequency f ref is no longer fixed. It is:

$$f_{ref1} = n * fr$$

where

$$n = 1,2, \ldots n_{max}$$
$$fr = f_{master} / k1$$
$$= \text{smallest frequency step}$$

The constant factor k1 is chosen in such a manner that the division k1 / n becomes for all n an integer. Then the dividers are realizable and the frequency steps become equidistant.

This reference generation is done by a divider and multiplexer block and in our system is n max = 5.

The offset frequency f off is not fixed as well. The following formula determines f off:

$$f_{off} = f_{lo} + f_{ref2}$$

where the frequency f ref2 is:

$$f_{ref2} = m * n_{max} * fr$$

and where

$$m = 1,2, \ldots m_{max}$$

The two formulas combined give:

$$f_{off} = f_{lo} + m * (n_{max} * f_r)$$

where

$m = -m_{max}, ..-2,-1,0,+1,+2, ..+m_{max}$
$n_{max} * f_r = f_{ref1\ max}$
$f_{lo} = q * f_{master}$
q = appropriate factor to get the desired frequency band

As m is positive, negative or zero the addition of the two terms is done by a single sideband mixer with a following switch where the lower or the upper sideband or only f lo can be coupled out.

The generation of f ref2 is done by a divider and multiplexer block too and in our system is m max = 4.

The generation of f off is a synthesizer of the direct type with moderate spurious performance.

3.2 Discussion of the PLL-Synthesizer

As a summery one can see that the output frequency is a function of the master frequency, the constants and the different divider ratios. It is:

$$f_{out} = f_{master} * (n * 1/k1 + q + m * n_{max} * 1/k1)$$

and rearranged

$$f_{out} = f_{master} * [(n + m * n_{max}) * 1/k1 + q]$$

In this mathematical deduction the output frequency is derived from only one single master frequency. An advantage is that the output is coherent to the master. A more important characteristic is that the offset mixer of the loop generates only spurious which are a multiple of fr and therefore no filters are needed in the f off path. All the unwanted signals are filtered out by the loopfilter.

A disadvantage is that the output frequency band is linked with the frequency step. To overcome this problem the f lo could be derived from a second oscillator, but special care must be taken that no spurious generated in the offset mixer of the loop fall into the loop bandwidth as they are not filtered out.

The total number of channels that generates one PLL-Synthesizer with this arrangement is:

$$\text{total channels} = [n_{max} * (2 * m_{max} + 1)] = 45$$

For the "TWIN-SYNTHESIZER" the total number is either 45 or 90 channels depending on the grouping.

4. Description of the components

The above described PLL topology was chosen because of the phase noise degradation by the dividers in the loop. As we have eliminated them we have achieved a constant phase noise over all channels and new components determine the phase noise of the loop, i.e. the VCO, the master oscillator and the phase detector.

4.1 The Voltage Controlled Oscillator

At this point one should determine what is the best operating frequency of the phase lock loop. We have found that an operating frequency around 1 GHz is the best one because VCO's can be built best there.

Our VCO is a LC-Colpitts oscillator. As inductance we use a coaxial ceramic resonator operated below its resonant frequency. This type of resonator leads to a high loaded Q of the resonace circuit and therefore optimum phase noise. Another advantage is that this resonator can be operated with a high power level which allows a low noise floor.

For a VCO at 1 GHz and a tuning bandwidth of 10% the phase noise measurements are:

- 85 dBc/Hz @ 1 kHz offset
-160 dBc/Hz noise floor

4.2 The Master Oscillator

The master oscillator operates around 100 MHz. As resonant element we use a 5th overtone AT cut quartz. It vibrates in the series mode and is used in a two port configuation. This allows us to control very precisely the amplitude and the phase in the oscillator loop.

The power dissipation in the quartz is high (near 1 mW). This gives a very low phase noise which is very important as it will be multiplied up to the output frequency of the synthesizer. Phase noise measurements are:

- 150 dBc/Hz @ 1kHz offset
- 168 dBc/Hz noise floor

To achieve this low phase noise one must be sure that the phase noise of the quartz itself is better than the above mentioned value.

4.3 The Phase Detector

The phase detector is another component which affects the phase noise in low noise loops as it modulates directly the VCO within the loop bandwidth. The lower the noise the better the loop. Therefore we have developed two detectors: a phase-frequency detector and a "pure" phase detector. During the switching from one frequency to another the phase-frequency detector is operating. Afterwards it is switched to the "pure" phase detector. The reason for this switching is that the "pure" phase detector has the better phase noise (approximately 7 dB).

These phase detectors are part of an ASIC in ECL technology which was especially designed for this PLL-Synthesizer.

4.4 The ECL ASIC

The ECL ASIC takes in not only the phase-frequency and the "pure" phase detector but the two divider and multiplexer blocks too. Futher it executes several auxiliary functions i.e. the timing for the phase detector switching or the voltage doubling for the tuning varactor.

The maximum input frequency is 650 MHz. It operates from a single +5 V power supply and power dissipation is 2.5 W. Its evident that it needs a special cooling.

5. Measurement Results

As this synthesizer was designed for the lowest possible phase noise this measurement is the most interesting one. With a loop bandwidth in the PLL of approximately 200 kHz we have got the optimum phase noise.

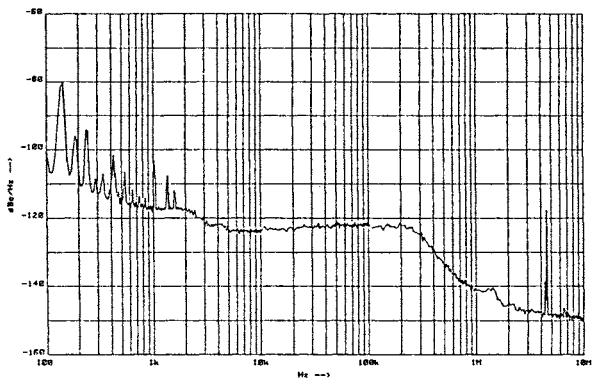


FIG.4 Phase Noise of the "Twin-Synthesizer"

Figure 4 shows a plot of it. The measurement was taken at a frequency in the L-Band (approximately 2 GHz) with the PLL-Synthesizer followed by a frequency doubler which degraded the phase noise by 6 dB. As measurement set up we used the carrier noise test set HP 11723C and the HP 8663 synthesizer.

The two "humps" at 2 kHz and 4 kHz need an explanation: They are caused by the 640 MHz reference frequency used in the measurement set up. This reference is generated by the multiplication of a very low noise 10 MHz source and quartz filters at 40MHz and 160 MHz and a SAW filter at 640 MHz reduce the multiplied noise. The two "humps" are the response of the quartz filter characteristic.

The "true phase noise" at 2 GHz and at 1 kHz offset is -123 dBc/Hz. Due to the frequency multiplication the noise floor is at -150 dBc/Hz. For radar applications the figure of the noise floor is not very critical. Here counts the phase noise close to the carrier. The longer the range of the radar system the closer to the carrier the phase noise must be good.

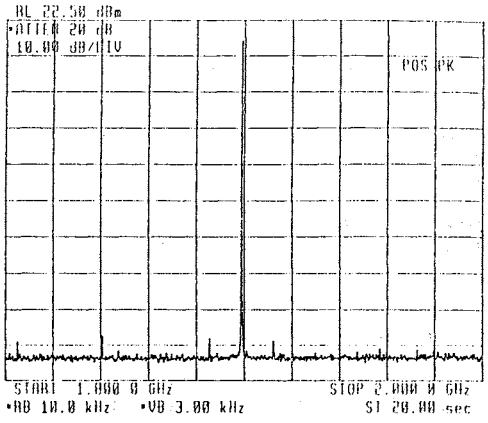


FIG.5 Spurious of the "Twin-Synthesizer"

Figure 5 shows the typical spurious performance of the "TWIN-SYNTHESIZER" at 2 GHz. The value of -80 dBc can only be achieved in a topology where all signals are derived from one single master oscillator.

The switching time is the time which elapses from the switching command till the full stability of the new frequency has reached which is in our system 65 dB of clutter rejection and which corresponds to a final phase error of 0.03 degrees.

In the "TWIN-SYNTHESIZER" we have two different switching times: First, the switching of the SPDT-switch from one PLL-Synthesizer to the other. This time is less than 10 us. Second, the switching time of the PLL itself. This time is 750 us max depending on the frequency step (see Figure 6).

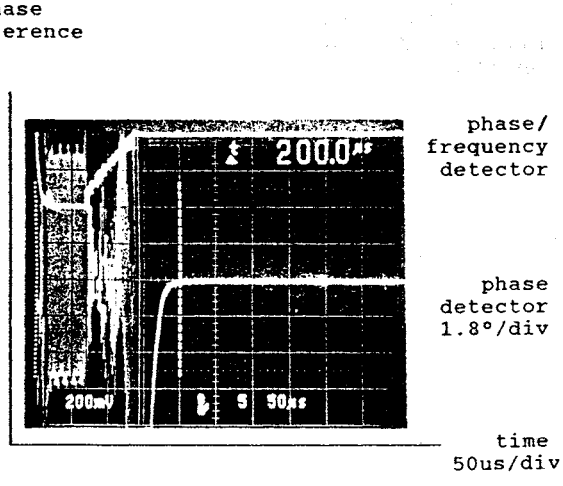


FIG.6 Frequency switching of the PLL-Synthesizer (Preset time)

With conventional methods it is very hard to measure dynamically phase differences down to 0.01 degrees. Therefore we use the signal processing of a radar receiver and determine the moment when the desired stability (clutter rejection = 65 dB) is reached.

6. Mechanical

A fully operational "TWIN-SYNTHESIZER" needs two PLL-Synthesizers, a master reference, the SPDT -switch and the output stage with the frequency multiplier. All this blocks fit in three shielded housings. The dimensions of one housing are: 240 mm x 160 mm x 40 mm.

The power consumption of the whole "TWIN-SYNTHESIZER" in the L-Band is 50 W.

7. Conclusion

A synthesizer topology, the "TWIN-SYNTHESIZER" has been described which is a combination of indirect and direct synthesizers. The direct type was optimised for fast switching and the indirect type for minimum phase noise. A PLL achitecture which has no divider in its loop and well designed building blocks i.e. VCO, master oscillator and phase detector have shown the expected ultra low phase noise.

8. References

[1] Vadim Manassewitsch, "Frequency Synthesizers Theory and Design", Second Edition, John Wiley & Sons, Inc., New York 1980.

[2] Roland C. Stirling, "Microwave Frequency Synthesizers", Prentice-Hall, Inc., New Jersey, 1987

[3] I.M.H. Williamson, P.J. Webb, P.W. Chilvers, "An Advanced Indirect Microwave Synthesizer", Proc. Military Microwaves 1990, MEP Ltd., London pp220-228

[4] D.A. Williams, T.C. Aliwell, "A Low Noise Microwave Frequency Synthesizer Design for a High Vibration Environment", Proc., Military Microwaves 1990, MEP Ltd. London pp229-235

Session 3B Number 20

Phase Noise of Regenerative Frequency Dividers

E. Rubiola* M. Olivier† J. Groslambert‡

Abstract

An analysis of noise related problems, based on an experimental description of the mixer, will be given for regenerative frequency dividers. Measurement techniques will be proposed and discussed. In some cases a noise reduction technique can be used. This is predicted by the theory and experimentally observed.

1 Introduction

Frequency dividers seem to have recently become again an interesting field of research. In frequency metrology, since research on frequency sources goes towards higher frequencies, low noise dividers are of increasing importance, especially if phase locking techniques are used. But the *noise* behaviour of dividers is less known than most of the other aspects in frequency synthesis.

Analog dividers show some advantages, compared with digital ones, due to their higher input frequency and lower noise. The regenerative divider seems to be a very low noise scheme, very likely the best one.

In the last few years, monolithic implementations of the regenerative divider have been proven to be feasible, by resorting to various technologies [1,2,3], thus making this scheme more attractive.

Although regenerative schemes first appeared in the literature a long time ago [4,5], and were better explained in more recent papers [6,7], only a few experimental results dealing with phase noise have been published [8,9]. No theory or euristic explanation of the phenomena has been given till now.

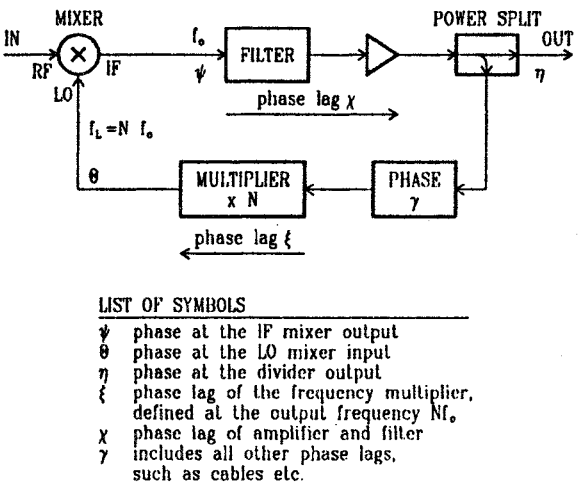


Figure 1: Block diagram of the regenerative frequency divider.

2 Frequency dividers operation

The block diagram of the regenerative frequency divider by N+1 is shown in Fig. 1, where most of the symbols used in this paper are also defined. The mixer is a nonlinear 3-port device; the combined action of mixer and filter is such that at the input of the amplifier the frequency is:

$$f_o = f_R - f_L \qquad (1)$$

Because of the multiplier in the feedback path, $f_L = N f_o$, the LO signal regenerates the IF one. Consequently, a synchronous mode of operation is estabilished, with $f_o = f_R/(N+1)$.

*Politecnico di Torino, Torino, Italy
†LPMO Besançon, France
‡LPMO Besançon, France

The signals at the mixer ports, *radio frequency* (RF), *local oscillator* (LO) and *intermediate frequency* output (IF) are the following, respectively:

$$\text{RF:} \qquad x_{R(t)} = A_R \cos(2\pi f_R t) \qquad (2)$$

$$\text{LO:} \qquad x_{L(t)} = A_L \cos(2\pi N f_o t + \theta) \qquad (3)$$

$$\text{IF:} \qquad x_{o(t)} = A_o \cos(2\pi f_o t + \psi) \qquad (4)$$

where A_o is a function of A_R and A_L. Taking the RF signal as the phase reference, at the output of an ideal mixer (i.e. a perfect analog multiplier) the phase is $\psi = -\theta$.

The closed loop stability constraints, unity gain and zero phase ($+2k\pi$, with integer k), are easily met. The desired gain is to be ensured by saturation either in the amplifier, or in the mixer, or in the frequency multiplier. The parameter θ is a degree of freedom of the system, and it will satisfy the phase condition if jumps multiple of 2π can be avoided.

Asynchronous modes of operation, either alone or in conjunction with a chaotic behaviour, are possible, as discussed in a qualitative but clear way in [10].

3 Phase noise in frequency dividers

In low noise applications it is convenient to use high level signals, of the order of 10 dBm. In this condition mixers are strongly perturbed by harmonic distortion, responsible of the *phase gain* phenomenon. This problem will be discussed, with some experimental techniques, in order to evaluate the phase noise of the frequency dividers. The case of the halver will be discussed in detail, giving some hints for the extension to different division ratios.

3.1 Mixers in the frequency halver scheme

When two signals of frequencies f_R and f_L are supplied at the inputs of a mixer, the output spectrum contains many signals, each one with its own amplitude, whose frequencies are:

$$f_{n,m} = nf_L + mf_R \qquad (5)$$

where n and m are integers. In order to build a divider, m and n must have opposite signs.

Since the two input signals are synchronous, many harmonics give the desired output frequency. Consequently the signal $x_{o(t)}$ comes from the sum of all the *mixing modes* satisfying the following relations:

$$n = 1 + 2h \qquad (6)$$

$$m = -h \qquad (7)$$

where h is integer. A further selection on n and m arises from the symmetry of the mixer, if there is any. In the double balanced devices, used in the experimental part of this paper, the even harmonics are strongly attenuated by the simmetry of both inputs.

The IF signal, at the frequency f_o, is given by the following sum of harmonics:

$$x_{o(t)} = \sum_{n,m} I_{n,m} \cos(2\pi f_o + \beta_{n,m}) \qquad (8)$$

I and β are the amplitude and phase of the n, m mixing mode. $I_{n,m}$ decreases rapidly with $|n|$ and $|m|$, with a law depending on the mixer and on its working conditions.

The terms $\beta_{n,m}$ come from the LO signal multiplied by n, and consequently any phase shift is multiplied by n too. This can be formulated as:

$$\beta_{n,m} = n\theta + \alpha_{n,m} \qquad (9)$$

where $\alpha_{n,m}$ is the *phase offset* of each mixing mode.

In a simplified model of the mixer, taking into account only the 1st and 3rd harmonics, only two mixing modes give contribution to $x_{o(t)}$: $f_R - f_L$ and $3f_L - f_R$. If the LO signal is shifted by $\mathrm{d}\theta$, these harmonics are shifted by $-\mathrm{d}\theta$ and $+3\mathrm{d}\theta$ respectively. The output phase shift is determined by the amplitude and phase relations between these two harmonics.

Two extreme situations, shown in Fig. 2, are to be considered:

Case A the signals are in phase, the output amplitude is maximum, and the resulting signal is rotated by $\mathrm{d}\psi$, being $|\mathrm{d}\psi| < |\mathrm{d}\theta|$,

Case B the signals are opposite, the output amplitude is minimum, but the effect of $\mathrm{d}\theta$ on the output phase ψ is stronger than $\mathrm{d}\theta$.

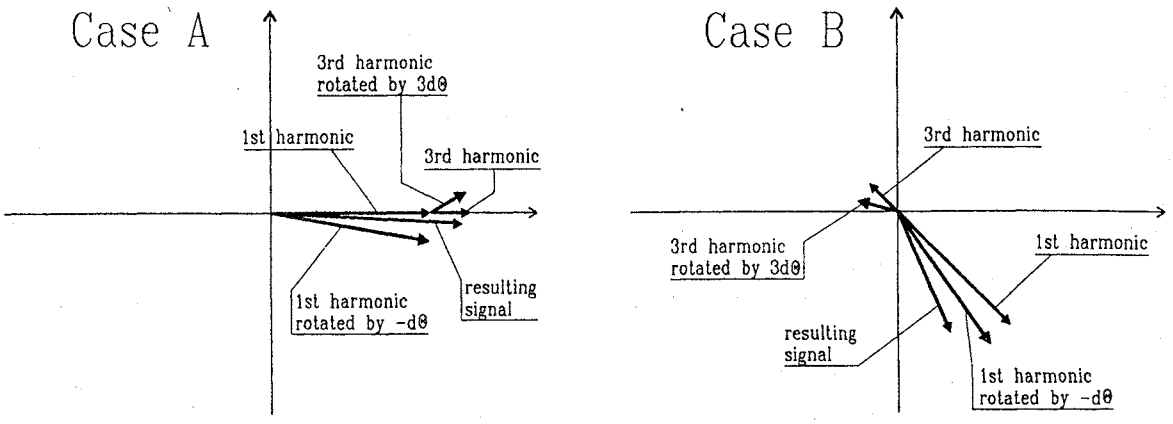


Figure 2: Phasor representation of the two main signals contributing to the output frequency f_o in a halver. The effects of a perturbation $d\theta$ imposed to the LO signal are considered for the two extreme situations.

The main conclusion taken from this model is that the mixer shows a phase gain $G = d\psi/d\theta$, defined for small perturbations, such as the phase noise.

Although G is usually a negative quantity, in some circumstances it can assume positive values: referencing to the phasors of Fig. 2, it happens that $G > 0$ in the case A if $3|I_{3,-1}/I_{-1,1}| \ll 1$. In an ideal mixer, without any contribution of higher order harmonics $G = -1$ always holds.

In the two-phasors restricted model G is maximum at $\theta = \theta_A$, when the output amplitude is maximum too. In the opposite condition, $\theta = \theta_B$, both G and output amplitude A_o are minimum. In the general case, where all the harmonics are taken into consideration, the phase gain G *tends* to be maximum or minimum *near* the above described conditions.

In principle, there is no difference in the resulting behaviour if the harmonics are actually originated inside the mixer, or come from the outside. In this way we can take into account other distortions.

3.2 Mixer measurement techniques

Mixers are complex devices, not well described in the literature in the case of synchronous signals. Thus, we have chosen an experimental characterization, following two different methods.

Whatever method is chosen, the mixer *must* be measured in the same impedence matching conditions of the final application. This constraint is particularly relevant for the output, given the many signals at other frequencies than f_o that can be reflected; the stronger one of these, at frequency $(2N+1)f_o$, has about the same power as $x_{o(t)}$.

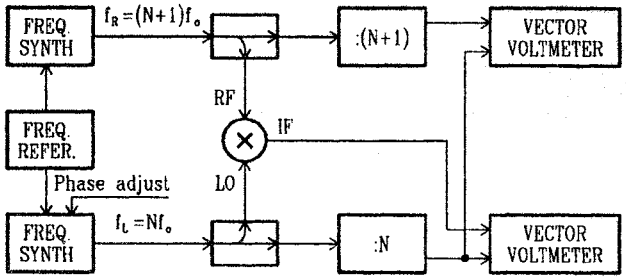


Figure 3: Phase gain measurement technique.

The methods here proposed are equivalent. The first one has some advantages of simplicity, while the other one is suitable to a wider frequency range.

3.2.1 First method

The mixer is driven by two synthesizers whose frequency ratio is $(N+1)/N$, phase locked to a common reference. The measurement is performed with two vector voltmeters, as shown in Fig. 3, by shifting the phase of one synthesizer output. This scheme simulates what happens in a mixer used in a frequency divider. The result is a plot in which the IF amplitude A_o and phase ψ are given as functions of the input phase shift θ. From this data the G is easily evaluated.

3.2.2 Second method

The signal $x_{o(t)}$ can be calculated knowing the amplitudes $I_{n,m}$. This method needs the assumption that all the terms $\alpha_{n,m}$ of equation (9) are zero, and consequently all the mixing modes are in phase for $\theta = 0$. The mixer is driven by two synthesizers whose frequencies are $(N+1)f_o$ and $Nf_o(1+\epsilon)$. The frequency ratio is *almost* $(N+1)/N$, with a small relative difference ϵ. The IF output is measured around the frequency f_o with a spectrum analyzer. The mixing modes that give contributions at frequency f_o when the frequency ratio is $(N+1)/N$ are now split around f_o and can be observed individually. This method relies on the hypothesis that the mixer is a wideband device, therefore a small relative frequency offset ϵ allows measurement in the same working conditions of the divider, with a negligible difference. On the spectrum analyzer, each mixing mode appears as a single spectral line whose power is proportional to $I^2_{n,m}$.

3.2.3 Comparison between the methods

A double balanced mixer TFM-10514 has been tested, simulating a frequency halver. All the parameters (A_o, ψ and G) have been measured with the first method, and evaluated with the second, taking into account all the harmonics down to -50 dB below the main signal. The experimental conditions were the following:

LO input	235 MHz	+10 dBm
RF input	470 MHz	+14 dBm

adding 20 kHz to the LO frequency when the second method is used.

From the results, reported in Fig. 4, the evaluated average amplitude is 0.2 dB higher than the measured one. Subtracting this bias, the amplitude difference is within ± 0.3 dB.

Comparing the two plots of ψ, there is a systematic difference of 13.5°, with variations of $\pm 3°$ (peak).

The measured and calculated phase gains G differ by about 0.2, excluding the narrow regions around the minima, but these regions are not meaningful for our purposes because the amplitude A_o has its minimum too: this extreme condition should be avoided in the divider because of the difficulty in keeping a stable dividing operation.

3.3 Expected phase noise

The noise is generated by the electronic circuits contained in the divider. We take into consideration only the amplifier because, in our experience, this is the most significant source of phase noise; other noise sources can be included in the model of the amplifier.

Being interested in the phase noise *near* the carrier, the measurement bandwidth of the phase noise will be assumed to be small compared to the divider bandwidth and to f_o.

Phase noise at the amplifier input can be considered as a small perturbation $\mathrm{d}\chi$ to the amplifier phase lag χ. Because of the narrow band hypothesis, a static analysis is valid. The period of the highest Fourier frequency is therefore long, in comparison with the signal transit time around the loop.

The closed loop phase condition of the divider (see Fig. 1, in which the multiplier is suppressed) is:

$$\psi + \chi + \gamma = \theta \tag{10}$$

and the output phase η is given by:

$$\eta = \theta - \gamma \tag{11}$$

The mixer output phase shift $\mathrm{d}\eta$ can be found by differentiating equation (10) and assuming $\mathrm{d}\gamma = 0$ because γ is constant. Thus:

$$\mathrm{d}\psi + \mathrm{d}\chi = \mathrm{d}\theta \tag{12}$$

From equation (11) we get $\mathrm{d}\eta = \mathrm{d}\theta$. Combining the above reported results, and remembering that $\mathrm{d}\psi = G\mathrm{d}\theta$, we find:

$$\frac{\mathrm{d}\eta}{\mathrm{d}\chi} = \frac{1}{1-G} \tag{13}$$

When the noise statistic is considered, equation (13) leads to the final result:

$$S_\eta(f) = \left[\frac{1}{1-G}\right]^2 S_\chi(f) \tag{14}$$

where $S_\eta(f)$ and $S_\chi(f)$ are the phase noise one-sided power spectral densities for η and χ. $S_\chi(f)$ comes from the device characterization, while $S_\eta(f)$ represents the phase noise generated by the divider.

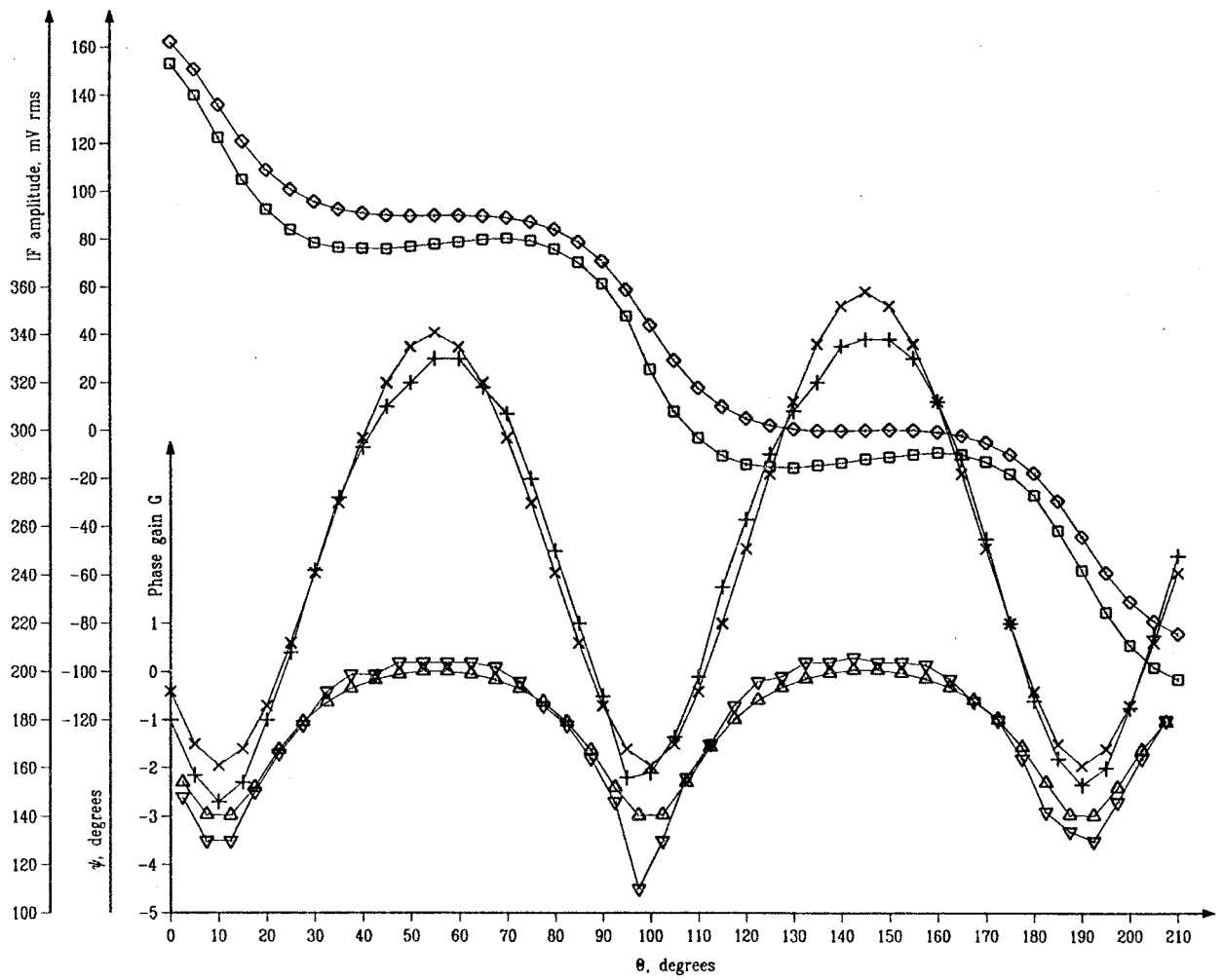


Figure 4: Comparison between measured (1st method) and evaluated (2nd method) parameters of a double balanced mixer. × amplitude calculated + amplitude measured ◇ phase calculated □ phase measured ▽phase gaincalculated △phase gainmeasured

We obtained this result by taking into account the instantaneous values of the phase shift originated by the noise. There is no reason, in principle, to restrict its validity to some particular types of phase noise, such as the flicker or the white floor.

3.4 Extension to different division ratios

Analyzing the phase noise of a divider by $N+1$ ($N \neq 1$) the same approach shown in section 3.3 can be used, with two main differences:

- the contribution of the frequency multiplier must be taken into account,
- the output phase noise instantaneous value $d\theta$ is multiplied by N at the LO input of the mixer.

Since multiplier and amplifier are separated devices, their contributions to $S_\eta(f)$ are statistically independent and can be added.

The measurement techniques described in section 3.2 are suitable, with minor changes.

Comparing the other dividers with the halver, higher harmonics give contribution to the IF output frequency. Since harmonics power decreases rapidly with $|n|$ and $|m|$, a weaker phase gain phenomenon is expected.

Some experiments have been done in order to evaluate G with the same mixer and in condition similar to those of section 3.2.3, but simulating a divider by 3. The phase gain does not depart significantly from -1.

4 Low noise design

If the mixer can be considered an ideal device, thus $\psi = -\theta$, the equation (14) leads to a noise level $S_{0\eta}(f) = S_\chi(f)/4$. This is an intrinsic noise reducing property of the regenerative divider.

The classical approach is to tune the phase lag of the dividers for the maximum output amplitude at the central input frequency. This choice brings some additional advantages: the operation bandwidth is larger, loop stability and startup conditions are easily met, a chaotic behaviour is rare, and so on. However, remembering the results given in section 3.3, when the amplitide A_o reachs its maximum, G is close to its maximum too. According to equation (14), this turns out to be the *worst* case for spectral purity: the output phase noise is $S_\eta(f) > S_{0\eta}(f)$.

Conversely, an improved version of the divider is possible choosing the minimum G, such that the output phase noise be $S_\eta(f) < S_{0\eta}(f)$. Unfortunately all the previously described advantages will be lost; practical considerations suggest to avoid the extreme condition.

The main way of acting on G is by adjusting γ. This *detunes* the divider out of the maximum amplitude. Moreover, it is possible to: (i) act on the levels A_R and A_L, thus affecting all the harmonics amplitudes; (ii) alter the impedence matching, which affects the mixing modes *not* leading to f_o at the IF output, consequently the power reflected by the filter enhances the power of the perturbing modes $(n, m \neq -1, 1)$.

From the previous considerations, two alternative design criteria seem to make sense:

- keep the harmonics small, in order to ensure $G \simeq -1$ for any θ. The output phase noise is close to $S_{0\eta}(f)$. This approach is recommended for wideband applications.
- choose a working point with a strong harmonic distortion in order to achieve $G < -1$, and thus improve the noise performances.

The first implies $|I_{n,m}/I_{-1,1}| \ll |1/n|$ for all the mixing modes but $n = -1$, $m = 1$.

The second style needs a careful design, built around a strongly saturated mixer. The condition $G < -1$ can be obtained only in a subrange of the possible values of θ, in partial conflict with the other design rules. Unless a cumbersome feedback network is used, a compromise can be achieved only in a narrow band, thus making this approach attractive for extremely low noise applications.

Any configuration in which the mixer is strongly perturbed by the harmonics, and no attention is paid to G, should be definitely avoided.

5 Experimental results

Some experiments have been done with two halvers, working at 470 MHz input frequency $2f_0$. The aim was twofold, to build a low noise

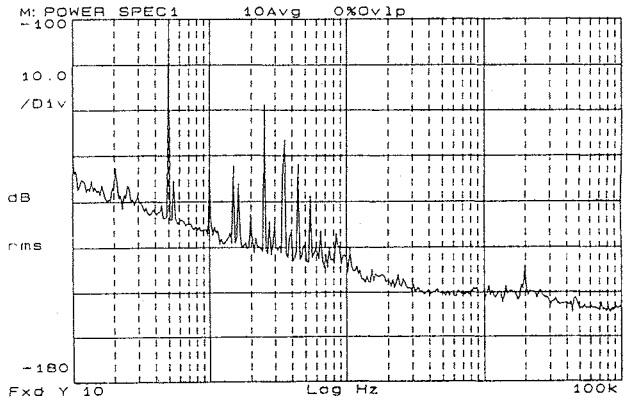


Figure 5: Phase noise for two dividers tuned for the maximum output amplitude (classical approach). Values are given in dB below 1 rad^2/Hz.

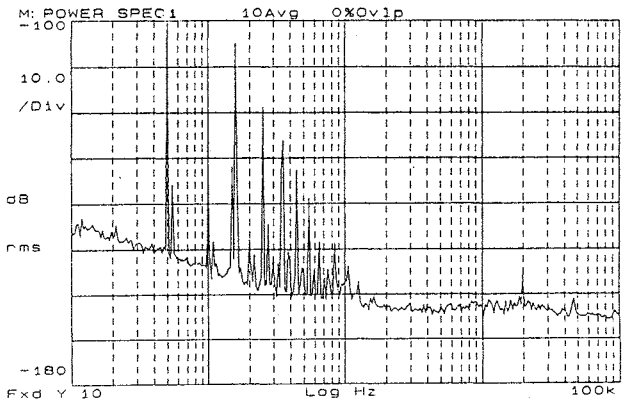


Figure 6: Phase noise for two dividers tuned for minimum noise, out of the maximum output amplitude working point (new approach). Values are given in dB below 1 rad^2/Hz.

divider with usual parts, and to get an experimental evidence of the noise reduction method given in section 4.

The significant components in a divider are the following:

device	type	manufacturer
mixer	TFM-10514	Mini Circuits
amplifier	AMP-77	Mini Circuits
filter	bw $\simeq$ 10 MHz	built in laboratory

Two situations are compared, in which the divider is:

- tuned for the maximum output amplitude,
- detuned adding to the phase lag γ a delay equialent to 50°. This value has been found with some attempts.

Since in our dividers most of the phase noise originates in the amplifiers, we have selected two

Reference	flicker	white
[8,9]	-125 dB f^{-1} ¶	-161 dB f^0 ¶
our divider (tuned)	-130 dB f^{-1}	-165 dB f^0
our divider (detuned)	-140 dB f^{-1}	-175 dB f^0 §
¶equivalent noise at $f_o = 235$ MHz		§inferred

Table 1: Comparison of the phase noise for one regenerative frequency divider. All values are in dB below 1 rad^2/Hz.

equally noisy devices, verifying that their phase noise spectral density $S_\chi(f)$ remains the same, in spite of the amplitude variation due to the detuning procedure. We have chosen as noise test source the phase flicker of the amplifiers (-125 dB rad^2/Hz f^{-1}) because it is the only type of noise significantly higher than the limit of the phase noise measurement equipment (-139 dB rad^2/Hz f^{-1}).

The phase noise spectrun has been measured comparing two dividers driven by the same source. The results are shown in Figg. 5 (tuned for the maximum amplitude) and 6 (detuned).

Comparing Figg. 5 and 6, a reduction of the flicker noise is evident. Taking into account the contribution of the test equipment, that raises by 0.3 dB the flicker of Fig. 5 and 1.8 dB the flicker of Fig. 6, the actual improvement is 10 dB.

As for the white phase noise, we expect a similar result. A qualitative improvement is evident, but a quantitative evaluation is not possible; since for the white noise the plot of Fig. 6 overlaps exactly to the test equipment limit, one can think that the noise generated by the dividers is at most 5–6 db smaller than the equipment limit. If this hypothesis is true and the thermal noise is adequate, the improvement is the same as that obtained for the flicker.

The noise of one divider can be obtained subtracting 3 dB from the plot of Fig. 6, since the circuits are the same and the components have been selected equally noisy.

Only two papers, reporting similar noise data, have been found in the literature [8,9]. A comparison with our results is given in Tab. 1. The conclusion is that the divider here described shows *state of the art performances when tuned* for the maximum amplitude, with a significant improvement if detuned as described.

As for different division ratios, a divider by

three has been tested. Unfortunately we could not find the experimental conditions for a phase gain $G \neq -1$, and consequently the noise reduction method failed. But this seems a practical problem, which can be solved in further experiments.

6 Conclusions

Regenerative dividers show very low phase noise, if carefully designed.

In the case of the halver a preliminary theoretical analysys of the noise related problems has been given, which can be extended to more complex cases.

The key idea is that, in a 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 the phase gain phenomenon, which is particularly evident in strongly saturated mixers. The divider is in a certain sense similar to a linear control system, in which the loop gain attenuates the phase perturbations arising inside the loop.

In some case it is possible to reduce the noise by choosing an unusual design approach. This is forseen by the theory and experimentally observed.

References

[1] Kunihiko Kanazawa, Masahiro Hagio, Masaru Kazamura, Gota Kano, "A 15 GHz single-stage GaAs dual-gate FET monolitic analog frequency divider with reduced input treshold," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-36 no. 12, December 1988, pp. 1908–1912.

[2] R. H. Derksen, H. M. Rein, K. Wörner, "Monolitic integration of a 5.3 GHz regenerative frequency divider using a standard bipolar technology," *Electronics Letters* vol. 21 no. 22, 24th October 1985, pp. 1037–1039.

[3] T. Ohira, K. Araki, T. Tanaka, H. Kato, "14 GHz band GaAs monolithic analogue frequency divider," *Electronics Letters* vol. 21 no. 22, 24th October 1985, pp. 1057–1058.

[4] R. C. Miller, "Fractional-frequency generators utilizing regenerative modulation," *Proc. IRE*, vol. 27, July 1939, pp. 446–457 (quoted by S. V. Ahamed et al.).

[5] R. L. Fortescue, "Quasi stable frequency dividing circuits," *J. Inst. Electrical Engineers (U.K.)*, vol. 84, 1939, pp. 693–698 (quoted by S. V. Ahamed et al.).

[6] G. Immovilli, G. Mantovani, "Analysis of the Miller Frequency divider by two in view of applications to wideband FM signals," *Alta Frequenza* vol. 17 no. 11, November 1973, pp. 313–323.

[7] R. C. Harrison, "Theory of regenerative frequency dividers using double balanced mixers," *IEEE Trans. on Microwawe Theory and Technology*, MTT-S Symp. Digest vol. 1, June 1989, pp. 459–462.

[8] M. M. Driscoll, "Phase noise performance of analog frequency dividers," *Proc. 43th Annual Frequency Control Symposium*, 1989, pp. 342–348.

[9] M. M. Driscoll, "Phase noise performance of analog frequency dividers," *IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control*, vol. UFFC-37 no. 4, July 1990, pp. 295–301.

[10] S. V. Ahamed, J. C. Irving, H. Seidel, "Study and fabrication of a frequency divider-multiplier scheme for high efficiency microwave power," *IEEE Trans. on Communications* vol. COM-24 February 1976, pp. 243–249.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 3B Number 21

A NARROW BAND FREQUENCY SYNTHESIZER WITH A NCO CIRCUIT

Peter Schumacher

Observatoire de Neuchâtel, CH-2000 Neuchâtel/Switzerland

Abstract

Frequency synthesizers are principal building blocks of precise time and frequency systems. This paper gives a description of a 100 MHz frequency synthesizer with narrow band ($1 \cdot 10^{-5}$) frequency deviation. This synthesizer can also be used as a phase stepper. A similar design has been used for the synthesis of the 5.316 MHz frequency for a Rubidium frequency standard.

The main components are a NCO (numerically controlled oscillator), a high speed D/A-converter, a microprocessor and low noise crystal oscillator with a PLL circuit. Frequency synthesis with these elements results in a very compact synthesizer, that has a low phase noise. Phase noise as a function of frequency deviation is discussed and phase noise mensurements are presented.

1. Introduction

Digital synthesizers are the principal building blocks of precise time and frequency systems. Direct frequency synthesizers which synthesize waveforms by digital techniques have become increasingly popular by the event of large scale integration circuits. Narrow band frequency synthesizers are used to correct the output frequency of different frequency standards. The synthesizer we have designed is used in a Rubidium frequency standard to correct an initial accuracy offset due to the fabrication process of the gas mixture in the Rubidium cell and the Rubidium lamp.

2. Synthesizer for Rubidium frequency standard

The principle schematics of a Rubidium frequency standard is shown in fig. 1. The natural frequency of the atoms in a Rubidium vapor cell has a nominal value of 6834.684 MHz. This frequency is generated by multiplication from a 20 MHz voltage controlled crystal oscillator. A multiplication factor of 9 x 38 gives a frequency of 6840 MHz. At the 180 MHz level the multiplier is phase modulated with a 5.316 MHz synthesizer frequency. The result is a subtraction of the 5.316 MHz frequency at the 6840 MHz level.

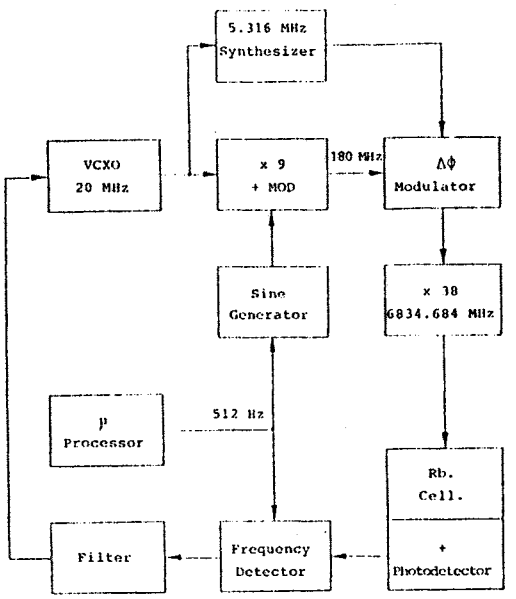


Fig. 1: Principle schematics of Rb frequency standard

The frequency of 5.316 kHz is obtained by a direct digital synthesizer (fig. 2). The main elements are a 20 MHz input X-tal oscillator (VCXO), a numerically controlled oscillator (NCO), a single chip processor/controller with a serial RS 232 input for remote control of the frequency setting and a digital analogic converter (DAC). The NCO is a standard large scale integration circuit from Stanford Telecom, the DAC is a 160 MHz device from Sony.

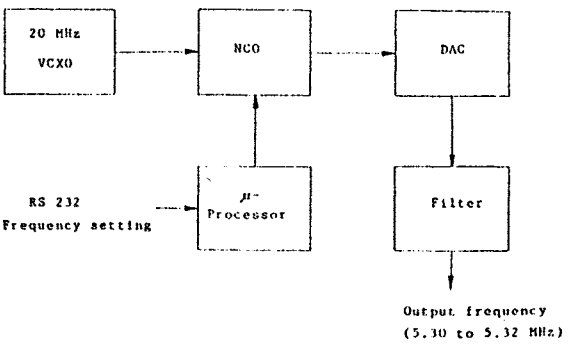


Fig. 2: Frequency synthesizer for Rb frequency standard.

The internal operation of the NCO is represented in fig. 3. At each clock cycle, the number that is stored in the Δ-phase register is added to the previous value of the phase accumulator. The output frequency is proportional to the hexadecimal value in the Δ-phase register. The numbers in the accumulator are generating a digital step function that is translated into a digital sine function by a ROM look up table. The 10 most significant bits of the NCO are translated to the DAC whose output is a truncated sine function (fig. 4).

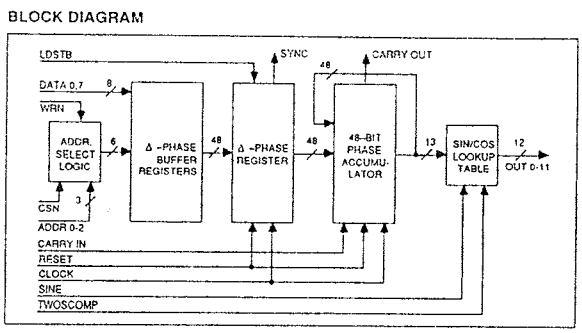


Fig. 3: Internal operation of NCO

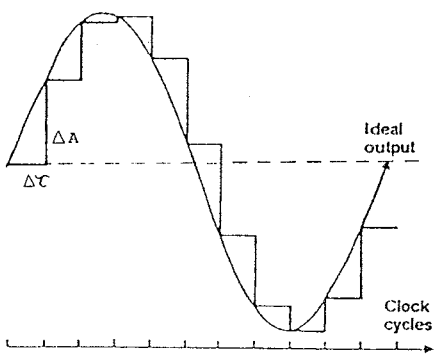


Fig. 4: Typical output waveform of sine output direct digital synthesizer.

The processor program converts the basic BCD frequency settings that are sent over the serial RS-232 input into a hexadecimal 48 bit number that is transmitted to the NCO.

The output signal has been measured with a Fourier signal analyzer. The result is represented in fig. 5. With a 10 bit DAC all spurious harmonics are below the 80 db limit of the analyzer. This limit should be sufficient for a Rubidium frequency standard. Actually there is a slight influence of the synthesizer to the output frequency of the Rubidium standard due to an interference between the 512 Hz interrogating frequency and some spurious harmonics of the synthesizer. This interference causes a nonlinear behaviour of specific frequency settings at a 10^{-11} level.

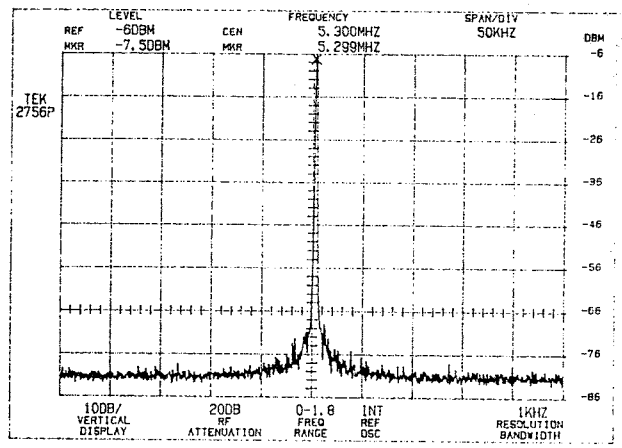


Fig. 5: Spectrum with 10 bit DAC. f CENTER = 5.300 MHz.

3. Measurements of the spectrum of the output frequency

In order to get a better understanding of the harmonic spurs of the synthesizer, the number of bits of the DAC has been reduced. The most significant bit (MSB) of the NCO is actually equivalent to a 1 bit square wave synthesizer. A picture of the power spectrum of a 1 bit synthesizer is represented in fig. 6. At first sight this 1 bit spectrum with a center frequency of 5.312501 MHz looks very good. In reality, however, the picture of the spectrum is smoothed by the span of the analyzer and what seems to be a single spectral line consists actually of a set of hundreds of lines.

One of the most interesting aspects of a digital frequency synthesizer is the following: If one moves the output frequency slowly through a narrow band of a few kilohertz the appearing spectrum patterns change their form in a surprising way. From one frequency setting to the other the spectral patterns are changing completely. This is demonstrated in the following pictures.

The center frequency main line in fig. 7 is at 5.3125 MHz. The highest subharmonic (of the input frequency) is at 1.25 MHz. This line is called the principal spur. For a 1 bit synthesizer this line is only about 10 db below the main line. Note that there is no line that is close to the carrier.

If the frequency is moved to a value of 5.3001 MHz one gets a picture with a complex spectral pattern whose envelope has several peaks (fig. 8). The first order principal spur is again 10 db below the carrier. We find however a second order spur that is more close to the carrier at 32 db. A third order principal spur is at 48 db. If the span of the analyzer is opened (fig. 9), there appears in addition a forth principal spur at a 70 db level that is very close to the carrier. At this point one is really wondering, where these complex patterns are coming from. A complete mathematical treatment of digital frequency synthesizers seems to be difficult. Therefore we want to develop only a few basic rules, that allow us a qualitative understanding of these spectra.

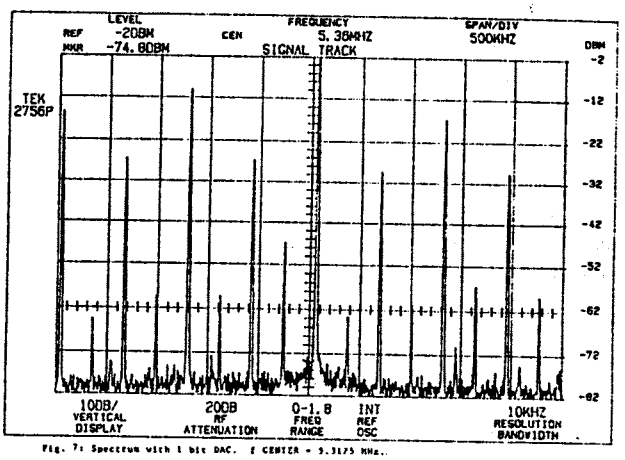


Fig. 7: Spectrum with 1 bit DAC. f CENTER = 5.3125 MHz.

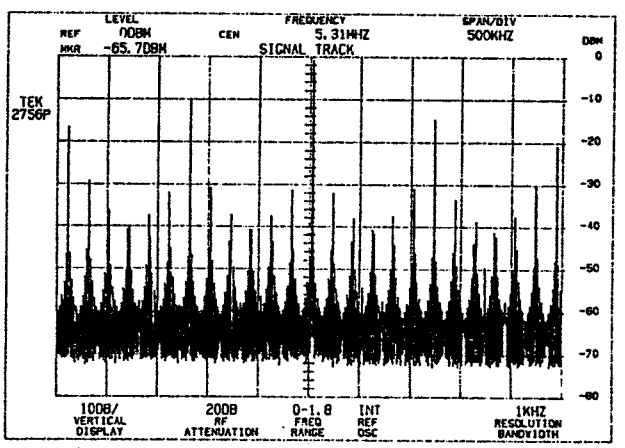


Fig. 8: Spectrum with 1 bit DAC. f CENTER = 5.3001 MHz, span = 500 kHz.

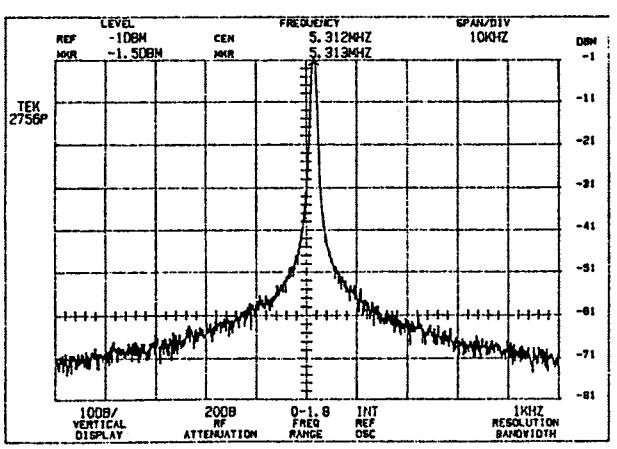


Fig. 6: Spectrum with 1 bit DAC. f CENTER = 5.3125C1 MHz.

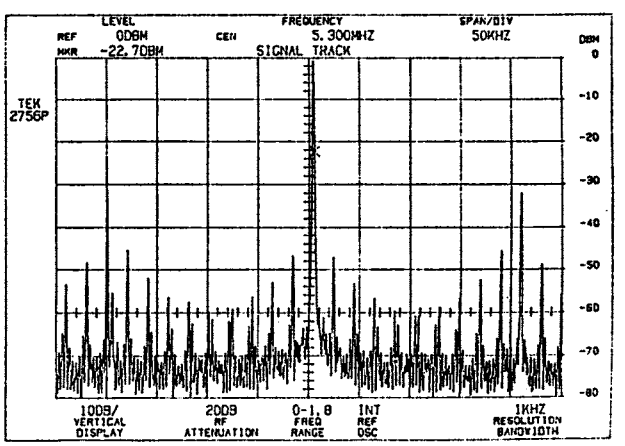


Fig. 9: Spectrum with 1 bit DAC. f CENTER = 5.3001 MHz, span = 50 kHz.

4. Mathematical model of a digital frequency synthesizer

A model of a 1 bit frequency synthesizer is represented in fig. 10. Similar models are found in the literature [1], [2], [3]. The output frequency is obtained by the addition of pulses with a rate that is determined by a number N of cascaded frequency dividers N_o, N_1, N_2...

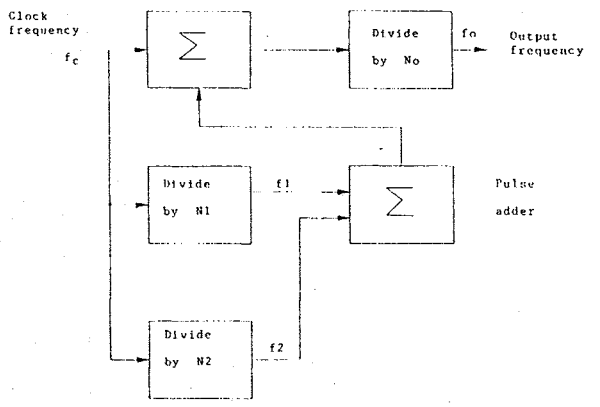


Fig. 10: Principle schematics of cascaded frequency division.

Now let us consider a simple "two-sound" synthesizer with two division factors N_o and N_1. It is easy to see that the averaged output frequency f_o is given by the equation

$$f_o = \frac{1}{N_o}\left(f_c + f_1\right) = \frac{1}{N_o} f_c + \frac{1}{N_o N_1} f_c ,$$

where f_c is the input clock frequency.

For an "N-sound" synthesizer with N division factors the output frequency is then

$$f_o = f + \Delta^1 f + \Delta^2 f \ldots$$
$$= \frac{f_c}{N_o} + \frac{f_c}{N_o N_1} + \frac{f_c}{N_o N_2} \ldots$$

If there is only 1 divider the output frequency is f_c/N_o. For 2 dividers one obtains a phase output as function of the time t, that is represented by a staircase function (fig. 11). The action of the second division factor N_1 is twofold. First it generates an average output frequency offset $\Delta^1 f$, equal to

$$\Delta^1 f = \frac{f_c}{N_o N_1}$$

and second it generates a phase modulation that is represented by a scanning wave or a sawtooth function.

The peak deviation of the scanning wave is half the unit period T_c of the clock frequency f_c.

The modulation frequency (f_{MOD}) of the scanning wave is

$$f_{MOD} = \frac{1}{N_1} f_c .$$

For a given set of monotonously increasing factors N_o, N_1, N_2..., there is always a well defined output frequency. Conversely, if one has a given output frequency, a series of division factors N_o, N_1, N_2 can be determined by a recursive fractional division algorithm. Different output/input frequency fractions can then be classified into 1, 2 or N-fold classes. For some fractions, the number N of the division factors is finite. It is clear that this number is equal to the number of principal spurs that have been measured (fig. 9).

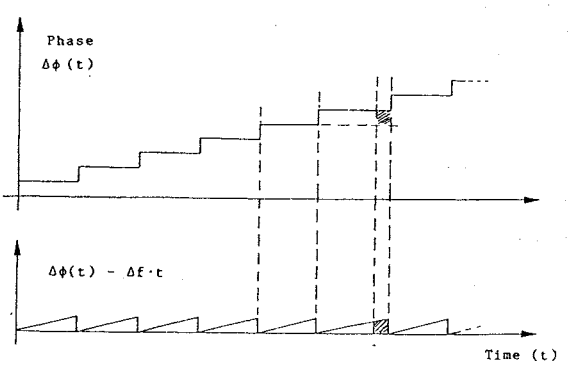


Fig. 11: Phase output $\Delta\phi$ with 2 dividers as a function of time. Δf is the average frequency drift.

The actual NCO is very similar to the mathematical model with the restriction that the division factors are limited to binary values 2^k. Amplitude steps (ΔA) that are due to the digital analog converter can be translated into phase steps ($\Delta\tau$) by the equation

$$\frac{\Delta A}{A} = 2\pi \cdot \frac{\Delta\tau}{\tau}.$$

This can be easily derived from fig. 4. With a 10 bit converter, one obtains phase steps that are $\frac{1}{2\pi}$ times 1024 smaller than the nominal period of the output frequency. An NCO with DAC can therefore be used as a digital phase stepper.

The scanning wave that we have considered so far has an elementary harmonic spectrum [4]. The fundamental modulation frequency has an amplitude of $\frac{2}{\pi}$. This translates into a modulation index m with $m = 2 \cdot \frac{\Delta\tau}{\tau}$.

The phase power spectral density S(f) is then equal to

$$S(f) = \frac{m^2}{2} = 2 \cdot \left(\frac{\Delta\tau}{\tau}\right)^2.$$

The one sided power spectral density L(f) with a low modulation index approximation [5] is equal to

$$L(f) = \frac{1}{2} S(f).$$

The overall behaviour of a digital frequency synthesizer can therefore be represented by the universal drawing in fig. 12. For an input/output frequency fraction $1/N_o$, where N_o is an integer number one concludes that there are no harmonic spurs. For every output frequency that is only a small factor (ε) away from a 1/N fraction, one gets immediately a set of subharmonic spurs with constant amplitude. These spurs can be very close to the carrier frequency and can therefore not be filtered by a phase-locked loop.

The frequency spacing (Δf) between two adjacent division factors N_o and N_o+1 is equal to $\Delta f = \frac{1}{N_o(N_o+1)}$. The amplitude of the first principal spur is proportional to N_o/N_1. Fig. 12 is self repeating, it is similar for a 10 bit converter with a frequency scale reducing factor of 2^{10}. The same figure applies also to case of the calculation of successive principal spurs. The frequency spacing is then $\Delta f = \frac{1}{N_1(N_1+1)}$ and the scale factor for the amplitude is proportional to $\frac{N_o}{N_2}$.

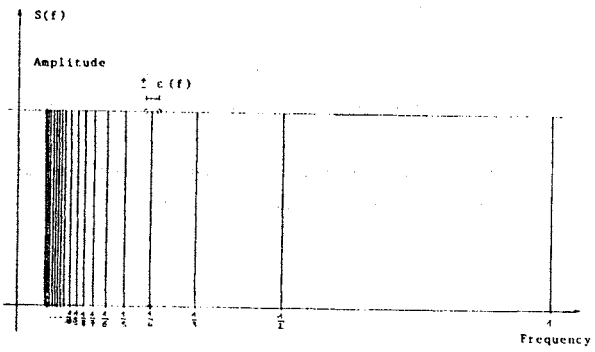


Fig. 12: Spectrum and frequency spacing of a digital frequency synthesizer.

Conclusion

As a conclusion we can state that we have designed a compact direct frequency synthesizer that is based on a numerically controlled oscillator. This synthesizer can be used in a Rubidium frequency standard.

A simple mathematical model has been presented that allows a qualitative understanding of the synthesizer spectrum and a quantitative estimate of successive principal subharmonic spurs.

Literature

[1] V. S. Reinhardt, Direct Digital Synthesizers, Proceedings of the 17th Annual Precise Time and Time Interval Applications and Planning Meeting (NASA/DOD), Washington, D. C., December 3-5, 1985, page 345-374.

[2] R. Schröder, Zur quasiperiodischen Frequenzteilung, Frequenz, Fachverlag Schiele & Schön GmbH, Vol. 30, March 1976, page 50-55.

[3] L. Euler, Einleitung in die Analysis des Unendlichen, Reprint, Springer-Verlag, 1983.

[4] C. E. Smith, Applied mathematics for Radio and Communication Engineers, Dover, New-York, 1961.

[5] J. D. Ryder, Electronic Fundamentals and Applications, Prentice Hall, page 443-468.

SESSION 4A

MATERIALS

Chairman : A. ZARKA

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4A Number 22

SYNTHETIC QUARTZ MATERIAL CHARACTERISATIONS. CONNEXIONS WITH THE QUALITY OF AT-CUT RESONATOR

J.J. BOY* , P. ZECCHINI** , B. VIARD**

*LCEP / ENSMM - Route de Gray - La Bouloie - 25030 Besançon Cedex - FRANCE
**LCCM - Faculté de Sciences - Route de Gray - 25030 Besançon Cedex - FRANCE

Summary

In order to examine the quality of different swept and unswept synthetic quartz material, both Y- and AT-cuts have been sawn in pure Z-zones. These various samples have been prepared according to IEC publication 758 recommendations.

Many characterizations have been carried out on these samples, including mechanical and physical defect analyses such as infrared and near-infrared absorption measurement, induction coupled plasma emission spectrometry spectrochemical analysis, X-ray topography, detection of inclusions by laser scattering and counts of etch pits and channels.

The electrical parameters of 10 MHz - 3rd overtone AT-cut resonators, fabricated from the same bars by using the closest region from the analysed part, have been checked to tentatively better understand the correlations existing between these measurements and the evaluation of the quartz material previously qualified. Two types of resonators have been used: the first one is a BVA_2 type plano-convex resonator, so comparisons are possible with well known specifications of BVA units manufactured in the C.E.P. laboratory. The other one corresponds to an ultraminiature (less than 10 mm^3) unit, making possible the characterization of blanks having a small volume.

The results of this evaluation of some different synthetic quartz crystals show that, if the commercial qualification of the material is the first step for an industrial choice of the material, the specificity of each bar or the sweeping can induce some non expected observations.

Introduction

Many works have been undertaken to better understand the behavior of quartz resonators. Various analyses were performed on cultured (or synthetic) and natural quartz to select the best quartz material for each particular application such as high stability resonator, accelerometer, temperature and/or pressure sensor, ultra miniaturized resonator manufactured by chemical etching, ...

Each application fields worker needs particular treatment or characterisation measurements. Furthermore, several techniques (such as sweeping or irradiation) are used to improve the quartz material quality and so the resonator behavior.

This paper deals only with a few specific characterizations on synthetic quartz and their consequences on certain behaviors of two kinds of AT-cut resonators.

The cultured quartz used in this investigation have been obtained from two different growers: the first one in the USA (called G1), the second one being S.I.C.N. (called G2) in France. These quartz bars have been bought (many years ago for the first one issued from G1!) to manufacture our "daily" BVA production. We order only pure Z-zones of lumbered quartz blocks and so we can not relate the complete growing method. In particular, we can not be sure where the dislocations take from: in the seed or the seed veil? Although, our first investigations take place in a important job of the quartz quality material determination in order to perform high stability resonator.

The block of G1 (which will be called, its too, G1) was bought in 1978 and its pure Z-zone had the following dimensions: 28 mm in the X and Z directions about 120 mm in the Y-direction

As for the G2 quartz block, more recently produced, we have asked that only half block must be swept. Before sweeping, the dimensions were:

16 mm along Z-axis and 34 mm along X-axis,
about 110 mm along Y-axis,

after sweeping, we had two bars, with identical dimensions and with squared section of 16*16 mm^2.

Preparation and measurements

The samples which are necessary to the examination of the material quality have been prepared according to IEC publication 758 recommendations. The Y- and AT-cut plates (used to the Infra-Red spectrometry measurements) are about 10 mm thick and are grinded and polished to avoid the disadvantages due to a bad surface profile. The Y-cut blanks are, after, used to analyse the amount of foreign atoms. These chemical impurities being as small as trace elements, we have chosen to use the Argon Plasma emission spectrometry called ICP (Induction Coupled Plasma emission spectrochemical analysis). For that, the quartz crystal is dissolved in a hydrofluoridric solution, the detail of the preparation being explained in (1).

The other AT-cut blanks are cut as plates of about 1 mm thick to perform X-ray topographies with Lang camera, etch pits and etch channels characterizations and finally the miniature or ultra-miniature resonators.

Macroscopic inspections

LASER SCATTERING OBSERVATION:

An inspection by laser scattering of the blocks dipped in benzilic alcohol (to avoid grinding and polishing operations) revealed that the amount of small inclusions (with diameter less than a few microns) is low: about 2 or 3 by cm^3. The two kinds of cultured quartz having been swept, numerous inclusions can damage the quartz around these. Indeed, we have seen any fracture. Finally, we can insure that the quartz blocks belong to Grade I i.e. inclusion density in one cm^3 does not exceed 3 in the size range : <10 - 30 µm>

This first observation, made rapidly, was followed by an other identical observation on the AT-cut blanks which will become resonators; it avoids to be confront with eventual problems on the resonators during their manufacture.

ETCH PITS AND ETCH CHANNELS:

The second macroscopic inspection concerns the count of etch pits and the observation of the narrow etch tunnels (2), (3), (4). The chemical etching of quartz blanks in fluoride-based etchants has been a common production practice for many years (5). Etching is performed for many reasons as different as:

- control quality of natural quartz blanks,
- post lapping adjustment technique (which avoid also the inevitable "dead-layer" due to mechanical lapping process),
- operation of mechanical polish lapped blanks,
- ...

Nevertheless, this chemical etching can create damages on the etch surface which result in the formation of etch pits and etch channels in crystal blanks.

AT-cut blanks are sawn with parallel faces and finely polished before etching. After polishing, the plates are of about 1 mm thick to permit the realisation of X-ray topographies and finally the manufacture of the resonators on the same sample.

As it is shown in (6), there is no significant difference in etch channel count when etching time exceeds one hour (i.e. material removal exceeds 60 microns) and so our etching time was limited to 1.5 hour in a saturated aqueous solution of ammonium bifluoride at 75 °C (according, on this last point, with the standard method specified in the April 1988 edition of IEC 758 publication).

Fig. 1 shows a lot of characteristic details on one side of an AT-cut sample. Almost of these pits were terminated by a large and narrow pipe (or channel) whose the angle of penetration, related to the angle of the crystal cut, is close to the Z-direction. Indeed, etch channels ALWAYS lie within a few degrees of the growth direction. This is true now that channel is produced by a dislocation coming from the seed, the seed veil or foreign inclusions (as "acmite" or "spaudumène" which contains Aluminium impurities due probably to Lithium doped growing method) (7) (8). See below the X-ray topographies of AT-cut samples.

We have found that these kinds of pipes are only seen on NON-SWEPT material. Having a few hundred of their in a squarred centimeter (so, the N-S G2 quartz belongs to the GRADE 5), they are just less than 10/cm^2 in SWEPT material. We can said it belongs to the grade1, as, moreover, the half-part quartz crystal from G2.

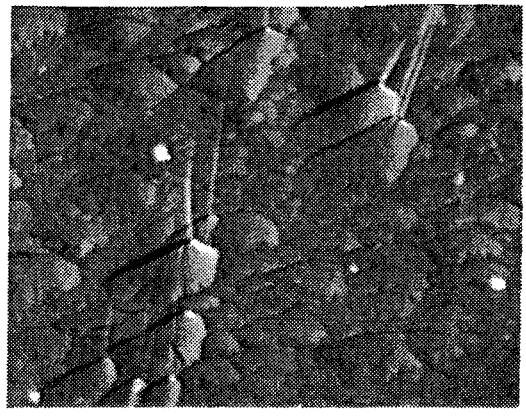


Fig. 1: etch pits and channels on NON-SWEPT G2 sample (75 X)

The following picture (takes in with a CCD camera and stores with "Digital Dark" soft in Macintosh Desk Computer) presents a particular shape, roughly hexagonal and identical to this of dislocations topographies. Furthermore, this link between etch pits and dislocations was already confirmed by Lang and Miuscov in 1967; they specially concluded that, with rare exceptions, pits formed at dislocations emergences and that no pits occurred not at dislocations. They explained also that dislocations organize into polygonal cell boundary network (9), (10).



Fig. 2: Etch pits on G1 sample

X-RAY TOPOGRAPHIC PATTERNS

The next pictures presented here are X-ray topographic patterns obtained with our own RIGAKU 100PL generator. All samples are AT-cut and, of course, the reticular plane is 2-1.0 which is about perpendicular to dislocation lines. The last picture shows (if it was necessary to confirm that!...) the existing of the same number of dislocations in SWEPT and NON SWEPT quartz material. So, we insure again that absence of etch channels does not mean absence of dislocations and etch channel counting cannot replace X-ray topographies with Lang camera. However, the presence of etch pits having been revealed with ammonium bifluoride explains the existing of "bundles" of dislocations.

Fig. 3: Dislocations in Swept. AT-cut (G2)

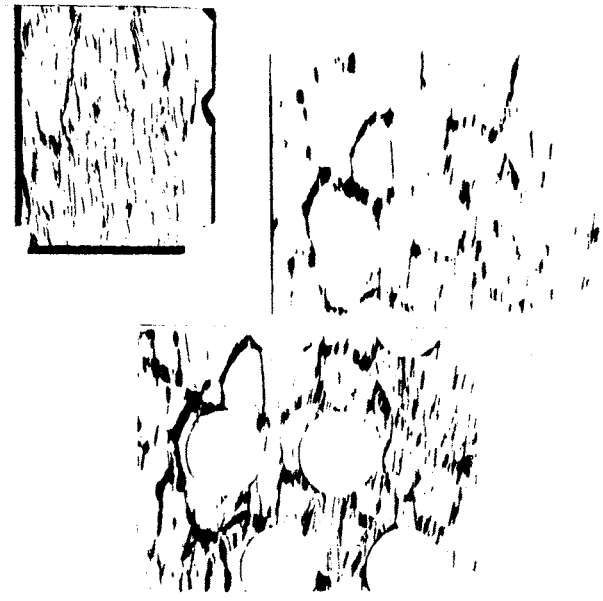


Fig. 4a and 4b: AT-cut from G1 - Same shape as etch pits picture. Before and after ultrasoning cutting process of ultra-miniature resonators.

Electrical Twin Observation:

When we manufacture resonators in cultured quartz, we do not look for optical or electrical twinning. The problem of their existings is not take in account. However, certain mechanical process can create damages in a crystal structure as rectification during the pre-cut operation. Sometimes, a few micro-twins can appear, but rarely electrical as we describe in this paragraph. The shape of one of this electrical twin is drawn is the next picture (Fig. 5). The surface shapes are not identical, nevertheless, this twin is in volume.

The shape of the two kinds of etch pits and the removed thickness by chemical etching describes really an electrical twin (See fig. 6a, 6b and Table 1).

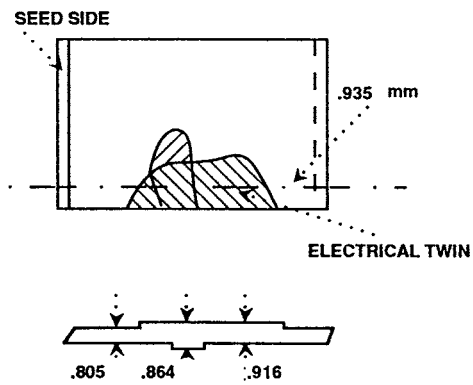


Fig. 5: Shape of one electrical twin on G1 sample (AT-cut)

CUT ANGLES	ETCHING VELOCITY	REMOVED THICKNESS
AT-cut	68 Å/s	65 µm/side
"AT-twin"	16 Å/s	14 µm/side

Table 1

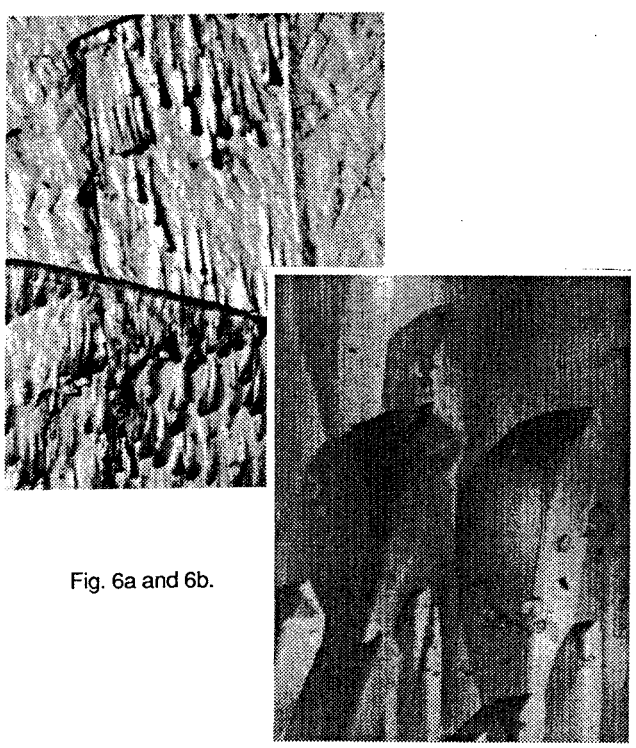


Fig. 6a and 6b.

Microscopic observations

INFRARED AND NEAR-INFRARED SPECTROMETRY:

The established commercial method for unswept cultured quartz quality evaluation and α_{3500} determination uses scans at specific room temperature OH absorption regions. While grown-in OH distributions can be specified from this method, low temperature measurements at 85 K provide improved sensitivity for determining OH and Al-OH absorption before and after sweeping or irradiation.

The samples used are rectangular sections with parallel x-, y- and z-faces. Their thickness is approximately 0.5 cm in the Y-direction after lapping and optical polishing.

- Room temperature infrared measurements:

Different regions of each samples have been measured. The α_{3500} calculated values are listed in table 2. Even though these calculations are not significant for swept quartz, their relative values can be usefully exploited to make comparison between swept crystals.

	G1(seed)	G1(ext)	unswept G2	swept G2
α_{3500}	0.0442	0.0332	0.0215	0.0199
grade	B	B	A	Aa
Q value*	$1.97.10^6$	$2.12.10^6$	$2.36.10^6$	$2.40.10^6$
H content	210 ppm	160 ppm	100 ppm	95 ppm

Table 2

In table 2, the relative Q values are listed too. They have been calculated by the equation established by the C.C.M. laboratory by extrapolations of the SICN's experimental curve. Furthermore, according to Brice and Cole (11), the hydrogen content could be determine using data from Lias and al. who give an α_{3500} value equal to 1 cm^{-1} for a 4800 ppm H relative to silicon content.

Infrared measurements performed successively at two constant wavelengths, 3800 or 3500 $cm.^{-1}$, and along the sample growth sweeping axis, the Z-axis, reveal the local variations of the α_{3500} value. For these measurements, the sample is moved perpendicularly to a 2x10 mm infrared beam. In figure 7, 32 data points calculated along the 28 mm G1 sample width are reported. A large inhomogeneity is observed. These measurements are nevertheless in good agreement with the previous results of the α_{3500} values given in table 2 and calculated in two half parts of the sample. In fact these two values are approximately the average of the data obtained in each region as shown in figure 7.

Measurements as described previously and performed along 16 mm width Y-cut plates of the both G2 samples indicate a perfect homogeneity of the pure Z-zone.

- Low temperature infrared measurements:

By measuring the infrared spectra at 85 K, the specific absorptions of the as-grown OH in unswept and swept crystals, and of the Al-OH in swept crystals can be increased.

Figures 8a and 8b show the effect of the sweeping in each part of the sample G1. A large difference is noted, particularly on the intensity of the "e_2" absorption at 3366 cm^{-1}

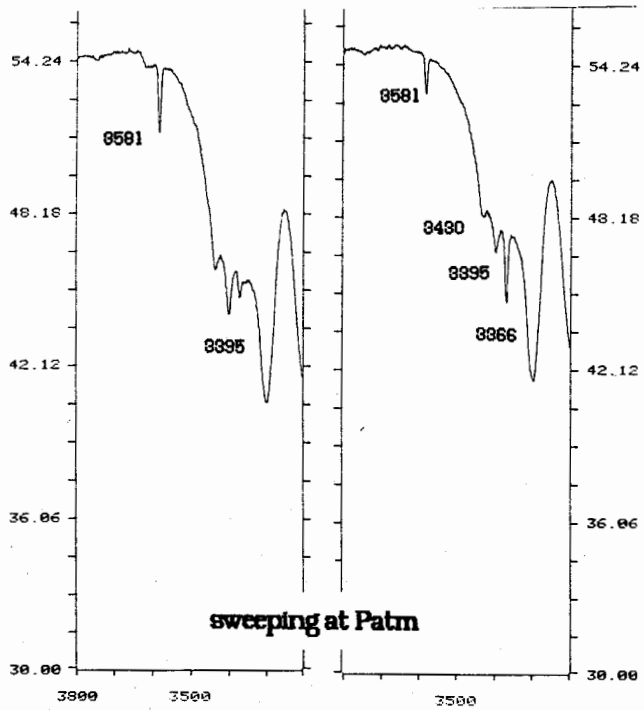


Fig. 8 a : G1, seed side

Fig. 8 b : G1, ext. side

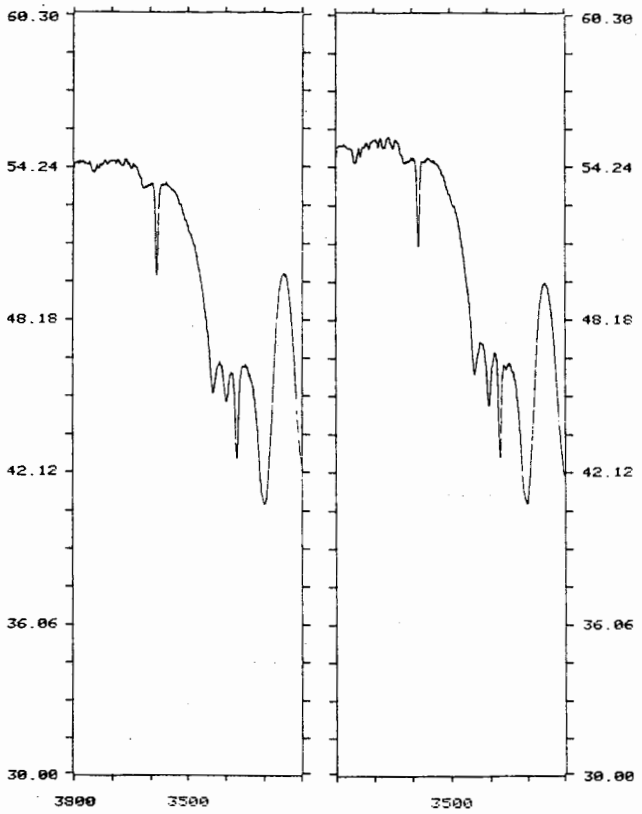


Fig. 8 c: G1,both sides
after irradiation

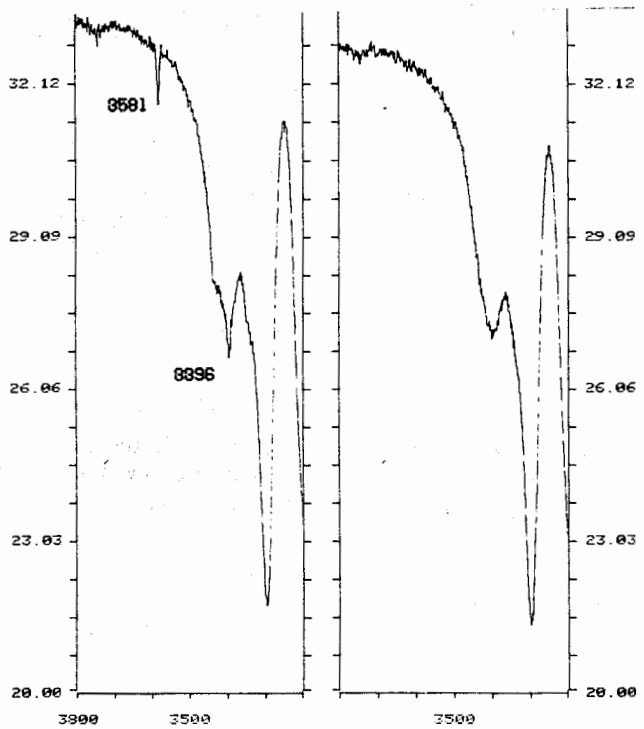


Fig. 8d : G2, unswept

Fig. 8e : G2, swept

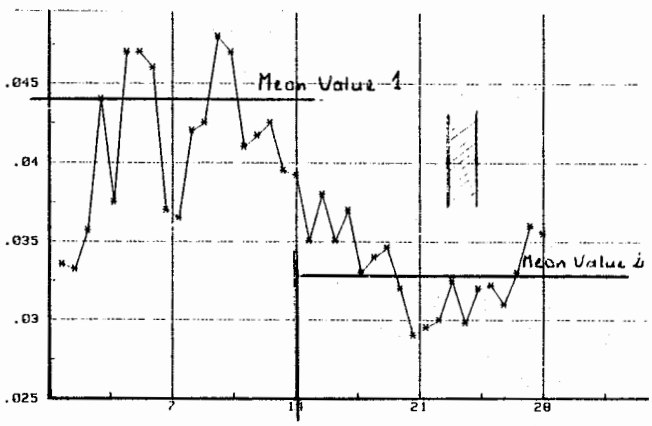


Fig. 7 - α_{3500} measured along the Z-axis

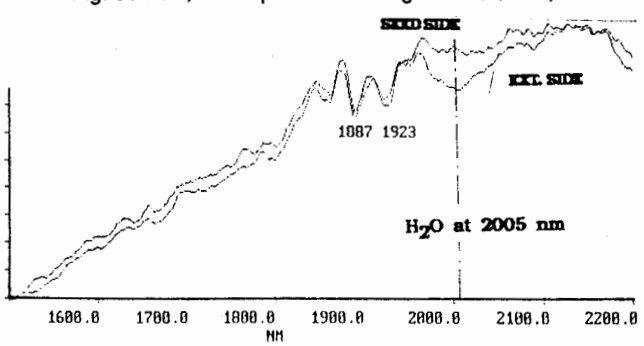


Fig 9a: near-infrared spectra on the two parts of G1

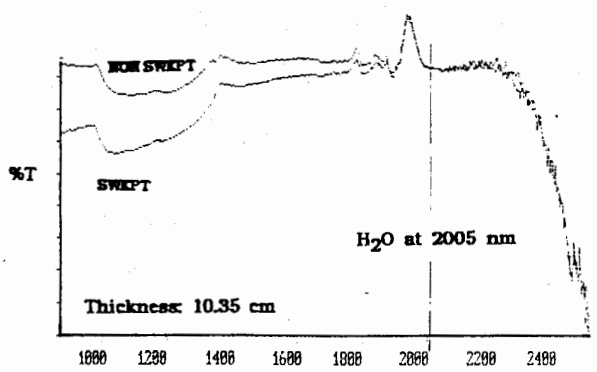


Fig 9b: near-infrared spectra on G2 (unswept and swept)

which corresponds to the Al-OH absorption. After a 1 Mrad ^{60}Co gamma irradiation dose, a new scan of this sample, figure 8c, shows that neither the Al content nor the OH content justifies the former difference. Only an incomplete air-sweeping, performed from the external side to the seed side, can explain these observations.

Figures 8d and 8e show the spectra obtained by scanning the unswept and the swept G2 samples. Very low OH-absorption at 3581 and 3396 cm^{-1} are observed in the unswept sample. These absorptions are no longer present after the vacuum-swept process, but the "e_2" Al-OH absorption is not detected too. This observation can have at least one of those two origins: a very low aluminum content or no OH-source within the crystal.

- Room temperature near-infrared measurements:

The measurement of the water content is performed at room temperature in the near-infrared at 2005 nanometers, using 10 cm thick samples. Observations are made with the incident beam align with the Y-axis. Figures 9a and 9b indicate the relative H_2O content of the samples G1(seed side and external side) and G2 (unswept and swept).The largest content is noted in the external side of the G1 sample. In both G2 samples, no absorption is detected at 2005 nanometers. As shown in figure 9b, the increasing absorption from 1900 to 1000 nanometers, and in the visible part of the spectra too, is stronger for the swept sample than for the unswept one. These both observations indicate that the G2 material can be considered as a "dry" crystal without OH-source, the sweeping induced blackness could be due to aluminum-hole formation.

ICP CHEMICAL ANALYSIS:

The low amount of impurities in the G1 and G2 synthetic quartz crystals requires to be quantitatively measured to use ICP chemical analysis. The analysed parts of the crystals are the IR Y-cut plates and both the adjacent plates from the AT-cuts which will be used to fabricate resonators. By avoiding all the causes of contamination as discussed in (1), the results (expressed in ppm) obtained by performing analyses on a few grams of quartz are given in table 3.

El.	G1(Seed)	G1(Ext.)	G2(N.S.)	G2(S.)
Al	15.95		5.79	1.68
Fe	0.83		0.81	1.41
Ca	3.47		2.03	2.52
Na	1.0	0.7	1.0	1.0
K	0.2	0.1	0.3	0.3
Li	0.02	0.02	0.02	0

Table 3

The amount of aluminum is very high in sample G1 and as a large OH-source is available, the sample remained clear after sweeping. For the crystal G2, the difference in aluminum content between the swept and the unswept sample is not yet explained. As the NIR scans might lead predictable, aluminum contained in the G2 sample can give a blackness to the sample by forming Al-hole under sweeping.

In all the analysed samples, a very low alcali content is measured. The expected consequences of the sweeping on a so low alcali amount are not very convincing, only a small reduction of the sodium content is noted in the external swept part of sample G1.

ALPHA MEASUREMENTS ALONG Z'-AXIS:

In figure 10, the α_{3500} value distribution is shown for an AT-cut plate. This last representation is used in order to correlate the calculated α_{3500} values with the position of the ultra-miniature resonators (figures 11a and 11b). The results are equivalent to the previous ones, the only difference being due to the fact that the IR beam reaches the surface of the sample with an angle of about 35 degrees (in fact equal to theta) around Y-axis.

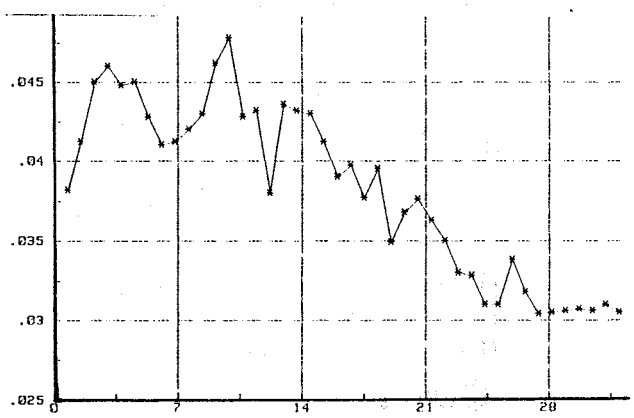


Fig. 10 - α_{3500} measured along the Z'-axis

TESTS ON AT-CUT RESONATORS:

We have achieved two kinds of resonators. Though we ordinary manufacture SC-cut BVA resonators in CEP laboratory, we have chosen to manufacture AT-cut for two reasons:

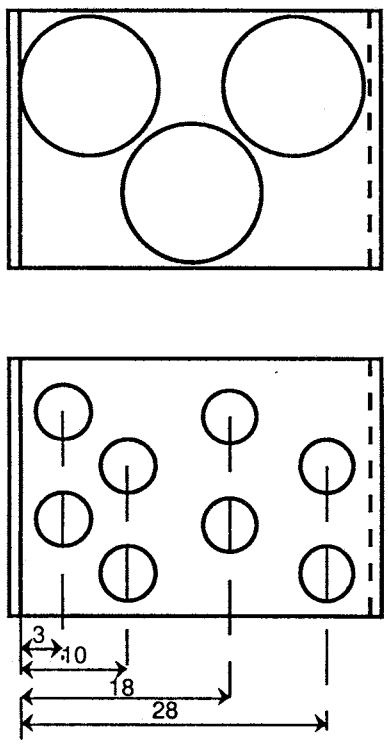
the first one is because the technical comittees in "Piezoelectric devices for frequency control and selection" recommend, as specified in IEC Publication 122-1, to evaluate the material and in fine the frequency versus temperature characteristics on AT-cut plate (with Theta equal to 35°13' +/- 0.5', the sign of Theta depending on the used convention). These evaluations permit, in particular, comparison between the quality factor of one given resonance frequency and the intrinsic Q-factor, deduced from the IR spectrometry analysis.

the second one is due to a LCEP team who has developed and realized AT-cut ultraminiature units (less

than 10 mm^3) permitting the characterisation of a very small volume of quartz blanks. The specifications of these resonators are presented in the thesis of C. Longet submitted to a referee in January 1991. We can add that their designs have been sufficiently optimised for that the electrical performances are repeatable and excellent: the mean Q-value of the 10 MHz 3rd-Ovr is better than .5 million.

Though a part of G2 quartz block has been unusable due to the electrical twin, we have tried to perform, at least, more than two identical resonators.

Fig.11a and 11b give the positions of each blank resonator which have been determined according to the ALPHA-values measured point by point along Z'-axis, and this, only for G2 sample.



Even though the ultra-miniature resonators work on the 3rd-Overtone at about 10 MHz, the BVA AT-cut (with a vibrating part diameter of 10.6 mm) has a design optimized to work on the 5th-Overtone.

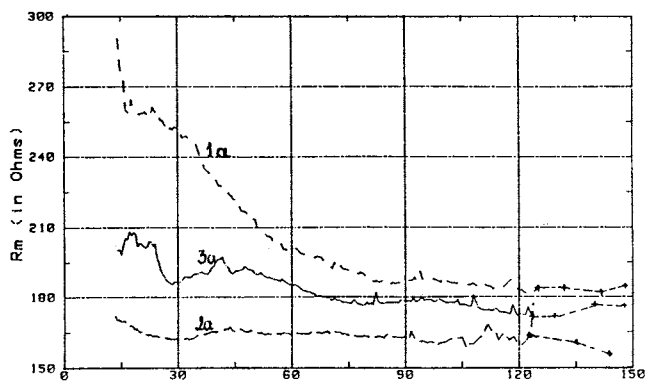
The following table presents the electrical parameters of each resonator whichfor the golden electrodes have been deposited by evaporation. The resonators have been mounted in a "flag" and the measurements performed under vaccum (about 6.5 mPa) at ambient temperature.

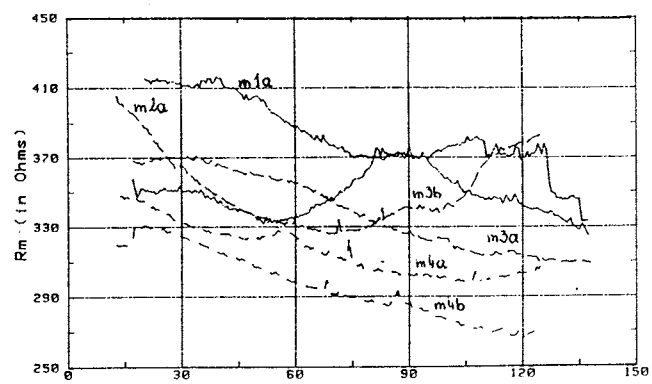
Last measurements performed on these resonators concern the temperature behavior of the motional resistance which was measured at atmospheric pressure. So, the values of these motional resistances are higher than under vaccum.

With a desk computer, we check the phase difference between the input and output levels on a "PI" bridge within which the quartz resonator is inserted.

Res. number	Rm (in Ω)	Q (* 10^6)
1a	96	.95
1b	Electrical twin (?)	
2a	68	1.22
2b	Electrical twin (!)	
3a	72	1.13
3b	71	1.12
m1a	280	.26
m1b	300	.25
m2a	250	.29
m2b	240	.31
m3a	240	.31
m3b	250	.30
m4a	210	.35
m4b	193	.38
Sa	84	1.02
Sb	78	1.10
NSa	65	1.35
NSb	74	1.18

Generally, the temperature behavior is in accordance with the last measurements. We can note again that the quality of the resonators made in Non-Swept material of G2 are slightly better than the others made in Swept material. At least, it must be pointed out that all these resonators have not been annealed before measurements (just kept at 140 °C during about two hours).





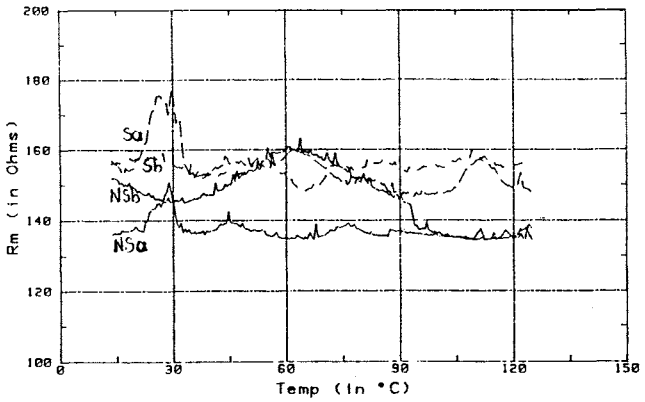


Fig. 12a, 12b and 12c

Conclusions

In this paper, the first results obtained in order to correlate the quality of the material and the resonator performances have been presented. This study of unswept and swept quartz crystals by using many different physical observation techniques and by manufacturing two types of AT-cut resonators closer as possible from the analysed material, shows all the difficulties which can occur, consequences of the inhomogeneity of the crystal, or of the sweeping or in the resonator manufacturing process.

It seems that the measurements on AT-cut resonators is more appropriate if the mechanical process is well known, i.e. if the results are repeatable. This parameter can become a problem in case of miniaturization. Whereas, the researchs of the LCEP about ultraminiature resonators permit to qualify a very small volume of quartz material.

Some of these results confirm the physical observations,and as it could be predicted, the better quartz material the better resonator performances. But, the measurements of the electrical parameters of the motional resistance of the "G2" resonators have to be performed again after annealing and aging, to appreciate better the real effect of the sweeping of crystals whichfor the amount of the as-grown alcali is already very low.

More complete chemical analyses have to be done using larger amounts of quartz crystal so that monitoring the change in the low impurity content before and after the sweeping process could be possible.

Acknowledgments

The authors would like to thank R.Bourquin and C. Longet from the CEP laboratory who have defined the dimensions and the contouring of the ultra-miniature AT-cut resonators, all the technical team of the CEP laboratory for its competent help in the fabrication of the resonators and B. Astric from the CCM laboratory for performing etch channels patterns in etched plates.

References

1. B. VIARD, J. DAUMET, M. KHAYAR, H. MERIGOUX: "Trace element analysis in quartz", Proc. 3rd European Frequency and Time Forum (1989), pp. 609 - 616

2. J. W. NIELSON, F. G. FOSTER: "Unusual etch pits in quartz crystals", The American Mineralogist, Vol. 45 (March / April 1960), pp.299 - 310

3. J. G. GUALTIERI: "The influence of temp. and electric field on the etch channel density in swept cultured quartz", Proc. 39th Annual Freq. Control Symp. (1985), pp. 247 - 254

4. R. A. LAUDISE and R.L. BARNS: "Perfection of quartz and its connection to crystal growth", IEEE Trans. on Ultra., Ferro. and Freq. Control, Vol. 35, N° 3 (May 1988), pp. 277 - 287

5. J. R. HUNT: "Etch pits and etch channels in swept AT- and SC-cut quartz", 41st Annual Freq. Control Symp. (1987), pp. 183 - 191

6. O. BIGNON, B. BROTTIER: "Etch channel in quartz monocrystals - Factors of influence", Proc. 4th European Frequency and Time Forum (1990), pp. 609 - 616

7. A. ZARKA, L. LIU, M. BUISSON: "Influence de la localisation sectorielle du germe sur la qualité cristalline de quartz de synthèse", J. Crystal Growth (1981), vol. 54, pp. 394 - 398

8. J. F. BALASCIO, A. F. ARMINGTON: "Development results for the production of high qulity quartz", Proc. 40th Annual Freq. Control Symp. (1986), pp. 70 - 75

9. A.R. LANG, V.F. MIUSCOV, Journal Applied Physics, 38 (1967), pp. 2477

10. G.R. JOHNSON, R.A. IRVINE, J.W. FOISE: "Origins of dislocations in single crystal cultured quartz", Proc. 3rd European Frequency and Time Forum (1989), pp. 337 - 345

11. J.C. BRICE, A.M. COLE:"The characterisation of synthetic quartz by using infra-red absorption", Proc. 32d Annual Freq. Control Symp. (1978),pp. 1, 10

12. H.G. LIPSON, F. EULER, A.F. ARMINGTON: "low temperature infrared absorption of impurities in high grade quartz", Proc. 32d Annual Freq. Control Symp. (1978), pp. 11, 23

Session 4A Number 23

Sebastian -

RS

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4A Number 24

THE CONTRIBUTION OF THE TRANSIT-TIME AND THE RESONATOR METHOD TOWARDS DETERMINATION OF THE NONLINEARITIES IN QUARTZ

Carl K. Hruska

Piezoelectricity Research Laboratory
Department of Mathematics and Statistics, York University
North York, Ontario M3J 1P3, Canada

ABSTRACT

The experimental data provided by the transit-time and the resonator method are related to 29 of the total number of 31 third-order nonlinear electromechanical constants of quartz. However, the analysis of the data can provide at most only the values of 5 constants and of 12 linear combinations of 24 other constants. Some of the results obtained independently by the two methods appear to be significantly different.

Insofar as the two methods are being used as a primary source of the electroelastic, electrostrictive and third-order dielectric constants, they can provide no useful information about the third-order elastic constants. Moreover, the latter constants must be known for the three former types of constants to be fully calculable. Some of their values are thus dependent on the third-order elastic constants. This dependence complicates the proper assessment of the standard errors of the calculated constants.

The degree of agreement in the values of the nonlinear constants calculated by the two methods is independent of the third-order elastic constants. As such it is a measure of the consistency of the two methods. Consequently, efforts to reconcile the existing contradictions should be directed towards a re-examination of the two methods.

The interpretation of the agreement in some of the values of the nonlinear constants as a confirmation of the absolute correctness of their values is conditional upon the reliability of the third-order elastic constants used for their calculation.

INTRODUCTION

The determination of the third-order nonlinear electromechanical constants of quartz is a gradual process. Additional constants are calculated as new experimental methods become available. Earlier results also participate in the process which makes the results obtained by different methods interrelated. While the nature of these relationships remains unclear it is difficult to pinpoint the source of discrepancies between comparable results. The amount of information each method can offer may also be unnecessarily reduced.

In order to clarify the above issues it is useful to state for each experimental method individually and in isolation from other methods what its real contribution towards the determination of the nonlinear constants in quartz is. This is done in this paper for two experimental methods based on interactions of quartz with a dc electric field: the transit-time method described by Reider, Kittinger and Tichy [1] and the resonator method by Hruska [2]. These two methods currently provide values of the electroelastic, the electrostrictive and the third-order dielectric constants. As their calculation also involves the use of the third-order elastic constants provided earlier by Thurston, McSkimin and Andreatch [3] their dependence on these constants is shown.

In dealing with the matter the work relies on the theoretical models of the transit-time and the resonator method taken from Kittinger, Tichy and Friedel [4] and from Kittinger and Tichy [5] and Brendel [6] respectively, with some modifications of the last made by Hruska and Brendel [7]. The models [4,5,6] are in turn said to be based on the work by Baumhauer and Tiersten [8], Nelson [9] and Tiersten and Ballato [10].

The second-order (linear) material constants of quartz needed in this work are the elastic constants taken from [11] and the piezoelectric and dielectric constants taken from [12].

All nonlinear material constants appearing in this paper are related to the state of zero strain and zero electric field. Their numerical values are stated for right-hand quartz and the frame of reference according to the IEEE Standard 176 of 1978 [13] and for room temperature.

THE INDIVIDUAL CONTRIBUTION OF AN EXPERIMENTAL METHOD

The total number of 31 independent constants describing the third-order electromechanical nonlinearities in quartz consists of fourteen third-order elastic constants

(1) c_{111}, c_{112}, c_{113}, c_{114}, c_{123}, c_{124}, c_{133}, c_{134}, c_{144}, c_{155}, c_{222}, c_{333}, c_{344}, c_{444},

eight electroelastic constants

(2) f_{111}, f_{113}, f_{114}, f_{122}, f_{124}, f_{134}, f_{144}, f_{315},

eight electrostrictive constants

(3) l_{11}, l_{12}, l_{13}, l_{14}, l_{31}, l_{33}, l_{41}, l_{44}

and of one third-order permittivity constant

(4) κ_{111}.

The nonlinear constants (1)-(4) are recorded using the traditional choice of matrix indices. The definition of these constants based on their position in the constitutive equations and on other considerations can be found in [14].

The analysis of the experimental methods used to determine the nonlinear constants typically leads to a system of m linear equations of the following type

$$(5)\quad L_i = M_{i1}x_1 + M_{i2}x_2 + \ldots + M_{ij}x_j + \ldots M_{in}x_n + A_{i0},$$

where $i = 1,2,\ldots,m$. L_i represent the measured quantities provided by the experimental method in question and x_j, $J = 1,2,\ldots,n$, commonly denote all of the third-order constants (1)-(4) which appear in Eq.(5). The coefficients M_{ij} and the absolute terms A_{i0} are known numbers. They are functions of the 2nd order (linear) constants, of the crystal specimen (crystallographic orientation and shape) and of

the conditions under which it is investigated. Their form is determined by the theory interpreting each particular experiment.

As the experimental quantities L_i are necessarily biased by experimental errors and the number of equations m usually exceeds the number of the nonlinear constants x_j to be calculated, system (5) is overdetermined and the nonlinear constants can be sought only in terms of their best fitting values. This is usually done using the least-squares fit.

Providing that the assumptions necessary for the applicability of the least-squares method are satisfied the least-squares method is an algebraic procedure which provides likely values of the unknown nonlinear constants x_j and translates the random errors in the experimental quantities L_i into the estimates of the standard errors of the determined nonlinear constants x_j .

Traditionally, when a linear system (5) pertaining to a new experimental method and data set becomes available it is used to calculate only those material constants unknown at the time of the solution. All other material constants appearing in (5) which have been determined before are substituted for using their known values.

Such substitutions may lead to three consequences:

(1) they may make the results of the new experiment dependent on the results of the previous work without the form of the dependence being known;

(2) they may preclude the possibility of the new experiment being used towards an independent calculation (verification) of the substituted quantities;

(3) in case that two methods produce significantly different results the substitutions may complicate the process of tracing the source of the disagreement.

For these reasons it is preferable to analyze each experiment by solving system (5) on the assumption that none of the nonlinear constants which appear in it is known i.e. making no substitutions. A set of values of the nonlinear constants thus obtained, which is dependent only on the particular method used, represents the individual contribution of the method towards the determination of the nonlinear constants (1)-(4).

The participation of any individual nonlinear constant (1)-(4) in the solution obtained in the above manner is determined solely by whether or not it is originally present in system (5). The form of its participation in the solution - as a pure constant or in a combination with other constants - depends exclusively on the properties of matrix M of the system.

In some cases the experimental data set obtained by a given experimental method is not fully representative of the variety of experimental data the method can offer. As a result, it may happen that the solution calculable using the data set does not correspond to the full potential of the experimental method.

The described solution represents the maximum amount of information about the third-order nonlinear constants a given method or a data set can offer independently of other data sets or methods. As such it is most suitable for comparisons with similar solutions provided by other methods or data sets.

As some of the values belonging to the solution may be accompanied with excessively large standard error estimates the above described solution must be dissociated from the set of values which will ultimately be adopted for actual later use.

THE INDIVIDUAL CONTRIBUTION OF THE TRANSIT-TIME METHOD

The transit-time method consists in observations of the travel time through bulk quartz of one quasi-longitudinal and two quasi-transverse acoustic

$$k^T_{111} = f_{111} + 2.310 .10^{-12} c_{111} - 2.310 .10^{-12} c_{112} - 0.727 .10^{-12} c_{114}$$

$$k^T_{113} = f_{113} + 2.310 .10^{-12} c_{113} - 2.310 .10^{-12} c_{123} - 0.727 .10^{-12} c_{134}$$

$$k^T_{114} = f_{114} + 2.310 .10^{-12} c_{114} - 2.310 .10^{-12} c_{124} - 0.727 .10^{-12} c_{144}$$

$$k^T_{122} = f_{122} + 2.310 .10^{-12} c_{111} + 2.310 .10^{-12} c_{112} + 0.727 .10^{-12} c_{114} + 1.454 .10^{-12} c_{124} - 4.620 .10^{-12} c_{222}$$

$$k^T_{124} = f_{124} + 2.310 .10^{-12} c_{114} + 6.930 .10^{-12} c_{124} - 0.727 .10^{-12} c_{155}$$

$$k^T_{134} = f_{134} + 4.620 .10^{-12} c_{134} - 0.727 .10^{-12} c_{344}$$

$$k^T_{144} = f_{144} + 2.310 .10^{-12} c_{144} - 2.310 .10^{-12} c_{155} - 0.727 .10^{-12} c_{444}$$

$$k^T_{315} = f_{315}$$

$$k^T_{11} = l_{11} + 1.205 .10^{-12} c_{111} - 0.603 .10^{-12} c_{112} - 0.379 .10^{-12} c_{114} + 0.379 .10^{-12} c_{124} + 0.060 .10^{-12} c_{144} - 0.603 .10^{-12} c_{222}$$

$$k^T_{12} = l_{12} - 1.205 .10^{-12} c_{111} - 0.603 .10^{-12} c_{112} - 0.379 .10^{-12} c_{114} - 1.138 .10^{-12} c_{124} + 0.060 .10^{-12} c_{155} + 1.808 .10^{-12} c_{222}$$

$$k^T_{13} = l_{13} + 1.205 .10^{-12} c_{113} - 1.205 .10^{-12} c_{123} - 0.759 .10^{-12} c_{134} + 0.060 .10^{-12} c_{344}$$

$$k^T_{14} = l_{14} - 2.411 .10^{-12} c_{124} - 0.379 .10^{-12} c_{144} + 0.379 .10^{-12} c_{155} + 0.060 .10^{-12} c_{444}$$

$$k^T_{31} = l_{31}$$

$$k^T_{33} = l_{33}$$

$$k^T_{41} = l_{41}$$

$$k^T_{44} = l_{44}$$

$$q^T_{111} = \kappa_{111} - 4.929 .10^{-35} c_{111} - 4.651 .10^{-35} c_{124} - 0.349 .10^{-35} c_{144} + 0.338 .10^{-35} c_{155} + 4.932 .10^{-35} c_{222} + 0.036 .10^{-35} c_{444}$$

Table I. Nonlinear parameters calculable by the transit-time method. The parameters are combinations of the third-order nonlinear material constants of quartz including electroelastic constants f_{ijk}, third-order elastic constants c_{ijk}, electrostrictive constants l_{ij} and third-order permittivity κ_{111}. The first eight parameters (k^T_{ijk}) are in N/(V.m), the next eight (k^T_{ij}) are dimensionless, the last parameter (q^T_{111}) is in F/V. The numerical coefficients at c_{ijk} in parameters k^T_{ijk}, k^T_{ij} and q^T_{111} are in m/V, m^2/N and Fm^2/NV, respectively. The number and choice of indices of k^T_{ijk}, k^T_{ij} and q^T_{111} correspond to the first material constant in their definition.

pulses which can generally propagate in any direction in the quartz continuum. The changes in the travel time are measured under the effect of a dc biasing field oriented in various directions relative to the wave normal of the pulses.

The existing experimental set of 24 transit-time observations originates from Kittinger, Tichy and Friedel [4].

The nonlinear parameters representing the individual contribution of the transit-time method have been shown by Hruska [15]. They consist of 17 quantities 5 of which are pure nonliner constants while the rest are their linear combinations as shown in Table I. Their numerical values and their estimated standard errors are in Table II.

k^T_{111}	2.33 ± 0.06	k^T_{11}	4.92 ± 6.04
k^T_{113}	0.17 ± 0.06	k^T_{12}	-8.56 ± 6.14
k^T_{114}	0.68 ± 0.04	k^T_{13}	14.60 ± 4.28
k^T_{122}	-0.73 ± 0.05	k^T_{14}	-5.93 ± 2.54
k^T_{124}	1.39 ± 0.02	k^T_{31}	2.32 ± 2.05
k^T_{134}	1.71 ± 0.03	k^T_{33}	-3.34 ± 4.39
k^T_{144}	0.04 ± 0.02	k^T_{41}	-1.44 ± 2.21
k^T_{315}	-0.89 ± 0.02	k^T_{44}	-0.74 ± 1.79
	$q^T_{111}.10^{20}$	3.71 ± 2.39	

Table II. Nonlinear parameters obtained by the transit-time method. Parameters k^T_{ijk} are in N/(V.m), k^T_{ij} are dimensionless, q^T_{111} is in F/V. The errors are standard errors. Given for room temperature, right-hand quartz and the frame of reference according to IEEE Standard of 1978[13].

THE INDIVIDUAL CONTRIBUTION OF THE RESONATOR METHOD

The resonator method consists in observations of the series resonance frequency of quartz resonators which changes under the effect of a dc field bias.

The experimental data currently used for the determination of the nonlinear constants (1)-(4) consist of 184 observations obtained by Hruska and Khogali [16], Hruska, Kucera and Merigoux [17], Hruska and Brendel [18] and by Kinigadner [19]. They pertain to the extentional mode of rods and to the three thickness modes of plates. The dc biasing field used was parallel to the resonator thickness.

The nonlinear parameters representing the individual contribution of the above data set have been shown by Hruska [20]. They consist of 15 quantities 3 of which are pure nonliner constants; the rest are linear combinations of 26 other nonlinear constants as shown in Table III. Their numerical values and estimated standard errors are in Table IV.

$$k^R_{111} = f_{111} + 2.310\,.10^{-12}\,c_{111} - 2.310\,.10^{-12}\,c_{112} - 0.727\,.10^{-12}\,c_{114}$$

$$k^R_{113} = f_{113} + 2.310\,.10^{-12}\,c_{113} - 2.310\,.10^{-12}\,c_{123} - 0.727\,.10^{-12}\,c_{134}$$

$$k^R_{114} = f_{114} + 2.310\,.10^{-12}\,c_{114} - 2.310\,.10^{-12}\,c_{124} - 0.727\,.10^{-12}\,c_{144}$$

$$k^R_{122} = f_{122} + 2.310\,.10^{-12}\,c_{111} + 2.310\,.10^{-12}\,c_{112} + 0.727\,.10^{-12}\,c_{114} + 1.454\,.10^{-12}\,c_{124} - 4.620\,.10^{-12}\,c_{222}$$

$$k^R_{124} = f_{124} + 2.310\,.10^{-12}\,c_{114} + 6.930\,.10^{-12}\,c_{124} - 0.727\,.10^{-12}\,c_{155}$$

$$k^R_{134} = f_{134} + 4.620\,.10^{-12}\,c_{134} - 0.727\,.10^{-12}\,c_{344}$$

$$k^R_{144} = f_{144} + 2.310\,.10^{-12}\,c_{144} - 2.310\,.10^{-12}\,c_{155} - 0.727\,.10^{-12}\,c_{444}$$

$$k^R_{315} = f_{315}$$

$$k^R_{11} = l_{11} + 1.205\,.10^{-12}\,c_{111} - 0.603\,.10^{-12}\,c_{112} - 0.379\,.10^{-12}\,c_{114} + 0.379\,.10^{-12}\,c_{124} + 0.060\,.10^{-12}\,c_{144} - 0.603\,.10^{-12}\,c_{222}$$

$$k^R_{13} = l_{13} + 2.00\,l_{44} + 1.205\,.10^{-12}\,c_{113} - 1.205\,.10^{-12}\,c_{123} - 0.759\,.10^{-12}\,c_{134} + 0.060\,.10^{-12}\,c_{344}$$

$$k^R_{14} = l_{14} - 2.411\,.10^{-12}\,c_{124} - 0.379\,.10^{-12}\,c_{144} + 0.379\,.10^{-12}\,c_{155} + 0.060\,.10^{-12}\,c_{444}$$

$$k^R_{31} = l_{31} + 2.00\,l_{44}$$

$$k^R_{33} = l_{33}$$

$$k^R_{41} = l_{41}$$

$$q^R_{111} = \kappa_{111} - 2.045\,.10^{-23}\,l_{12} - 2.465\,.10^{-35}\,c_{111} + 1.233\,.10^{-35}\,c_{112} + 0.776\,.10^{-35}\,c_{114} - 2.329\,.10^{-35}\,c_{124} - 0.367\,.10^{-35}\,c_{144} + 0.244\,.10^{-35}\,c_{155} + 1.234\,.10^{-35}\,c_{222} + 0.038\,.10^{-35}\,c_{444}$$

Table III. Nonlinear parameters calculable by the resonator method. The parameters are combinations of the third-order nonlinear material constants of quartz including electroelastic constants f_{ijk}, third-order elastic constants c_{ijk}, electrostrictive constants l_{ij} and third-order permittivity κ_{111}. The first eight parameters (k^R_{ijk}) are in N/(V.m), the next six (k^R_{ij}) are dimensionless, the last parameter (q^R_{111}) is in F/V. The numerical coefficients at c_{ijk} in parameters k^R_{ijk}, k^R_{ij} and q^R_{111} are in m/V, m^2/N and Fm^2/NV, respectively. The numerical coefficient at l_{12} in q^R_{111} is in F/V. The remaining numerical coefficients are dimensionless. The number and choice of indices of k^R_{ijk}, k^R_{ij} and q^R_{111} correspond to the first material constant in their definition.

Given the multitude of the modes of vibration no exhaustive study of the true potential of the resonator method is available. While there is some doubt that it can be extended to include the isolated third-order elastic constants, there is little doubt that the application of biasing fields in the directions other than the resonator thickness can provide additional results to equal the contribution of the transit-time method.

COMPARING CONTRIBUTIONS OF THE TWO METHODS

The nonlinear constants obtained from one experimental data set are frequently verified by demonstrating how well they predict another data set. However, the predicting functions are typically quite complicated. They do not show explicitly which

k^R_{111}	2.38 ± 0.05	k^R_{11}	-3.19 ± 0.91
k^R_{113}	0.29 ± 0.07	k^R_{13}	-8.62 ± 2.59
k^R_{114}	0.63 ± 0.04	k^R_{14}	-2.43 ± 0.54
k^R_{122}	-0.74 ± 0.03	k^R_{31}	-12.40 ± 2.99
k^R_{124}	1.38 ± 0.02	k^R_{33}	-8.39 ± 6.79
k^R_{134}	1.71 ± 0.03	k^R_{41}	-4.42 ± 0.67
k^R_{144}	-0.03 ± 0.03	$q^R_{111}.10^{20}$	-0.29 ± 0.24
k^R_{315}	-0.79 ± 0.03		

Table IV. Nonlinear parameters obtained by the resonator method. Parameters k^R_{ijk} are in N/(V.m), k^R_{ij} are dimensionless, q^R_{111} is in F/V. The errors are standard errors. Given for room temperature, right-hand quartz and the frame of reference according to IEEE Standard of 1978[13].

nonlinear constants are sufficiently strongly represented to be verified if the prediction is successful. Also the verification may apply to a linear combination of constants rather than to their individual values. For this reason the verification by prediction may be misleading and can be easily misintepreted.

A direct comparison of nonlinear constants calculated independently from each data set is much preferred as it provides an automatic protection against these inherent risks. Such a comparison is presented in Table V for the results obtained independently by the resonator and the transit-time method.

The definitions of the nonlinear parameters in Tables I and III indicate that the results obtained by the transit-time method and the resonator method are not fully comparable in the form shown in Tables II and IV. To provide a complete set of comparable results for Table V, the columns of matrix M in system (5) pertaining to the transit-time method had to be suitably rearranged and the least-squares fit re-done for the adjusted system.

PURE MATERIAL CONSTANTS

As can be seen from Tables I and III the potential of both the transit-time method and the resonator method to produce the values of pure material constants (1)-(4) on their own is limited to one electroelastic constant and several electrostrictive constants. The rest of the results provided by the two methods are only values of nonlinear parametersformed by linear combinations of the pure constants. These combinations cannot be resolved without assistance from a source external to the two methods.

Such assistance is available in the form of a complete set of the third-order elastic constants [3]. Using their values the linear combinations of the third-order elastic constants can be removed from the calculated parameters k^T_{ijk}, k^T_{ij}, q^T_{111}, k^R_{ijk}, k^R_{ij}, and q^R_{111} in Tables II and IV to isolate the pure values of the electroelastic, electrostrictive and third-order permittivity constants (2)-(4). The results obtained for the transit-time and the resonator method are shown in Table VI and VII, respectively.

There exists no procedure of similar simplicity to transform the standard error estimates of the parameters in Tables II and IV into the standard errors of the pure material constants. As a results the available standard errors of the pure constants

resonator method		transit-time method	
k^R_{111}	2.38 ± 0.05	k^T_{111}	2.33 ± 0.06
k^R_{113}	0.29 ± 0.07	k^T_{113}	0.17 ± 0.06
k^R_{114}	0.63 ± 0.04	k^T_{114}	0.68 ± 0.04
k^R_{122}	-0.74 ± 0.03	k^T_{122}	-0.73 ± 0.05
k^R_{124}	1.38 ± 0.02	k^T_{124}	1.39 ± 0.02
k^R_{134}	1.71 ± 0.03	k^T_{134}	1.71 ± 0.03
k^R_{144}	-0.03 ± 0.03	k^T_{144}	0.04 ± 0.02
k^R_{315}	-0.79 ± 0.03	k^T_{315}	-0.89 ± 0.02
k^R_{11}	-3.19 ± 0.91	k^T_{11}	4.92 ± 6.04
k^R_{13}	-8.62 ± 2.59	$k^T_{13} + 2k^T_{44}$	13.13 ± 3.75
k^R_{14}	-2.43 ± 0.54	k^T_{14}	-5.93 ± 2.54
k^R_{31}	-12.40 ± 2.99	$k^T_{31} + 2k^T_{44}$	0.85 ± 3.18
k^R_{33}	-8.39 ± 6.79	k^T_{33}	-3.34 ± 4.39
k^R_{41}	-4.42 ± 0.67	k^T_{41}	-1.44 ± 2.21
$q^R_{111}.10^{20}$		$(q^T_{111} - 2.045.10^{-23}\, k^T_{12}).10^{20}$	
	-0.29 ± 0.24		3.72 ± 2.41

Table V. Comparison of the nonlinear parameters calculated by the resonator and the transit-time method. Each set of parameters is obtained independently of any previously determined nonlinear constants. Parameters k^R_{ijk} and k^T_{ijk} are in N/(V.m), k^R_{ij} and k^T_{ij} are dimensionless, q^R_{111}, q^T_{111} and the numerical coefficient at k^T_{12} are in F/V. The errors are standard errors. Given for room temperature, right-hand quartz and the frame of reference according to IEEE Standard of 1978[13].

(2)-(4) can be given only for those constants which appear as pure constants already in Tables II and IV.

Most of the values of the electroelastic, electrostrictive and third-order dielectric constants in Tables VI nad VII cannot be determined without the values of the third-order elastic constants [3]. As such they cannot be considered to be an exclusive product of the transit-time or the resonator method. They depend on the results and the attendant analysis of another experiment. The degree of the dependence varies from case to case. In some cases it is quite substantial. Consequently, any uncertainties in the values of the third-order elastic constants must also be viewed as uncertainties in the present values of the remaining nonlinear constants (2)-(4). This illustrates the prominent role played by the third-order elastic constants in the process of determination of the third-order nonlinearities in general.

The values of the pure nonlinear constants listed in Tables VI and VII can be obtained directly by the least-squares fit performed on the linear systems (5) which produced the values of the nonlinear parameters in Tables II and IV. It only requires that the values of the third-order elastic constants (3) be substituted into (5) before the least-squares fit is executed. However, this approach obscures the nature of the individual contribution of the transit-time and the resonator method as well as the character of the dependence of the nonlinear constants in Tables VI and

f_{111}	2.14	l_{11}	4.83
f_{113}	-0.54	l_{12}	-8.33
f_{114}	0.24	l_{13}	14.24
f_{122}	-1.12	l_{14}	-5.89
f_{124}	0.76	l_{31}	2.32 ± 2.05
f_{134}	1.64	l_{33}	-3.34 ± 4.39
f_{144}	0.09	l_{41}	-1.44 ± 2.21
f_{315}	-0.89 ± 0.02	l_{44}	-0.74 ± 1.79
κ_{111}	$3.71 \cdot 10^{-20}$		

Table VI. Nonlinear material constants obtained by the transit-time method. f_{ijk} are electroelastic constants in N/(V.m), l_{ij} are electrostrictive constants (dimensionless) and κ_{111} is third-order permittivity in F/V. The errors are standard errors. Given for room temperature, right-hand quartz and the frame of reference according to IEEE Standard of 1978[13].

VII on the third-order nonlinear constants. It also leads to a misinterpretation of the calculated standard errors which are then erroneously associated with the pure constants in Tables VI and VII.

f_{111}	2.18	l_{11}	-3.28
f_{113}	-0.42	$l_{13} + 2l_{44}$	-8.98
f_{114}	0.19	l_{14}	-2.38
f_{122}	-1.14	$l_{31} + 2l_{44}$	-12.40 ± 2.99
f_{124}	0.76	l_{33}	-8.39 ± 6.79
f_{134}	1.64	l_{41}	-4.42 ± 0.67
f_{144}	0.02	$\kappa_{111} - 2.045 \cdot 10^{-23}\, l_{12}$	$-0.29 \cdot 10^{-20}$
f_{315}	-0.79 ± 0.03		

Table VII. Nonlinear material constants obtained by the resonator method. f_{ijk} are electroelastic constants in N/(V.m), l_{ij} are electrostrictive constants (dimensionless) and κ_{111} is third-order permittivity in F/V. The errors are standard errors. Given for room temperature, right-hand quartz and the frame of reference according to IEEE Standard of 1978[13].

COMPARISON OF THE PURE CONSTANTS

Tables VI and VII invite a comparison of the the electroelastic, electrostrictive and third-order dielectric constants determined by the transit-time method and by the resonator method. It can be seen that some of them are in a nearly perfect agreement; they are the values of f_{111}, f_{114}, f_{122}, f_{124}, f_{134} and also f_{144}. In other cases, such as $l_{13} + 2l_{44}$ or $l_{31} + 2l_{44}$, and, to a lesser degree, also elsewhere, no satisfactory agreement can be demonstrated.

Most of the constants in Tables VI and VII which are now compared depend on the third-order elastic constants. However, the degree of their agreement is completely independent of these constants. This will be sufficiently illustrated using one example.

Suppose it is required to compare the combination of constants $l_{13} + 2l_{44}$ obtained by the resonator method (Table VII, second column) with a similar combination formed by the results obtained obtained by the transit-time method (Table VI). Using Table III valid for the resonator method, it can be written (6)

$$l^R_{13} + 2l^R_{44} = k^R_{13} - 1.205 \cdot 10^{-12} c_{113} + 1.205 \cdot 10^{-12} c_{123} + 0.759 \cdot 10^{-12} c_{134} - 0.060 \cdot 10^{-12} c_{344} .$$

Using Table I valid for the transit-time method, one can find

$$(7)\quad l^T_{13} + 2l^T_{44} = k^T_{13} + 2k^T_{44} - 1.205 \cdot 10^{-12} c_{113} + 1.205 \cdot 10^{-12} c_{123} + 0.759 \cdot 10^{-12} c_{134} - 0.060 \cdot 10^{-12} c_{344} .$$

Superscripts R (standing for the 'resonator' method) and T (standing for the 'transit-time' method) have been added to the quantities on the left-hand side of Eqs.(6) and (7) to indicate their origin.

The difference between Eqs.(6) and (7)

$$(8)\quad (l^R_{13} + 2l^R_{44}) - (l^T_{13} + 2l^T_{44}) = k^R_{13} - (k^T_{13} + 2.00\, k^T_{44})$$

formally represents the difference between comparable nonlinear constants determined by the resonator and the transit-time method. As can be seen the difference depends exclusively on the parameters k^R_{13}, k^T_{13} and k^T_{44}, i.e. on the values which have been determined solely by the resonator and the transit-time method with the exclusion of any external influence.

A similar comparison can be made for all 15 material constants or their combinations appearing in Table VII and with the same outcome: the difference between the nonlinear constants in Tables VII and VI does not depend on the third-order elastic constants used for their calculation. The difference will remain the same for any random choice of the third-order elastic constants.

CONCLUSION

The individual contribution of the experimental data sets provided by the transit-time and the resonator method towards the determination of the electromechanical nonlinearities in quartz consists of electromechanical constants, respectively. In addition to it each data set provides the values of twelve combinations of other third-order electromechanical constants. The results produced by the two methods are independent of each other and of any interference from another source of nonlinear constants of quartz. For this reason any discrepancies observed there represent a problem the solution of which must be sought within the two experimental methods.

Neither the transit-time method nor the resonator method are particularly well suited for the determination of the pure third-order elastic constants. Given that the third-order elastic constants can be taken from another source, the two methods are used most effectively to determine the rest of the electromechanical nonlinearities, i.e. the complete tensors of electroelasticity, electrostriction and third-order permitivity.

An excellent agreement has been obtained for most of the components of the electroelastic tensor. On the other hand several electrostrictive constants and the third-order permittivity are at variance.

The fact that a substantial portion of the values of the above constants depends on the third-order elastic constants imparts to the latter constants a position of prominence and underlines the fundamental importance of the work by Thurston, McSkimin and Andreatch [3] completed a quarter of a century ago.

The dependence on the third-order elastic constants complicates the proper assessment of the

standard errors of the calculated constants. The errors which are calculable indicate that some of the values are not particularly reliable.

It is interesting to note that the differences between the electroelastic, electrostrictive and third-order permittivity constants as calculated by the transit-time and the resonator method are not affected in any way by their dependence on the third-order elastic constants. This is just another way of arriving at the conclusion stated above: that the responsibility for the observed discrepancies rests with the resonator and the transit-time method.

The degree of agreement in the values of the nonlinear constants calculated by the two methods is, first of all, a measure of the intrinsic consistency of the two methods. Its interpretation as a confirmation of the correctness of their values in an absolute sense is conditional upon the reliability of the third-order elastic constants used for their calculation.

Before the problem of the nonlinearities in quartz is solved, the source of the discrepancies shown in this paper needs to be located and explained. Some of the standard errors should also be reduced to make the calculated constants more reliable. To this end the execution and interpretation of the resonator and the transit-time experiments will have to be further scrutinized.

It should also be noted that the numerical values obtained for the third-order nonlinear constants necessarily depend on the choice of the second-order (linear) constants. Their effect should also be properly investigated.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of Pauline Hruska with the final stages of preparation of this paper. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Faculty of Arts, York University.

REFERENCES

[1] Reider G. A., Kittinger E., Tichy J., J.Appl.Phys. 53(1982), pp.8716-8721.
[2] Hruska K., Czech.J.Phys. B12(1962), pp.338-353.
[3] Thurston R. N., McSkimmin H. J. and Andreatch, J.Appl.Phys. 37(1966), pp. 267-275.
[4] Kittinger E., Tichy J., and Friedel W., J.Appl.Phys. 60(1986), pp.1465-1471.
[5] Kittinger E. and Tichy J., J.Acoust.Soc.Am., 83(1988), pp.647-651.
[6] Brendel R., J.Appl.Phys. 54(1983), pp. 5339-5346.
[7] Hruska C. K. and Brendel R., J.Appl.Phys. 67(1990), pp.1676-1679.
[8] Baumhauer J. C., Tiersten H. F., J.Acoust.Soc.Am. 54(1973), pp.1017-1034.
[9] Nelson D. F., J.Acoust.Soc.Am. 63(1978), pp. 1738-1748.
[10] Tiersten H. F., and Ballato A., J.Acoust.Soc.Am. 73(1983), pp.2022-2033.
[11] McSkimin H. J., Andreatch P. Jr, Thurston R. N., J.Appl.Phys. 36(1965), pp.1624-1632.
[12] Bechmann R., Phys.Rev. 110(1958), pp. 1060-1061.
[13] IEEE Standard on Piezoelectricity, IEEE Std 176-1978. Available from The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017.
[14] Hruska C. K., J.Appl.Phys. 68(1991), pp.41-45.
[15] Hruska C. K., Proc. IEEE 1990 Ultrasonics Symposium, Honolulu, Dec. 1990, in press. Available from The Institute of Electrical and Electronics Engineers, Inc., New York, NY. 10017.
[16] Hruska K. and Khogali A., IEEE Transactions on Sonics and Ultrasonics, vol. SU-18, 1971, pp.171-176.
[17] Kinigadner A., Ph.D.Thesis, University of Innsbruck, Austria, 1978.
[18] Hruska C. K., Merigoux H., Kucera M., J.Can.Ceram.Soc. 57(1988), pp.53-57.
[19] Hruska C. K. and Brendel R., J.Appl.Phys. 65(1989), pp.715-717.
[20] Hruska C. K., Proc. 12th Piezoelectric Devices Conference and Exhibition, Kansas City, Sept. 1990, pp. 8-14. Available from the Electronics Industries Association, Washington, DC.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4A Number 25

PRESENT STATE OF $GaPO_4$-RESEARCH

P. Krempl, J. Stadler, W. Wallnöfer, W. Ellmeyer, R. Selic

AVL LIST Ges.m.b.H., Physics Division, Graz, Austria

Abstract

Due to the strong stability of the α-phase and the larger electromechanical coupling, as compared with quartz and berlinite, quartz-homeotypic $GaPO_4$ has been recognized as one of the most promising piezoelectrics allowing the extension of quartz technologies to higher temperatures as well as larger bandwidths.

Besides the great efforts, which are necessary to make such materials producible, a lot of basic research has to be performed in view of future applications. This contribution gives a short review of the present knowledge of the physical properties of $GaPO_4$ together with some new data on thermal, electroelastic and optical constants.

Further progress has been achieved in investigating $GaPO_4$ as a material for future sensors and resonators. Studies concerning the temperature dependence of thickness-shear resonators underline the attractivity of the new material for elevated temperatures. The basic technologies for SAW-devices have been adapted to $GaPO_4$.

Introduction

Galliumorthophosphate has been presented to this forum since 1988 [1-3] together with other quartz-homeotypic materials less known than berlinite ($AlPO_4$). Among its most attractive properties, its large piezoelectric strain constant d_{11} and its high temperature stability of the alpha-quartz-phase to the β-cristobalite transition at 1206 K [9] make it a serious candidate for future sensor applications with challenging tasks.
Besides the availability of a sufficient amount of crystals with good quality, the precise knowledge of many physical properties is one of the most important preconditions for the development of new products. The main purpose of this paper is to give a collection of data on the physical properties of $GaPO_4$, most of them arising from the research work performed at the AVL and cooperating institutes.

Since the knowledge of the elastic constants is an indispensable prerequisite for most applications, a new determination of these constants was necessary. The temperature dependence of these properties will be the subject of further studies.

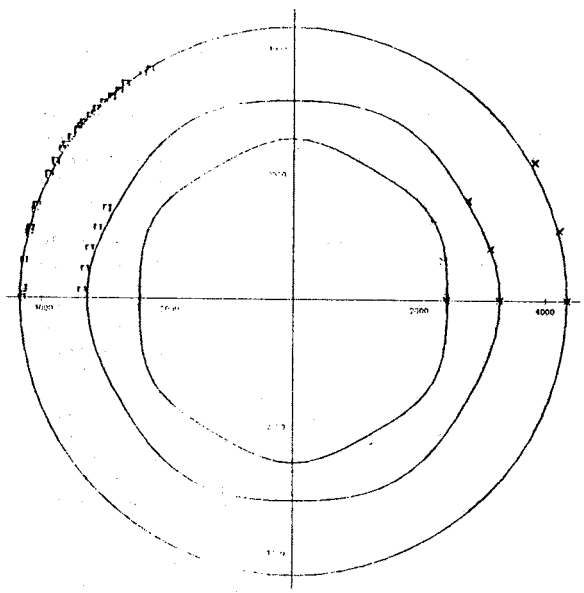
Determination of elastic tensor with Brillouin spectroscopy

To determine the complete elastic tensor of $GaPO_4$ the Brillouin spectra were studied. The measurements were done with 514 nm light on a triple-pass Fabry-Perot interferometer with capacitive resonator regulation and automatic optimization of the mirrors' parallelity.
The samples were cubes with 1.5 - 2 mm side length. All measurements were performed in 90° scattering geometry to minimize the background scattering. The sample was rotated up to 30° in both directions; this allowed measurements of sound velocities in directions deviating up to 18° from the bisectors of the surface normal vectors.
On some spectra, only two of the three acoustic modes were visible. Since birefringence is small (~ 1%), a mean refractive index of 1.62 was used to calculate the sound velocities from the frequency shifts of the Brillouin doublets.
The measured velocities in the three main planes are shown in Fig. 1. The drawn lines are calculated from a best fit of the elastic tensor to our data. For comparison, the velocity data from Brillouin measurements by Huard [11] are included, which are in good agreement with our measurements. However, in the xy-plane (Fig. 1) Huard interpreted the transversal mode which he measured as the slowest mode, not as the middle one as we do. Thus he derives a quite different set of elastic constants. We confirmed our interpretation by careful measurements of the transversal modes in many directions and by comparison with our resonator measurements.
In the xz-plane (Fig. 1b), the ranges around the x and z axes were measured with a cube rotated 45° around the y axis. Thus velocities in all directions could be measured.
In the yz-plane (Fig. 1c), the pure transverse mode is the one most used for resonators. A comparison of our resonator frequency measurements with the Brillouin data shows a good accordance.
The elastic constants c^E which give the best fit for the velocities are shown in Table 1.

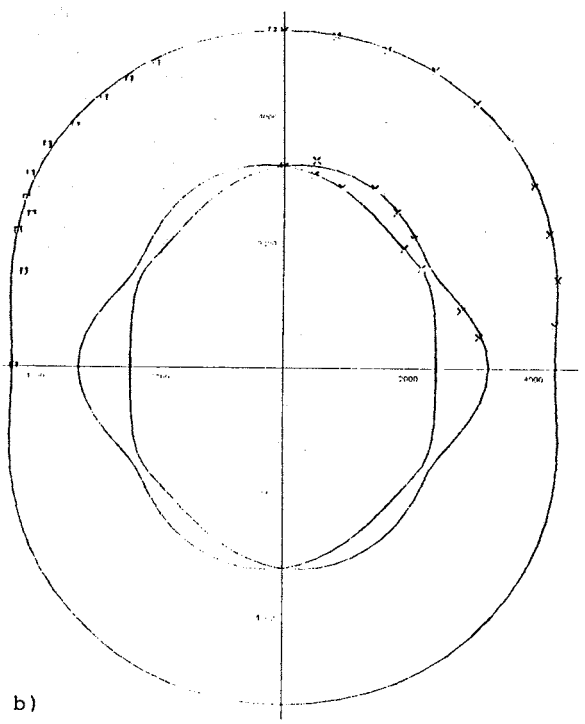
Uncertainties are about 1 GNm^{-2} for all constants. This set disagrees significantly with earlier results from resonator measurements [2]. A comparison with the isotypes quartz and $AlPO_4$ [12] shows remarkable differences. c_{12} and c_{13} are greater in $GaPO_4$ by a factor of 2 or more, while c_{14} is a quarter or less. This suggests considerable deviations in the elastic and thermoelastic properties.

Physical properties of $GaPO_4$

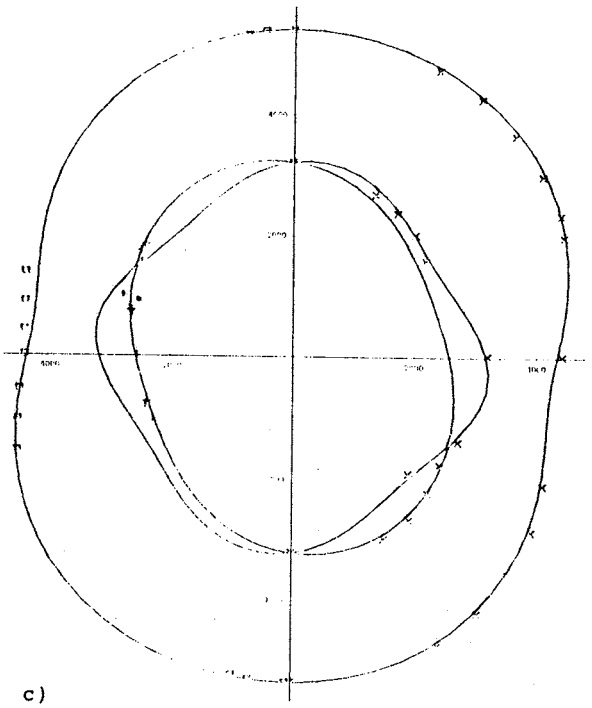
Table 1 displays a consistent set of electromechanical constants of galliumorthophosphate. The piezoelectric strain constants and the dielectric permittivity for constant stress have been determined by the direct piezoelectric effect and capacitance measurements.



a)



b)



c)

Fig. 1: Sound velocities in the main planes of $GaPO_4$; lines: calculated;
x: Brillouin, this work
□: Brillouin [11]
+: resonators, this work
a) xy-plane; b) xz-plane; c) yz-plane

From a theoretical point of view, the elastic stiffness coefficients $c^E_{\mu\nu}$ and the $d^2_{11}/\varepsilon^S_{11}$-term could be directly obtained from the sound velocities in the main directions. However, this attempt failed due to the sensitivity of the systems of equations to small experimental errors. The evaluation has therefore started with an estimated piezoelectric stiffening and iterative refinement of the piezoelectric stiffening with the new values for the stiffness constants. The small value of the elastic compliance constant s_{14}, which is about the half in $GaPO_4$ compared to quartz, indicates together with the high stability of the α-quartz form with a theoretical T_c at the melting temperature the improved stability of $GaPO_4$ against ferrobielastic twinning.

The thermic conductivity was measured by the comparative method [5]; a $GaPO_4$ and a quartz plate were put in series in a heat flow and the temperature decrease at both was measured.

The specific heat c_p was determined in two ways, with agreeing results:
- by directly measuring the amount of added heat and the temperature increase [5];
- by a differential scanning calorimeter [8].

The thermal expansion data were taken graphically from [13].

For certain applications piezoelectric devices can be exposed to nuclear radiation. In these cases, the resistivity against radiation damage is important.
A good measure of this resistivity is the inverse of the metamict damage cross-section D_M which has been measured for several α-quartz homeotyps by bombardement with 3 MeV argon-ions [6]. The derived resistivity gives 1 nm^{-2} for α-quartz and 1.4 nm^{-2} for $GaPO_4$, indicating that galliumorthophosphate is about 40% more resistant against nuclear radiation than α-quartz.

As known from infrared measurement as well as from theoretical calculation, the band gap of $GaPO_4$ is larger than that of quartz. This should also result in an increase of the intrinsic electrical resistance. However, due to impurities and crystal imperfection the experimental values of the specific electrical resistances have the same order of magnitude as measured in quartz, i.e. greater than 10^{14} Ωcm.

Studies on the linear and nonlinear optical properties of $GaPO_4$ [7] indicate that no phase matching for second harmonic generation occurs at frequencies of the visible light. The d_{11} (SHG) of $GaPO_4$ was measured to be about three times greater than the corresponding constant of α-quartz at RT. d_{11}(SHG) decreases from RT to 1038 K nearly linear with a temperature coefficient $1/d_{11}\ d(d_{11})/dT = -9.10^{-5}/K$. A transformation of α-$GaPO_4$ to the cristobalite form was found at 1350 K, i.e. 144 K about the transition temperature reported in the literature [9]. This shows that short heating above the transition temperature of 1206 K will not destroy the α-quartz form of $GaPO_4$.

Table 1: Physical properties of $GaPO_4$

Cristallographic Properties

Point group:	32		
Space group:	$P3_121$ (righthanded) $P3_221$ (lefthanded)		
Molecules/unit cell:	3		
Molecularweight (AMU)	164.69		
Lattice parameters at RT:		[4]	
a (nm)	0.4901		
c (nm)	1.1048		
Density (kg/m^3)	3570		
Mean tetrahedral tilt angles (degrees)	20.8(GaO_4) 25.6 (PO_4)	[1]	

Thermophysical Properties

Thermal expansion at RT ($10^{-6}/K$)	α_{11} α_{33}	5.3 1.2		
Heat capacity (J/kgK)	c_p	630	[5]	[8]
Thermal conductivity (10^{-4} W/mK)	λ_{11} λ_{33}	3.9 6.1		[5]
Transition temperature (K)		1206		
Transition to:		β-Cristobalite ($F\bar{4}3m$)		
Melting point (K)		1943		[9]
Radiation damage resistivity (Ar, 3MeV) (nm^{-2})		1.4		[6]

Electrical Properties

Electrical resistivity	$>10^{14}$		
Permittivity at 1 kHz	$\varepsilon^T_{11}/\varepsilon_0$ $\varepsilon^T_{33}/\varepsilon_0$	6.5 6.7	

Dielectric and Piezoelectric Properties

Permittivity	$\varepsilon^S_{11}/\varepsilon_0$ $\varepsilon^S_{33}/\varepsilon_0$	6.3 6.7	
Piezoelectric strain constants (pC/N)	d_{11} d_{14}	4.5 1.9	
Piezoelectric stress constants (C/m^2)	e_{11} e_{14}	0.20 0.03	

Elastic Properties

c^E_{11} (GNm^{-2})	67
c^E_{33} (GNm^{-2})	102
c^E_{44} (GNm^{-2})	37
c^E_{12} (GNm^{-2})	21
c^E_{13} (GNm^{-2})	25
c^E_{14} (GNm^{-2})	-4
s^E_{11} ($10^{-12}N^{-1}m^2$)	17.8
s^E_{33} ($10^{-12}N^{-1}m^2$)	11.4
s^E_{44} ($10^{-12}N^{-1}m^2$)	27.6
s^E_{12} ($10^{-12}N^{-1}m^2$)	-4.8
s^E_{13} ($10^{-12}N^{-1}m^2$)	-3.2
s^E_{14} ($10^{-12}N^{-1}m^2$)	2.4

Optical Properties

Indices of refraction (RT)	n_0	n_e	[7]
440 nm	1.6207	1.6396	
530 nm	1.6135	1.6319	
600 nm	1.6074	1.6255	
670 nm	1.6031	1.6211	
1060 nm	1.5935	1.6112	

Second harmonic generation (10^{-12} m/V)	d_{11}	0.994	[7]
Temperature coefficient d_{11} (SHG)	$1/d_{11} \cdot dd_{11}/dT$	$9 \cdot 10^{-5}$	

Y-cut BAW resonators

In previous papers we showed the attractivity of $GaPO_4$ in resonator applications. We demonstrated the existence of temperature compensated cuts with a piezoelectric coupling larger than quartz [2].
Last year's paper dealt with our method of measuring the equivalent circuit parameters, the manufacture of the blanks and some other aspects of the energy trapping concept applied to $GaPO_4$ [3].
In this paper we give an example of a BT-cut frequency/temperature behaviour. We also present the first $GaPO_4$ resonator measurements at temperatures up to 670°C.

$GaPO_4$-resonator with BT-like frequency/temperature behaviour

The $GaPO_4$ Y-17,4° resonator possesses a parabolic frequency/temperature characteristic with a maximum of frequency at T_0 = 40.5°C.
From a quadratic fit of the frequency normalized to the inflection temperature T_0 according to $(f(T)-f(T_0))/f(T_0) = a(T-T_0)^2$ we get the second order coefficient a.
We found $a = -18.28 \cdot 10^{-9}\ K^{-2}$, which is less than half of the corresponding coefficient for quartz ($a = -40 \cdot 10^{-9}\ K^{-2}$, [10]).
The comparison with the quartz AT- and BT-cut is shown in Fig. 2 and Fig. 3.
From the coefficient a and Fig. 3 we see the lower temperature dependence of the $GaPO_4$-resonator as compared with a quartz BT-cut.

$GaPO_4$-resonators of temperatures up to 670°C

We made the first measurements at higher temperatures.
The frequency response of a $GaPO_4$-resonator at a temperature of 670°C is shown in Fig. 4.

This measurement on the admittance plane is presented in Fig. 5.
The frequency/temperature behaviour of a $GaPO_4$-resonator up to 630°C is given in Fig. 6.

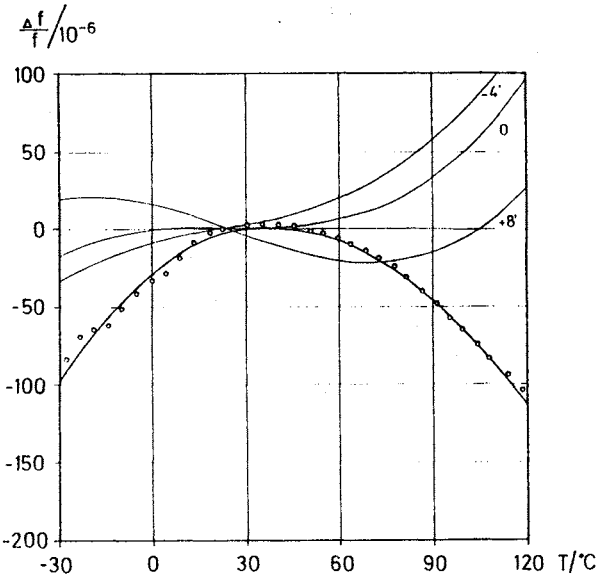


Fig. 2: Frequency temperature behaviour of $GaPO_4$ Y-17.4° resonator in comparison to AT-quartz cuts

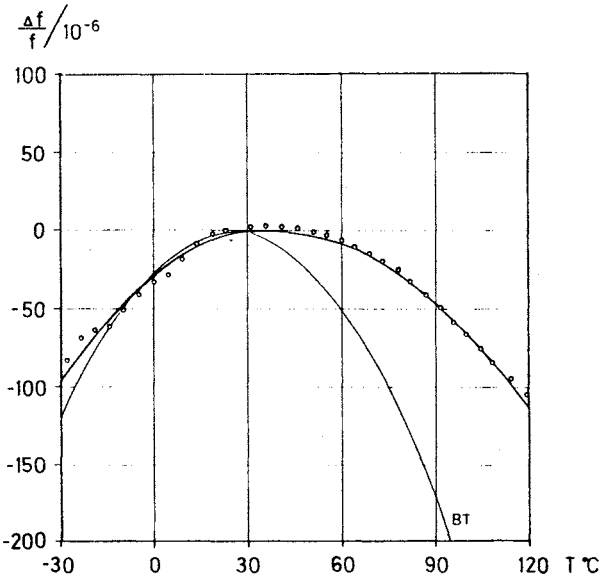


Fig. 3: Frequency temperature behaviour of $GaPO_4$ Y-17.4° resonator in comparison to BT-quartz cut

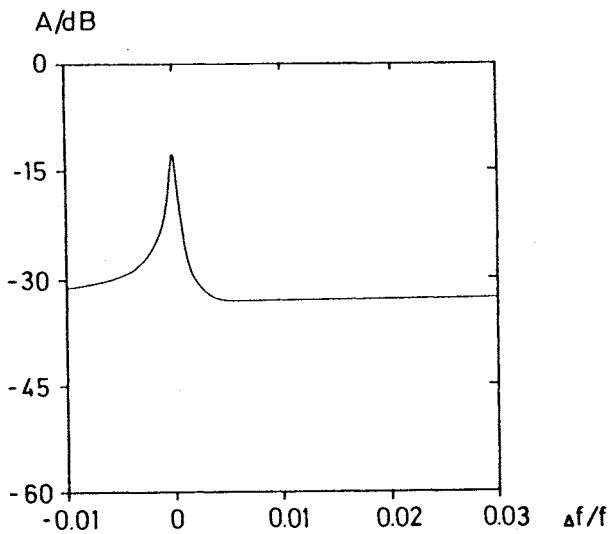


Fig. 4: Frequency response of a $GaPO_4$ Y-cut resonator at a temperature of 670°C

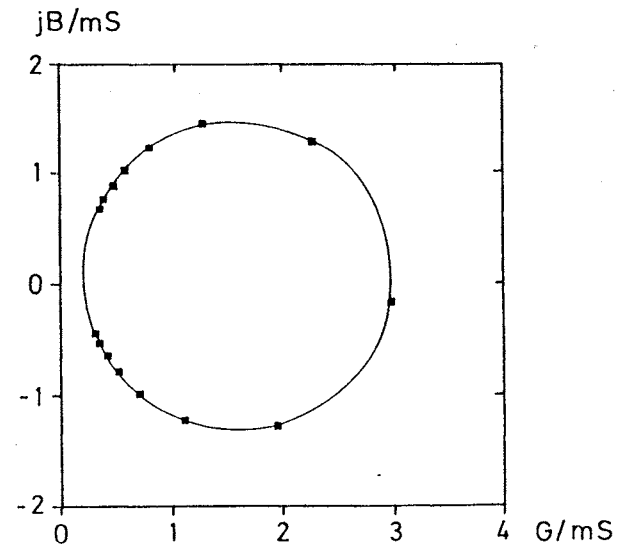


Fig. 5: Admittance circle of a $GaPO_4$ Y-cut resonator at a temperature of 670°C

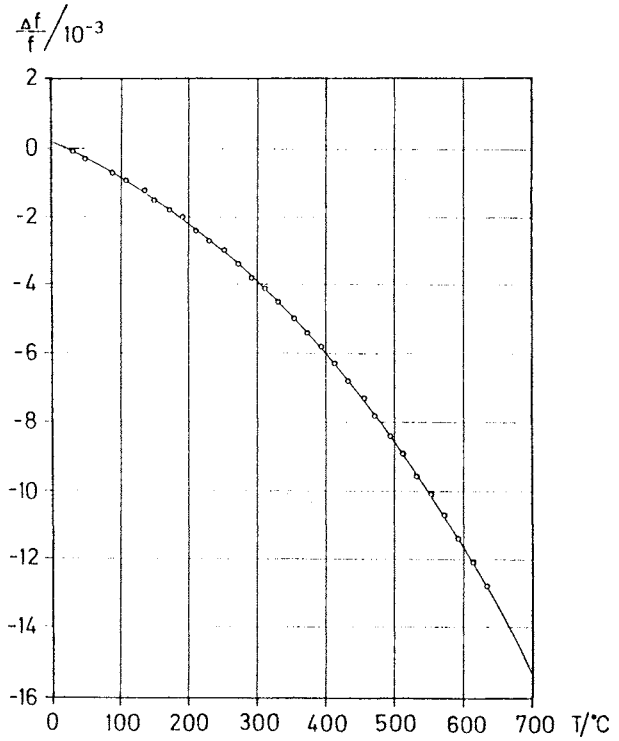


Fig. 6: Thermal behaviour of a $GaPO_4$ Y-18.2° resonator up to about 630°C

Surface acoustic waves

Surface acoustic wave devices are not only an important potential application of $GaPO_4$, but also a valuable means of studying the properties of the material.
Interdigital transducers were applied on a polished $GaPO_4$ surface by UV-lithography with standard AZ resist and lift off technique. Resist bulges at the edges of the small substrates could be eliminated by bevelling the edges. An HMDS film was used to improve the adhesion of the resist film. Structures with line widths down to 1 μm were produced. Measurements of the SAW velocities and couplings are on the way.

Conclusion

The physical constants reported in this contribution are in some cases quite different than those reported in earlier work. Despite our feeling that the new data have a higher confidence level than these reported in the literature, further research work has to be done to improve the accuracy and to show, which constants are influenced by the crystal quality.

The BT-cut of $GaPO_4$ shows a better frequency/temperature behaviour as compared to the quartz BT-cut.
Furthermore, resonance measurements of rotated $GaPO_4$ Y-cut resonators up to 670°C demonstrate the suitability of $GaPO_4$ for high temperature applications.

Acknowledgements

We acknowledge the collaboration of A. Asenbaum at performing the Brillouin measurements and of F. Seifert and his group at the SAW study.
The work on resonators was supported by the Austrian Forschungsförderungsfonds für die Gewerbliche Wirtschaft (Project-No. 3/7394), which is gratefully acknowledged.

Literature

[1] G.F. Engel, P.W. Krempl
Proc.2nd Europ.Frequency and Time Forum (Neuchâtel, 1988) p. 827

[2] G. Engel, P. Krempl, J. Stadler
Proc.3rd Europ.Frequency and Time Forum (Besançon, 1989) p. 50

[3] J. Stadler, P. Krempl, G. Engel
Proc.4th Europ.Frequency and Time Forum (Neuchâtel, 1990) p. 567

[4] O. Baumgartner, A. Preisinger, P. Krempl, H. Mang
Zeit.Kristallogr.168 (1984) p. 83

[5] S. Defregger
AVL Physik-Notiz IV-4, 54/89

[6] F.G. Karioris, K. Ramasami, K.A. Gowda, L. Cartz
Phil.Mag.A52, No. 4 (1985) p. 525

[7] S. Defregger, G. Engel, P. Krempl
to be published in Phys.Rev.B

[8] W. Lacom
AVL Int.Report (OEFZS-A-1327)
Österr.Forschungszentrum Seibersdorf 1988

[9] E.C. Shafer, R. Roy
J.Am.Ceramic.Soc.39, 10 (1956) p. 330

[10] B. Parzen
Design of crystal and other harmonic oscillators
Wiley, New York (1983) p. 106

[11] F. Huard
Dissertation Montpellier 1985

[12] Landolt-Börnstein, Tables, New Series, III/11 and 18

[13] K. Kosten, H. Arnold
Z.Krist.152 (1980) p. 119

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4A Number 26

ELECTRODIFFUSION (SWEEPING) AND IRRADIATION EFFECTS IN CRYSTALLINE QUARTZ

Harish Bahadur

National Physical Laboratory, Hillside Road, New Delhi-110012, INDIA

ABSTRACT

A study of electrodiffusion (sweeping) and irradiation 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 the hydrothermal synthesis of high quality cultured quartz crystals. In particular, various OH^- related defects have been monitored using infrared absorpiton measurements in the 3100-3700 cm^{-1} range. Essentially, crystals of both origins have same spectral features except that the bands are much more intense in the Brazilian quartz than the Arkansas quartz studied in the present investigations. Some of the low strength bands in the Brazilian crystal are not observable in the Arkansas sample. Thus, the Arkansas quartz used was a puer material in relation to the Brazilian one. Irradiation effects at 77K and 300K on as-received (or Li-swept), Na-swept and H-swept samples of matching thickness from crystals of both origins have been investigated and compared. The study included major bands designated N_1-M^+, where M^+ is either Li^+ or Na^+ and e_1 and e_2 (Al-OH^-) bands. Measurements carried out on both Arkansas and Brazilian crystals show that for the same radiation dose (2Mrad used at every stage) the N_1-M^+ centers exhibit a relatively fast decay at 300K irradiation than at 77K. On the other hand the N_1-H^+ centers do not show any depletion at 300K irradiation while at 77K irradiation their strength is significantly reduced to about 10% of their original value. Both Al-OH^- bands in natural quartz show their irradiation characteristics that are identical to those of high purity cultured crystals. For both samples, the N_1-Na^+ centers show a faster depletion to the similar extent than the corresponding N_1-Li^+ centers at 300K irradiation. Occasionally, after irradiation there appear some low strength OH^- infrared bands in all versions of swept and unswept natural quartz. This occurs at the expense of already existing OH^- defect centers.

INTRODUCTION:

For various accurate timing devices the quartz oscillator crytals play a very significant role as precision oscillators. 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 as well as 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 demand 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 interest. Also, it will be of interest to investigate the role of different impurities present in natural quartz which in most cases is the starting material for growth of high purity cultured quartz.

A variety of interesting point defects are found in quartz crystals (2,4-11). 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 respectively are simply designated as Al-M^+ where M^+ is either Na^+, Li^+ or K^+ and Al-OH^- in case of an interstitial proton compensating the aluminum; the protons forms an OH^- molecule with the adjacent oxygen. The Al-hole centers are designated as Al-h^+ or $[Al_e^+]^0$. In addition to the Al-associated centers there are other types of point defects in quartz (8). They include , for example, oxygen vacancies, Ge-associated centers and numerous OH^- related defects.

A variety of techniques are used to characterize the point defects in quartz. For example, the 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 the ESR techniques. Martin (12) has shown that the acoustic loss peaks at 23K, 100K and 135K are related to the production of Al-hole centers.Recently, an acoustic loss peak at 246K in 5MHz Ge-doped quartz has been found by Euler and Kahan (13) to be due to a Ge-Li-electron center.

Most of the OH^- related defects in quartz are the stretching vibrations of OH^- molecules trapped at a variety of unknown sites and are detected by infrared absorption technique.The OH^- intensities are very sensitive to 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 defects bands are to be monitored.

While cultured quartz has always a set of "growth defects" which trap hydrogen and give rise to infrared absorptin bands at 3350, 3400, 3437 and 3581 cm^{-1}, the number of bands in natural quartz can be quite high and uncorrelated in strength (14-16). Often the bands in cultured quartz are designated as s_1, s_2, s_3, and s_4 (17-23). Bahadur (14) has shown that in a variety of natural quartz of different origin a set of major bands always present is N_1-Li^+ at 3476 cm^{-1}, Al-OH^-; e_1 at 3306 and e_2 at 3367 wavenumbers. 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 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 atmosphere or irradiation by ionizing radiations at temperatures above or about 200K (15,24,25). On the other hand, electrodiffusion or irradiation in natural quartz increases the strength of Al-OH^- centers in the limit to saturate all the substitutional aluminum compensated with hydrogen (14). Except for the Al-OH^- bands, defects responsible for other infrared bands have not been identified in terms of any specific atomic model. In cultured samples, recently, Martin et. al.(26) have reported that these defects can also trap alkalis. While the hydrogen trapped at these "growth defect" sites does not cause any anelastic relaxation, Li or Na 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 (5,7,27-40). While the protons under radiation field are mobile at all temperatures, the movement of alkalis starts only about and above 200K (5,7,25,41). If a quartz crystal is irradiated above 200K, the Al-M^+ centers are converted into a mixture of Al-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 ($\sim$77K), hydrogen atoms are formed due to dissociation of OH^- molecules followed by displacement of hydrogen to an isolated interstitial site where it becomes stably trapped (42). Also, in case of natural quartz where Al-OH^- centers are present in the as-received unswept state or cultured quartz (H-swept or irradiated), the low temperature irradiation converts the Al-OH^- centers into Al-hole centers (4,5,7).

Halliburton et.al (5) have carried out a remarkable and systematic parallel IR and ESR study on high purity quartz both unswept and H-swept. These authors (5) investigated radiation-induced mobility of alkali and hydrogen ions. In their experiment, they first monitored the intensity of the ESR signal of the Al-h^+ center and infrared bands in the as-received state of the crystal at 77K. The crystals were then irradiated at this temperature and measured. Later, the crystals were warmed to 300K and irradiated at this temperature. Al-hole and Al-OH^- centers were monitored. Finally,the crystals were again irradiated at 77K to monitor the relative strength of these centers. Annealing studies on these samples were also conducted. The various treatments were found to simply exchange one type of compensator for another at the Al-sites and the sum of the aluminum centers remained constant for a given sample. The direct accountability of all the aluminum ions in hydrogen swept quartz further confirmed the assignment of 3306 and 3367 cm^{-1} bands to the Al-OH^- center.These authors (5) found through their spectroscopic studies that the aluminum content of randomly selected bars of high quality quartz can vary by an order of magnitude (4,5,7).

In a recent study on infrared characterization of natural and cultured quartz, Bahadur (14) has shown that natural and quartz crystals have a number of infrared bands which are different from those observed in cultured quartz. Most of these bands in natural quartz in their as-received state were found to be associated with the presence of a Li^+ ion except the Al-OH^- bands. Exchange of Li^+ with Na^+ changed the band positions in natural crystals only; the band position in cultured crystals remained unaffected. H-sweeping resulted in enhancement of Al-OH^- bands in natural quartz while they were introduced in cultured crystals. In natural quartz crystals of different origins, it has been found (14) that the bands always present in the significant strength are at 3476, 3367 and 3306 wavenumbers. While the latter two bands are the known Al-OH^- defect centes, the band at 3476 cm^{-1} has been found (14) to be associated with the presence of Li and H together. Na-sweeping results in a prominent band at 3451 cm^{-1} along with the Al-OH^- centers. In the H-swept condition the major band positions occur at 3468 cm^{-1} and the Al-OH^- bands. In ref.(14) these major bands dependent on alkali and H or H-sweeping were designated as N_1-Li^+ for 3476 cm^{-1}, N_1-Na^+ for 3451 cm^{-1} and N_1-H^+ for the 3468 cm^{-1} band. Intense irradiation at 77K was done to suppress these bands significantly and their isochronal anneal data showed that they all recover by about 760K.

In the previous study (14) the effect of irradiation was carried out on unswept natural Brazilian quartz crystal on the OH^- related defects in the way done by Halliburton et.al (5,43). The crystals were measured at 77K, irradiated at this temperature and measured, warmed to and irradiated at 300K and measured Finally, they were again irradiated at 77K, following a 300K irradiation, and measured.Since the infrared spectra of unswept or Li-swept crystals were essentially the same, it will therefore be assumed that irradiation infrared characteristics of the two types of crystals, unswept and Li-swept, would also be the same. During the entire sequence of irradiation, the radiation dose was essentially the same at each step; 2 Mrad.

In the present paper we further describe the previously unreported results and show a comparison of our investigations carried out on the infrared absorption spectra of Brazilian and Arkansas quartz.The studies include the effect of irradiation at 77K on unswept (or Li-swept), Na and H-swept versions of the crystals. All the spectra presented in the following were recorded at 77K. In addition to the major bands the following description also incorporates comments on some of the smaller but prominent bands as well. The broad bands at 3200 and 3300 cm^{-1} in all spectra, which are due to intrinsic overtones of Si-O stretching vibrations of the quartz lattice, remained unchanged throughout the sequence of irradiations.

EXPERIMENTAL DETAILS:

Two natural stones, one from Brazil and the other from Arkansas mines were 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 thec-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 77K with temperature controllable to $\pm$ 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. 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 $\mu A/cm^2$ .The distance of the sample in cryostat and the window of the accelerator was 5in.; 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 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 30 V/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 20-24 hours for 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:

Figure 1 shows the infrared spectra in unswept or as-received versions of Arkansas and Brazilian quartz used in the present investigation. 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 significant strength in the Brazilian quartz are not observable in the Arkansas quartz. Thus, Arkansas quartz used in the present study is of better quality in relation to Brazilian quartz. Figure 2 represents the major bands observed in the Arkansas sample in its different swept versions. It may be observed that the N_1-Na^+ and N_1-H^+ bands arise respectively due to incorporation of Na and H ions and exchange of Li^+ ions from its unswept state. It can be seen that the alkali sweeping does not affect the position of e-bands.On the other hand, hydrogen sweeping increases the strength of both the e-bands indicating that the sweeping process has replaced the alkalis from Al-M^+ centers with protons. In fact, during the hydrogen sweep run all the aluminum has been compensated with protons as will be shown later on irradiation data of the H-swept crystals.

From the above, it has been observed that the IR spectrumof natural quartz is significantly altered after sweeping. In the case of cultured quartz the situation is somewhat different. While the spectra are not altered by alkali sweeping, the H-sweeping introduces additionally the e_1 and e_2 bands; the other band positions are unaffected. Such a case is clearly represented in Figure 3.

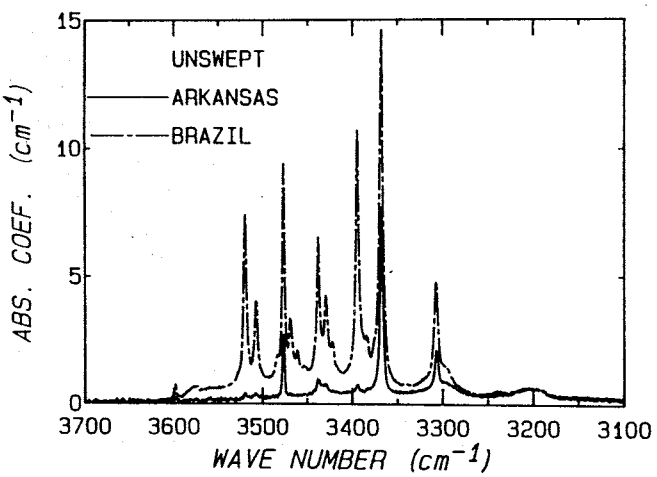


Figure 1: Infrared absorption spectra of natural Brazilian and Arkansas quartz crystals in their as-received states.

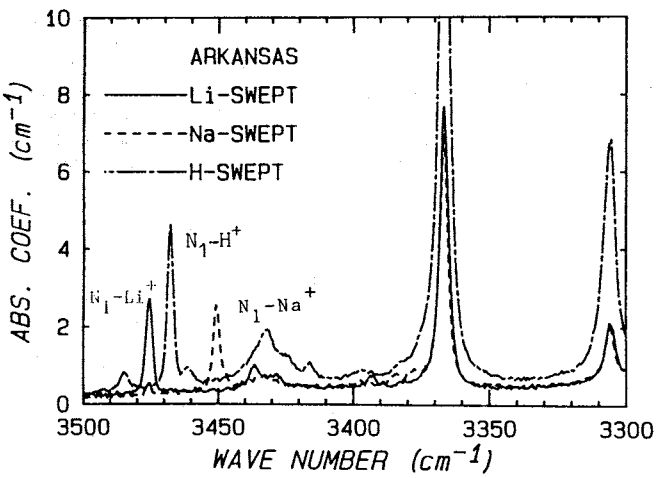


Figure 2: Infrared absorption spectra of the Arkansas natural quartz in unswept, Na-swept and H-swept states. The spectra depict all the major bands present in different versions

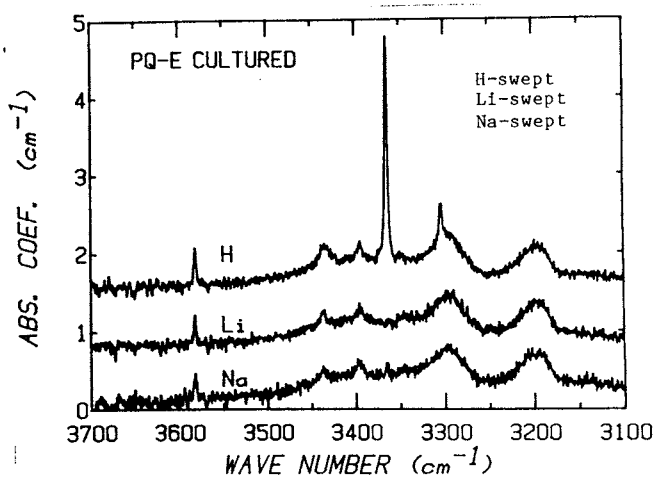


Figure 3: Infrared absorption spectra of a cultured quartz crystal in different swept versions. While the positions of growth-defect bands remain unaltered by any sweeping the H-sweeping introduces e and e bands in addition to the growth-defect bands.

A. Comparative Studies of Irradiation Characteristics of Unswept Brazilian and Arkansas Quartz:

Figures 4 and 5 respectively show the spectra upon irradiation sequence of unswept Brazilian and Arkansas quartz. In the following are described the effect of irradiation sequence on the infrared bands of these crystals.

At the first 77K irradiation the general feature was that all bands decreased in their strength significantly except the e_1 which had been found to grow upon low temperature irradiation both in natural (14) and cultured quartz (25). The small 3595 cm^{-1} band, which in the present investigation was found to exist in all swept versions of natural quartz, was removed in both the natural samples. There was an appearance of two new small bands at 3625 and 3618 wavenumbers in the Brazilian crystal only; the Arkansas quartz did not show any such bands. Perhaps, the non-appearance of these new small bands in the Arkansas quartz would be due to the fact that even to start with many of the low strength bands present in the Brazilian quartz were not observable in the Arkansas quartz. In both the cases the band at 3395 cm^{-1} exhibited a modest decrease at low temperature irradiation. A small shoulder at 3322 cm^{-1} may be seen at the higher frequency side of e_1 band in the Brazilian quartz only.

It may be seen from the spectra obtained after the first low temperature irradiation that while the 3476 cm^{-1} N_1-Li^+ band was reduced by about 10% in the Brazilian sample, the decrease in the strength of this band in the Arkansas quartz was relatively fast; about 35%. Regarding the other major bands, the 3367 cm^{-1} e_2 band decreased in nearly the same proportion to about 39% of their as-received strength in both the cases. The 3306 cm^{-1} e_1 band increased more in the Brazilian quartz than the Arkansas quartz. While the increase in the e_1 band intensity in Brazilian quartz was more than double (X 2.25), in the Arkansas quartz it did not increase beyond 1.5 times of its as-received strength.

Upon warming the quartz samples to room temperature, the 3595 cm^{-1} band in both the crystals was recovered. Small bands at 3625 and 3618 wavenumbers introduced at 77K irradiation stayed back. All other bands showed in general a recovery upto 85-95% of their as-received strength.

On room temperature irradiation the small bands at 3625 and 3618 wavenumbers which were introduced at low temperature irradiation in the Brazilian sample and stayed upon warming to room temperature, vanished at this stage. Another small band at 3595 cm^{-1} which got removed upon 77K irradiation in both Arkansas and Brazilian quartz remained unaffected at this irradiation. The 3520 cm^{-1} band in the Brazilian crystal was

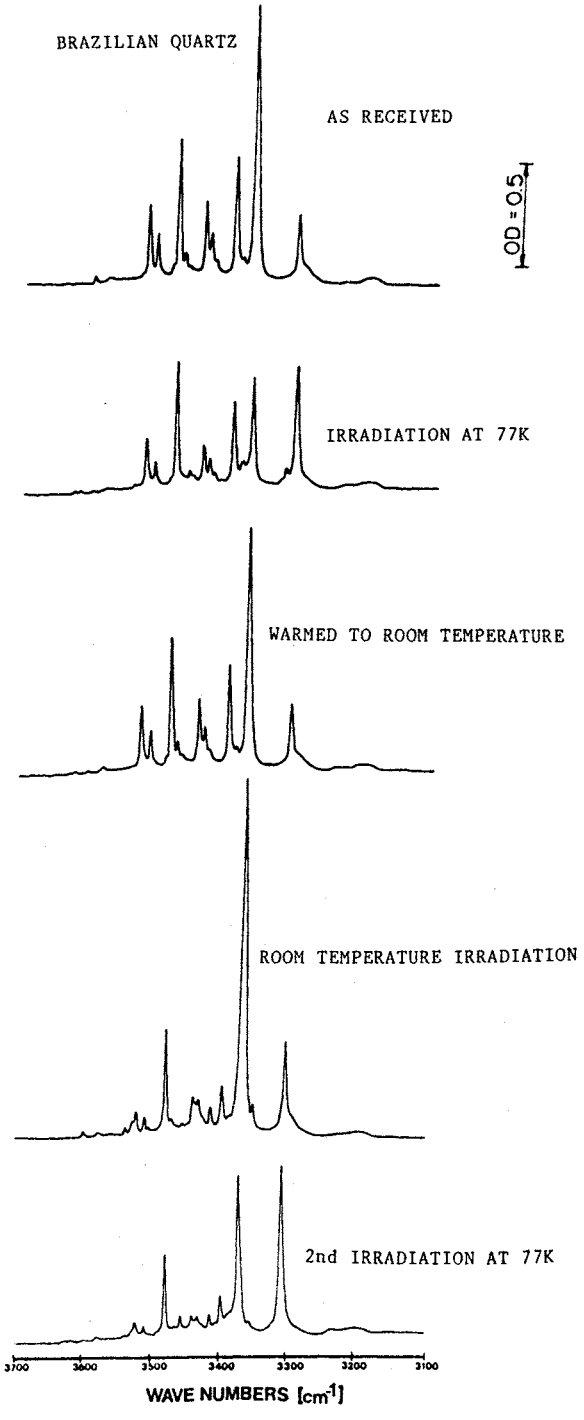


Figure 4: Infrared spectra showing the effect of irradiation sequence on unswept Brazilian quartz.

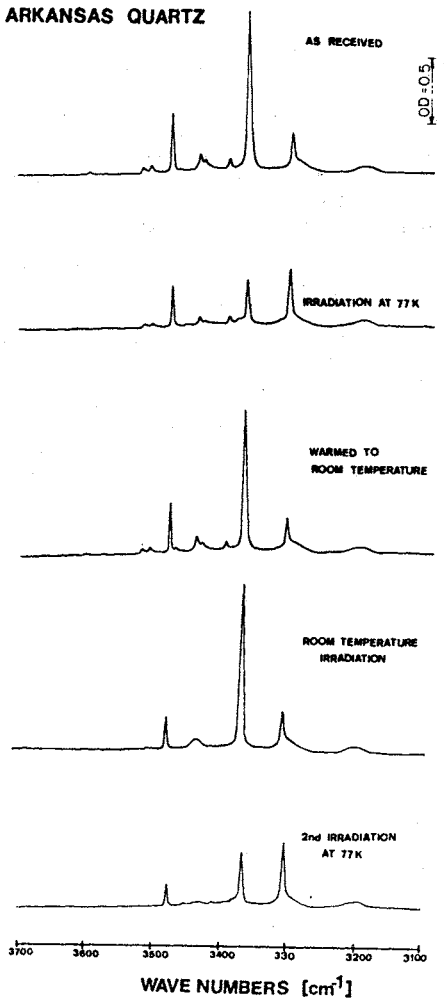


Figure 5: Infrared absorption spectra showing the effect of irradiation sequence on the unswept Arkansas quartz crystal.

reduced to one-third of its as-received strength while it was removed or lost in the background in Arkansas sample. In the Brazilian sample, two new small bands at 3537 and 3405 wavenumbers appeared and a large reduction occured in the bands at 3430 cm^{-1}. The 3507 cm^{-1} band was reduced to one-third of its as-received strength in the Brazilian sample while in the Arkansas sample it decayed much faster to about two-third of its as-received strength. The small 3395 cm^{-1} which showed a modest decrease at 77K was reduced fast at 300K irradiation; in the Arkansas sample it was lost in the background. A band of small strength at 3405 cm^{-1} was introduced in both the samples upon irradiation at 300K.

Among the major bands, on 300K irradiation, the N_1-Li^+ band was reduced by about 25% of its as-received strength in the Brazilian sample. while the decay of this band in the Arkansas sample was faster for identical radiation dose. The Arkansas quartz (Figure 5) exhibited the decay of N_1-Li^+ centers by about 45% of their as-received strength. The integrated growth of Al-OH^- centers was 1.2 times in the Brazilian sample while it was 1.07 times in the Arkansas quartz for the same radiation dose of 2 Mrad. For this high a radiation dose it is expected that the aluminum related defects will get saturated. This would therefore mean that used in the present investigation the Brazilian quartz has somewhat more aluminum than the Arkansas quartz.

Final irradiation at 77K showed that in both the samples the small 3595 cm^{-1} band, which got removed at the first 77K irradiation and recovered upon warming to 300K, behaved essentially the same as on first 77K irradiation. While all other bands above N_1-Li^+ were reduced significantly in Brazilian quartz, they disappeared in the Arkansas quartz. A new band in small strength appeared at 3452 cm^{-1} in both samples. All other bands except the e_1 at 3306 wavenumbers showed a reduced strength.

Among the major bands, at the final 77K irradiation, the strength of N_1-Li^+ centers in the Arkansas quartz was reduced much more than in the Brazilian sample. While the post-irradiated strength of the N_1-Li^+ centers in the Arkansas quartz was about one-third of their as-received strength, that in the Braz-

ilian quartz remained to about two-third of the original value. In the Brazilian quartz, the e_2 band was knocked down to about 50% of its strength attained after 300K irradiation and the e_1 band band grew to almost double of its strength. In the Arkansas quartz, e_2 band was knocked down to about one-third of its as-300K irradiated value and the e_1 bnad grew to more than double. Finally, the relative strength of the two bands in both the cases was nearly equal. This suggests that both the long and short bond oxygens around aluminum are approximately equally populated at this stage.

B. Comparative Studies of Irradiation Characteristics of Na-swpept Brazilian and Arkansas Quartz:

Figures 6 and 7 respectively depict the irradiation effect on the infrared spectra of Na-swept Brazilian and Arkansas quartz samples of matching thickness. The change in the spectral features as a result of electrodiffusion have been described previously (14). In the following we describe the results when these samples were irradiated at different temperatures.

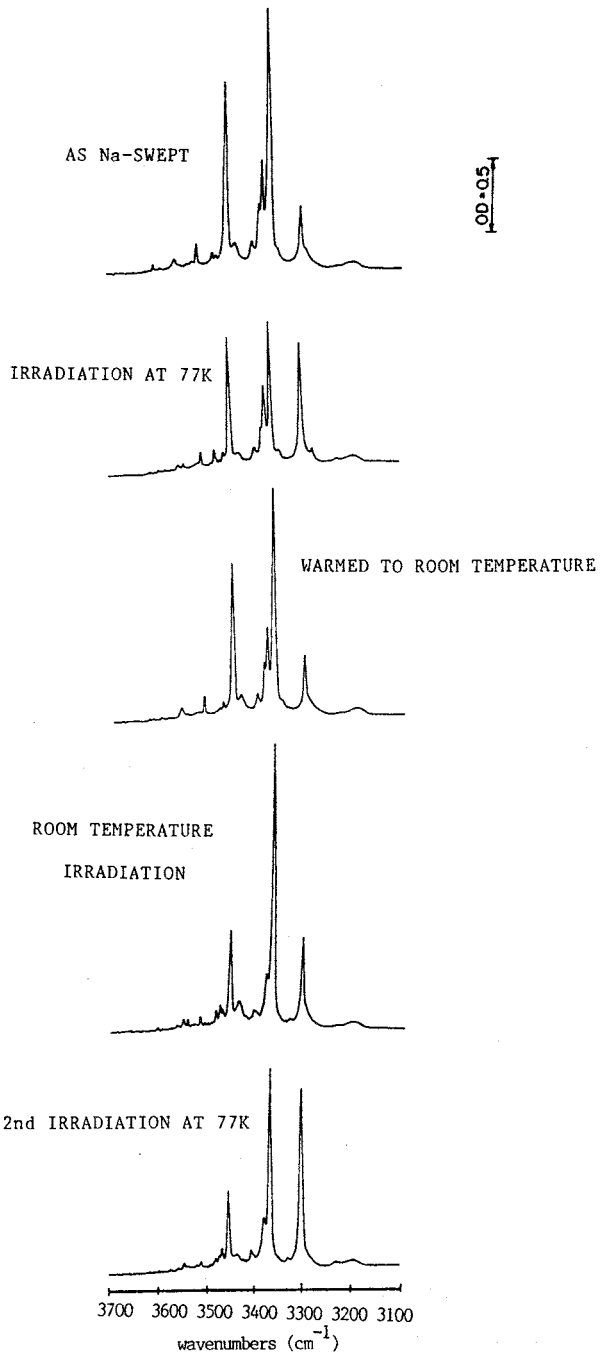


Figure 6: Infrared absorption spectra showing the effect of irradiation sequence on the Na-swept Brazilian quartz.

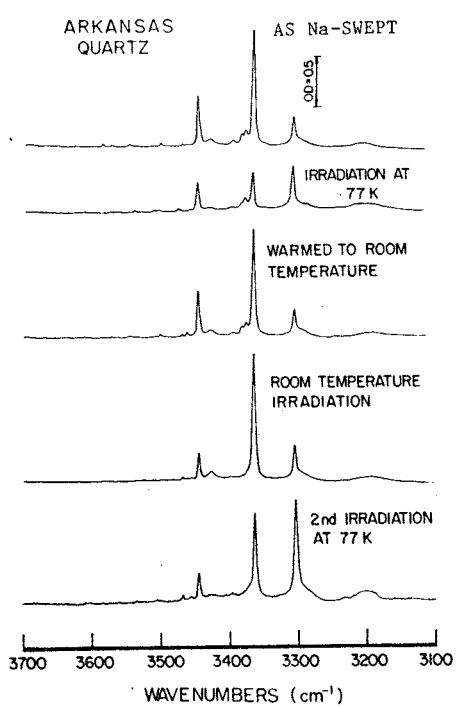


Figure 7: Infrared absorption spectra showing the effect of irradiation sequence on the Na-swept Arkansa quartz.

At the first 77K irradiation the small 3595 cm^{-1} band was decreased by about 60% in the Brazilian crystal and disappeared in the Arkansas quartz. It may be compared that in case of unswept condition this band was removed or lost in the background in both the samples. Some new low strength bands on the higher frequency side of N_1-Na^+ band such as 3512, 3545 cm^{-1} appear at the expense of other bands in this region.

Among the major bands the strength of 3451 cm^{-1} band representing the N_1 -Na^+ centers in both the Brazilian and Arkansas quartz depleted in almost the same proportion; about one-third of its as-swept strength. It may be mentioned here that while the N_1-Li^+ band in the unswept Arkansas quartz depleted about the same at the first 77K irradiation, that for the unswept Brazilian sample exhibited a much lower decrease of the order of 10-15%.The Al-OH^- bands in the Na-swept Arkansas quartz for the same radiation dose showed a similar character as exhibited in its unswept state. The e_2 band was reduced to about 25% and e_1 increased to about 1.5 times. However, the results for the Brazilian sample showed that while the e_1 band grew in nearly the same proportion as in the unswept condition, the reduction in the e_2 band was only to about 50%. It may however be seen from Figure 8 that the intensity of e_2 band in the Na-swept Brazilian sample is higher than that from the unswept sample of this crystal. This is perhaps due to the fact that the other bands near 3400 wavenumbers straddle the 3367 cm^{-1} band. Apparently, the Al-OH^- band strength should not be affected by alkali sweeping. This fact is observed in case of Arkansas quartz (Figure 9) where the strength of Al-OH^- bands is not affected by alkali sweeping.

Upon warming to room temperature the spectra in both the cases returned by and large to the "as-swept" condition with about 80-90% of strength of as-swept bands. Again,since the spectra restore to the original it shows that the low temperature irradiation-induced changs were not connected with the high range mobility of protons.

At room temperature irradiation it was interesting to note that in case of major bands, the N_1-Na^+ band at 3451 cm^{-1} was reduced to nearly 50%. The magnitude of this reduction was of similar extent in both Brazilian as well as Arkansas quartz. It may be

noted that the room temperature irradiation did not reduce to this extent the strength of the N_1-Li^+ centers in the unswept quartz of both the origins.

The integrated increase in the Al-OH^- strength after room temperature irradiation in Brazilian sample was 1.18 times of their as-Na-swept strength while in the Arkansas sample it was 1.12 times of the Na-swept strength. At this stage it may be compared that in the unswept version of the Brazilian crystal the Al-OH^- bands grew upon room temperature irradiation by a factor of 1.13 while the unswept Arkansas quartz showed almost similar increase.

In the case of Brazilian quartz the small 3595 cm^{-1} band started reappearing after room temperature irradiation. Another small 3507 cm^{-1} band showed a reduction to about 50% at this stage. A change can be noted in the significant increase of the band strength at 3470 cm^{-1} and the left shoulder of 3451 cm^{-1} band in the case of Brazilian sample only, while these bands show a modest reduction in Arkansas quartz. Among the other smaller bands a modest increase of the order of 15% may be observed for 3430 cm^{-1} band in both samples.

Upon final irradiation at 77K while the N_1-Na^+ band in the Brazilian sample was reduced to about 33% of its as-swept strength, in the Arkansas quartz it was reduced to the order of 25% of its as-swept strength. The e_2 band was about 1.2 times larger than e_1 in the Brazilian sample while in the Arkansas quartz both the bands attained almost similar strength. The small 3470 cm^{-1} band in Arkansas sample was unaffected while it was reduced in the Brazilian sample to the extent that it was lost in the background.

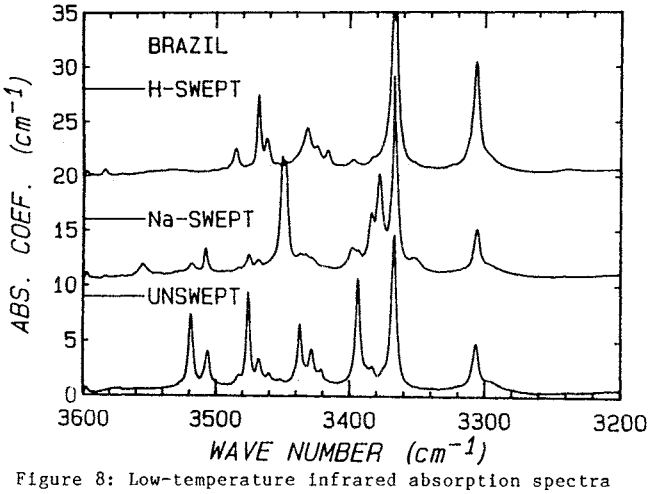


Figure 8: Low-temperature infrared absorption spectra of a Brazilian natural quartz crystal (Z plate) in its unswept, Na-, and H-swept conditions. The spectrum of Li-swept was exactly similar to the unswept condition and is not shown for the sake of clarity.

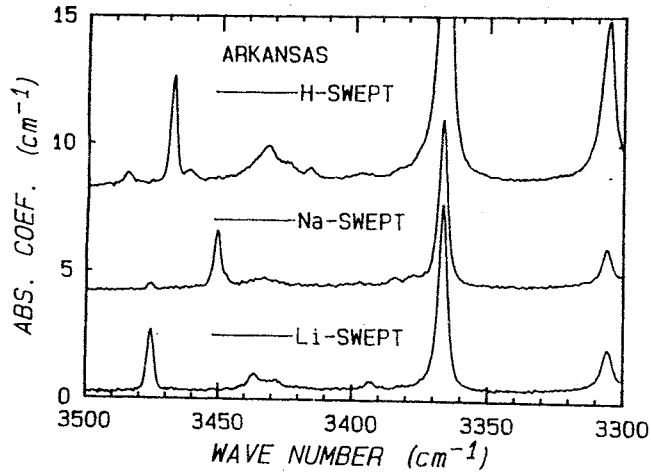


Figure 9: Low-temperature IR spectra of an Arkansas natural quartz (Z plate) in its Li-swept, Na-swept and H-swept conditions. The spectrum of unswept sample was exactly similar to its Li-swept condition and is not shown for the sake of clarity.

C. Comparative Studies of Irradiation Characteristics of H-Swept Brazilian and Arkansas Quartz:

Figures 10 and 11 respectively depict the irradiation effects on the infrared spectra of H-swept Brazilian and Arkansas quartz samples. The changes in the spectral characteristics as a result of electrodiffusion have been described previously (14). In the following we describe the results when these samples were irradiated at different temperatures.

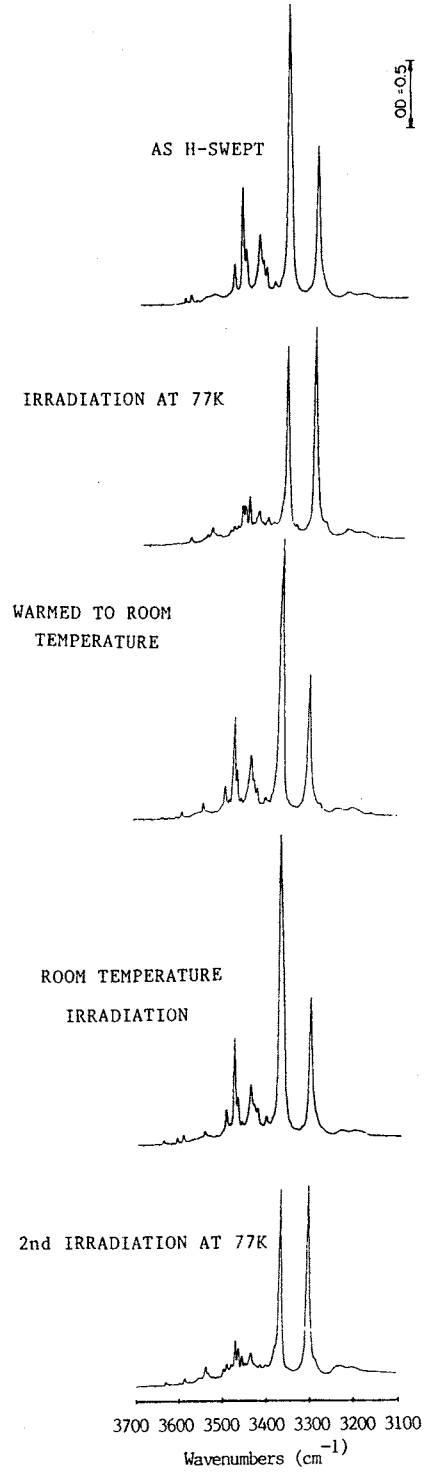


Figure 10: Infrared absorption spectra showing the effect of irradiation sequence on the H-swept Brazilian quartz crystal.

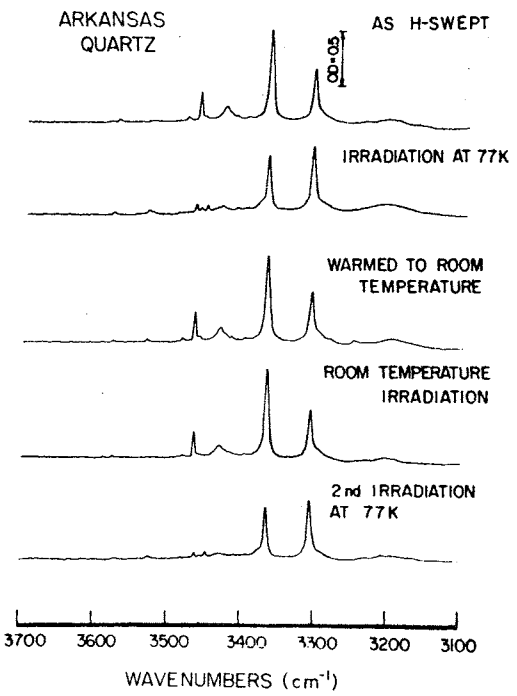


Figure 11: Infrared absorption spectra showing the effect of irradiation sequence on the H-swept Arkansas natural quartz.

At the first 77K irradiation, as in the other versions, the small 3595 cm^{-1} band was removed in both samples. Another small 3583 cm^{-1} band in both the cases remained unaffected. While in the Brazilian sample the 3535 cm^{-1} band grew to about three times upon 77K irradiation, it started showing up in a small strength in the Arkansas sample. It may be mentioned here that in the H-swept Arkansas sample this band is not observable in the as-swept condition.

A new band in significant strength at 3452 cm^{-1} appeared in both the H-swept samples at this irradiation. The intensity of this band in the Brazilian sample was nearly four times higher of that in the Arkansas sample. The band at 3430 cm^{-1} in both the samples decreased in same proportion; nearly 25% of their as-swept strength. Interestingly, while the 3415 cm^{-1} band in the Brazilian sample decreased significantly, it remained unaffected in the Arkansas quartz. The 3395 cm^{-1} band in both the samples got lost in the background upon the low temperature irradiation.

Among the major bands the strength of the N_1-H^+ centers at 3468 cm^{-1} was reduced to about 10% in the Brazilian quartz and in the Arkansas quartz it was down to about 25%. Interestingly, while the shoulder of the 3468 cm^{-1} band in the Brazilian quartz decreased, the corresponding shoulder in the Arkansas quartz increased to double. The behavior of Al-OH^- bands was in accordance with the observations of Sibley et.al (25) on cultured quartz and our results reported above on the other swept versions of the natural quartz. In the Brazilian quartz the e_2 (3367 cm^{-1})band showed a very significant reduction at low temperature irradiation. The band was down to about 65% of its as-swept strength. In the Arkansas quartz this band showed almost a similar reduction; about 60% of the H-swept strength. The decrease in the e_1 (3306 cm^{-1}) band intensity in the crystals of both the origins was also very different. In the Brazilian sample the e_1 band increased on low temperature irradiation by a factor of 1.44 and in the Arkansas sample the corresponding increase was by a factor of about 1.35 times.

Despite the fact that the relative magnitude of change of e-bands in Brazilian and Arkansas quartz was different, it is interesting to note that in both cases the final intensity of these bands is almost relatively in the same proportions for both the samples. This shows that both long and short bond oxygens around the aluminum are approximately equally populated with protons at this stage.

Upon warming the 77K irradiated H-swept natural quartz to room temperature the 3535 cm^{-1} band was retained and became more well defined in both the samples. It may be observed that in the Brazilian quartz this band already existed prior to low temperature irradiation and showed a significant increase upon irradiation while in the Arkansas quartz it only came up after low temperature irradiation. The N_1-H^+ center band in both samples of quartz attained a value about 10% higher than the as-swept strength. All other bands showed a recovery of the order of 85-95%.

Upon room temperature irradiation a new small band appeared at 3630 cm^{-1} in the Brazilian quartz. There was however no appearance of this band in the Arkansas quartz. The 3430 cm^{-1} small band in both types of samples was only slightly reduced. The 3395 cm^{-1} band showed a recovery. While the recovery in the Arkansas sample was only marginal, it was nearly 1.5 times of as-swept strength in the Brazilian quartz.

The major bands viz. e_1 and e_2 and N_1 -H^+ were almost unaffected at this irradiation. It is known that for completely H-swept quartz crystals all the aluminum ions are charge compensated by protons. The irradiation process then carried out does not form stable Al-hole centers (4,5,7). The formation of stable Al-hole centers in such cases is prevented because the protons released from Al-OH^- centers (e_1 and e_2 bands) do not find trapping sites and therefore quickly return to the aluminum sites. It seems that almost a similar situation seems to be taking place with regard torelease of protons from N_1-H^+ centers at room temperature irradiation.

At the final 77K irradiation, the spectrum was almost the same as on the first 77K irradiation except that the band intensities showed a very nominal reduction. In the Brazilian sample, a band is introduced at 3630 cm^{-1} at room temperature irradiation that continued to stay on the subsequent low temperature irradiation.

DISCUSSION:

This paper has described the investigations on OH^- infrared bands in different versions of natural quartz viz. unswept, Li-swept, Na-swept and H-swept. The description includes previously unreported results of irradiation effects at low and room temperatures on the major bands as well as some smaller but prominent bands. The major bands include N_1-Li^+ in unswept (or Li-swept) condition, N_1-Na^+ in Na-swept and N_1 -H^+ in H-swept conditions and the Al-OH^- (e_1 and e_2) bands. While, the e_1 and e_2 bands have been undisputed assigned in literature (5,10-12,25,43,44) to represent Al-OH^-centers with hydrogen trapped at short and long bond respectively, there has not been any specific model assigned so far for the remaining OH^- bands. The N_1-Li^+, N_1-Na^+, and N_1-H^+ bands have been only characterized for their low temperature irradiation and anneal characteristics (14). It has been shown that N_1-M^+ as well as N_1-H^+ centers following a low temperature irradiation recover completely during annealing by about 760K. These bands seem to represent a specific configuration of LiOH, NaOH and H_2O molecules perturbed by SiO_4 tetrahedra in natural quartz lattice. It has been established by now that OH^- molecules in natural and cultured crystals vary in their specific configuration (14-16).

A number of interesting features have emerged from the present work. Let us first consider the irradiation behavior of the major bands present in the natural crystals.

At the outset it may be mentioned that the Al-OH^- centers in natural crystals show similar irradiation characteristics as those observed in cultured crystals (5,25). While these bands are present in the as-received unswept natural quartz (14), they are produced in cultured crystals only after irradiation at temperatures

beyond 200K or upon sweeping in hydrogen atmosphere (5, 14,25,41,43). The process of irradiation or H-sweeping in natural quartz crystals increases the Al-OH^- band strength in the limit to saturate all the aluminum with hydrogen (14). Low temperature irradiation like what happens in cultured quartz equalizes the strength of the e-bands indicating that both short and long bonds in the AlO_4 tetrahedra become equally populated with protons. Our isochronal anneal data (14) following a strong low temperature irradiation on these crystals also showed that like in the case of cultured quartz (25) these bands grow in their relative strength until 230K. Above 230K, they quickly recover to their original strength. Sibley et.al (25) interpreted that these bands arise from OH^- molecules in two different but closely related sites relative to aluminum ions and that a transfer from one site to the other takes place beyond 230K such that long bonds in the AlO_4 tetrahedra become more populated than the short bonds.

The present studies on irradiation characteristics of N_1-Li^+, N_1-Na^+, and N_1-H^+ have shown that irradiation effects on these bands depend upon the presence of alkali or hydrogen present at this defect center. The N_1-M^+ centers (where M^+ is an alkali ion either Li^+ or Na^+) show a relatively large decay at room temperature irradiation than at low temperature irradiation. It has been found (5,43) that at temperatures 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). In the case of N_1-M^+ centers the present results suggest that the process of radiation induced dissociation of alkali ions from N_1-M^+ centers is similar to the temperature dependence of radiation induced mobility of alkali and hydrogen ions from aluminum.

An important point may be mentioned here for unswept (or Li-swept) and Na-swept crystals.The magnitude of the depletion of N_1-Li^+ centers under irradiation in Brazilian and Arkansas quartz was different. But in the Na-swept version of crystals of both origins the depletion in the N_1-Na^+ centers at the first low temperature irradiation and also at the room temperature irradiation was in the same proportion. At the stage of room temperature irradiation both the crystals exhibited a reduction of about 50% in the strength of the N_1-Na^+ centers. In the unswept or the Li-swept version of the Brazilian crystal the magnitude of depletion at this stage of accumulated radiation dose was much low (~25-30%) and in the Arkansas quartz it was comparable (45%) with the Na-swept version. It may be seen from the above that there are some apparent complexities involved in the case of Brazilian crystal used.

We shall now turn our discussion on the radiation effects on the major bands in H-swept crystals. In sharp contrast to the N_1-M^+, the N_1-H^+ 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 the N_1-M^+ centers in alkali swept versions. The Al-OH^- centers also behaved on the same pattern. Interestingly, all the major bands in H-swept natural quartz following a low temperature irradiation show almost complete recovery upon warming to room temperature. This suggests that like cultured quartz (25) the mobility of hydrogen is not of long range order at low temperature irradiation. It is also known (4,5,7) that in the case of H-swept cultured crystals the room temperature irradiation does not form stable Al-hole centers. Instead, the hydrogen which is knocked from aluminum under the influence of radiation diffuses back to the aluminum. In the process of back diffusion the transient frequency offsets are observed (28-36). It seems that in the H-swept natural crystals an almost similar situation exists with regard to radiation induced removal of protons from N_1-H^+ centers at room temperature irradiation which does not allow the formation of N_1-e^+ centers from their precursors. Thus, like Al-OH^- centers the N_1-H^+ cenetrs also hold hydrogen firmly so that no post-room temperature irradiation steady state effects are observable.

A new band at 3452 cm^{-1} in significant strength was observed in both H-swept natural quartz samples at low temperature irradiation and a small band at 3630 cm^{-1} on room temperature irradiation in Brazilian sample only.

Experimental results on radiation effects on smaller bands in natural quartz are intricate. In some cases there are new bands observed either at the expense of other bands or as a result of electrodiffusion. An interesting observation may be mentioned here. A very small band at 3595 cm^{-1} is found to exist in all the versions of unswept and swept natural quartz. This band removed upon irradiation at 77K and recovers upon warming to room temperature. Room temperature irradiation does not affect the intensity of this band. Irradiation characteristics of this band as a function of temperature are similar to N_1-H^+ defect band. At this stage it is extremely difficult to identify all these bands and assign models to them.

SUMMARY:

In summary we have found the following:

(1) It has been observed that the natural quartz drawn from Arkansas and Brazilian origins which are generally used as starting material in hydrothermal synthesis of high purity cultured quartz have essentially the same spectral features in their as-received state. Some of the low strength infrared absorption bands present in the Brazilian sample are not observable in the Arkansas stone. Thus, the particular Arkansas quartz used in the present work was a purer material in relation to the Brazilian quartz employed in the investigation.

(2) Irradiation at low temperatures (~77K) suppresses 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 suppressed bands. This suggests that mobility of hydrogen released from these low strength bands under the influence of irradiation at low temperature is of short range order. On the other hand major bands show a complete recovery by 760K (14) showing a long range order mobility of protons from these defects.

(3) The effect of irradiation at 77K as well as at 300K on the Al-OH^- centers in natural quartz have similar characteristics as are shown up by all grades of cultured quartz.

(4) For the same radiation dose (2 Mrad used at every stage the N_1-M^+ centers showed a relatively large decay upon irradiation at 300K than at 77K. On the other hand the N_1-H^+ cenetrs do not show any depletion at 300K while at 77K irradiation their strength is significantly reduced to about 10% of their original value.

(5) A small band at 3595 cm^{-1} is observed in all unswept and swept versions of natural quartz.Thus this band is independent of any sweeping. This band is removed by irradiation at 77K and recovered upon warming to 300K. Also, irradiation at 300K did not affect this band. The irradiation characteristics of this band are thus identical to N_1-H^+ centers.

(6) During irradiation sequence on both the samples it was observed that at the stage of 300K irradiation the extent of depletion in the N_1-Na^+ centers was larger than the N_1-Li^+ centers. The N_1-Na^+ centers depleted to about half of their as Na-swept strength while the reduction in the strength of N_1-Li^+ centers was much less. Future work in this direction would lead to a more exact understanding of the role of the alkali impurity at the N_1-M^+ defect centers in the natural quartz.

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

ACKNOWLEDGMENTS:

The author expresses his gratitude and appreciation to Professor Joel Martin of Oklahoma State University, Stillwater, OK for supporting this work. He is also grateful and appreciative to Dr. J.C. King of Sandia

National Laboratories, Albuquerque, NM for his encouragement to investigate natural crystals also. The Arkansas quartz of the present was given by Professor Martin and Brazilian 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, 1985 Vol.1 pp. 147-159.

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, 1985, Vol.1 pp. 1-46.

3. J.J. Martin, Proc. 43rd Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ (1989). Copies available from IEEE-445 Hoes Lane, Piscataway, NJ 08854.

4. L.E. Halliburton, Proc. Int. Conf. Insulators (1984), University of Utah, Salt Lake City, Utah.

4. L.E. Halliburton, N. Koumvakalis, M.E. Markes and J.J. Martin, J. Appl. Phys. 52, 3565 (1981).

6. J.A. Weil, Radiation Effects 26, 261 (1975).

7. L.E. Halliburton,M.E. Markes and J.J. Martin, Proc. 34th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ 1, (1980). Copies available from Electronic Industries Association, 2001 Eye Street, Washington, DC 20006.

8. J.A. Weil, Phys. Chem. Minerals 10, 149 (1984).

9. S.W.S. McKeever, Radiation Protection Dosimetry, 8, 81 (1984).

10. L.E. Halliburton, J.J. Martin and W.A. Sibley, 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, Oklahoma State University, RADC-TR-83-309, Jan. 1984. Copies available from Rome Air Development Command, Hanscom, AFB, MA 01731.

12. J.J. Martin, J. Appl. Phys. 56, 2536 (1984).

13. F. Euler and A. Kahan, Phys. Rev. B 35, 4351 (1987).

14. Harish Bahadur, J. Appl. Phys. 66, 4973 (1989).

15. R.N. Brown and A. Kahan, J. Phys. Chem. Solids 36, 467 (1975).

16. A. Kats, Philips Res. Rep. 17, 133 (1962).

17. H.G. Lipson, F. Euler, A.F. Armington, Proc. 32nd Ann. Symp. Frequency Control, U.S. Army Electronics Command, Fort Monmouth, NJ 11 (1978). Copies available from the Electronic Industries Association, 2001 Eye Street , N.W., Washington, DC 20006.

18. F. Euler, H.G. Lipson and P.A. Ligor, Proc. 34th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ, 72 (1980). Copies available from Electronics Industries Association, 2001 Eye Street, N.W., Washington,DC 20006.

19. H.G. Lipson, A. Kahan, R.N. Brown and F.K. Euler, Proc. 35th Ann. Symp. Frequency Control, U.S.Army Electronics Research and Development Command, Fort Monmouth, NJ 329 (1981). Copies available from Electronic Industries Association, 2001 Eye Street, N.W., Washington, DC 20006.

20. F. Euler, H.G. Lipson, A. Kahan and A.F. Armington, Proc. 36th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ, 115 (1982). Copies available from Electronic Industries Association, 2001 Eye Street, N.W., Washington, DC 20006.

21. H.G. Lipson, A. Kahan, and J.O'Connor, Proc. 37th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ, 169 (1983). Copies available from IEEE-445 Hoes Lane, Piscataway, NJ 08854 as document no. 83CH1957-0.

22. H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci. NS-31, 1223 (Dec. 1984).

23. H.G. Lipson and A. Kahan, J. Appl. Phys. 58, 963 (1985).

24. J.J. Martin, IEEE TRans. Ultrason, Ferroelec. Frequency Control 35, 288 (1988).

25. W.A. Sibley, J.J. Martin, M.C. Wintersgill and J. D. Brown, J. Appl. Phys. 50, 5449 (1979).

26. J.J. Martin, Ho B. Hwang, H. Bahadur and G.A. Berman, J. Appl. Phys.65,4666 (1989).

27. J.C. King, Bell Syst. Tech. J. 38, 573 (1959).

28. B.R. Capone, A. Kahan, R.N. Brown and J.R. Buckmelter, IEEE Trans. Nucl. Sci., Vol. NS-17, 217 (1970).

29. R.A. Poll and S.L. Ridgway, IEEE Trans. Nucl. Sci., vol. NS-13, 130 (1965).

30.T.J. Young, D.R. Koehler, R.A Adams, Proc. 32nd Ann. Symp. Frequency Control,U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ 34 (1978). Copies available from Electronic Industries Association, 2001 Eye Street, N.W., Washington,DC 20006.

31. A. Kahan, F. Euler, H.G. Lipson, C.Chen and L.E. Halliburton, Proc. 41st Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ. 216 (1987).

32. P. Pelligrini, F. Euler, A. Kahan, T.M. Flanagan and T.F. Wrobel, IEEE Trans. Nucl. Sci., NS-25, 1267 (1976).

33. J.C. King and H.H. Sander, IEEE Trans. Nucl. Sci., NS-19, 23 (1972).

34. J.C. King and H.H. Sander, Radiation Effects. 26, 203 (1975).

35. J.C. KIng and H.H. Sander, IEEE Trans. Nucl. Sci., NS-20, 117 (1973).

36. D.R. Koehler and J.J. Martin, J. Appl. Phys. 57, 5205 (1985).

37. J.R. Norton, J.M. Cloeren and J.J. Suter, Proc. 38th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ,., 63 (1984). Copies available from IEEE-445 Hoes Lane, Piscataway, NJ 08854.

38. J. Suter and R. Maurer, Proc. 40th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ 134 (1986).

39. J.J. Suter and R.H. Maurer, IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, UFFC-34, 667 (1987).

40. T. Flanagan, R. Leadon, and D. Shannon, Proc. 40th Ann. Symp. Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ., 127 (1986). Copies available from IEEE-445 Hoes Lane, Piscataway, NJ 08854.

41. M.E. Markes and L.E. Halliburton, J. Appl. Phys. 50, 8172 (1979).

42. R.A. Weeks and M. Abraham, J. Chem. Phys. 42, 68 (1965).

43. L.E. Halliburton, M. Markes, J.J. Martin, S.P. Doherty, N. Koumvakalis, W.A. Sibley A.F. Armington and R.N. Brown, IEEE Trans. Nucl. Sci., NS-26, 4851 (1979)

44. B. Subramaniam, L.E. Halliburton and J.J. Martin, J. Phys. Chem. Solids, 45, 575 (1984).

SESSION 5A

RESONATORS II

Chairman : J. DETAINT

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 5A Number 27

ON THE INFLUENCE OF A FABRICATION IMPERFECTION ON THE IN-PLANE ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS WITH RECTANGULAR SUPPORTS*

H.F. Tiersten and Y.S. Zhou

Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Abstract

An analysis of the influence of a relative displacement of the center of the mode shape with respect to rectangular supports on the in-plane acceleration sensitivity of contoured quartz resonators is performed. In earlier work of a similar nature on the normal acceleration sensitivity an orientation of the support rectangle of an SC cut, for which this effect exhibits a strong minimum, was found. The calculated influence of the offset of centers in the in-plane case is shown to be an order of magnitude smaller than in the normal case. Consequently, the orientation with the strong minimum in the normal case is practical.

1. Introduction

It has been shown[1-3] that both the normal and resultant in-plane acceleration sensitivities of contoured quartz resonators with rectangular supports vanishes when the centers of the mode shape and support rectangle coincide. This was shown analytically and is a consequence of the symmetry of the biasing state and the mode shape. The result holds for any symmetric shape for which the support configuration exhibits the required symmetry with respect to the center of the mode shape. Since it is extremely difficult to realize this situation in practice, an analysis of the influence of an offset of the centers on the normal acceleration sensitivity was performed in earlier work[1,2]. It was shown that for small offsets the normal acceleration sensitivity increases linearly with offset and for a particular aspect ratio of an SC cut an orientation of the support configuration for which this effect exhibits a strong minimum was found.

In view of the foregoing in this work an analysis of the influence of an offset of the centers on the in-plane acceleration sensitivity is performed. The extensional biasing deformation is determined by means of a variational approximation procedure using the variational principle in which all conditions appear as natural conditions. The resulting extensional biasing states are employed in an existing perturbation equation[4] along with the mode shapes of the contoured resonators[5] to calculate the in-plane acceleration sensitivities. It is shown that for small offsets of centers, which is the practical case, each component of in-plane acceleration sensitivity increases linearly with offset, as was the case for normal acceleration. The calculations show that the linear increase in the in-plane acceleration sensitivity with offset tends to be about an order of magnitude smaller than in the case of the normal acceleration sensitivity. Consequently, the aforementioned aspect ratio and orientation of the support configuration for which the linear increase in the normal acceleration sensitivity with offset for an SC cut has a strong minimum is practical for minimizing the influence of the type of fabrication imperfection considered.

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

$$\hat{K}^n_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M}\,, \tag{2.3}$$

where

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM}\delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN} E^1_{KN} + c_{2L\gamma KM} w_{\alpha,K} + c_{2LKM\alpha} w_{\gamma,K}\,, \tag{2.4}$$

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

$$g^\mu_\gamma \equiv \frac{u^\mu_\gamma}{N_\mu}\,, \quad \hat{f}^\mu = \frac{\hat{\varphi}^\mu}{N_\mu}\,, \quad N^2_\mu = \int_V \rho u^\mu_\gamma u^\mu_\gamma\, dV\,, \tag{2.6}$$

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. Formulation of Biasing Problem

Since the plate is thin, for in-plane acceleration the equations of anisotropic plane stress apply. For anisotropic plane stress the stresses K_{2L} vanish and only K_{AB} exist, where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2. The equations for anisotropic plane stress may be written in the form

$$K_{AB,A} - \rho a_B = 0 , \tag{3.1}$$

where a_B denotes the acceleration in the plane of the plate and the anisotropic constitutive equations for plane stress may be written in the form

$$K_{AB} = \gamma_{ABCD} E_{CD}, \quad E_{CD} = \frac{1}{2} (w_{C,D} + w_{D,C}), \tag{3.2}$$

and γ_{ABCD} are Voigt's anisotropic plate elastic constants, which are given by

$$\gamma_{RS} = C_{RS} - C_{RW} C^{-1}_{WV} C_{VS}, \quad R,S = 1,3,5 ;$$
$$W,V = 2,4,6 . \tag{3.3}$$

From the relaxation of the stresses K_{2L}, we have

$$E_W = - C^{-1}_{WV} C_{VS} E_S , \tag{3.4}$$

which enables us to obtain all the strains E_{KL} when a solution is known as well as the rotation Ω_{13} from the relation

$$\Omega_{13} = \frac{1}{2} (w_{3,1} - w_{1,3}) . \tag{3.5}$$

However, the rotations Ω_{2A} cannot be obtained from a solution because of the relaxation of the K_{2L}. Nevertheless, the flexural type plate rotations Ω_{2A} are taken to vanish in the case of uncoupled anisotropic plane stress, which is the kinematic uncoupling assumption for anisotropic plane stress. Accordingly, we have

$$\Omega_{2A} = 0 , \tag{3.6}$$

and now with (3.2) - (3.4), (3.6) and

$$w_{K,L} = E_{KL} + \Omega_{LK} , \tag{3.7}$$

we have the three-dimensional displacement gradients for the extensional biasing deformation field, which is required for the evaluation of $\hat{c}_{L\gamma M\alpha}$ in (2.4).

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 systems. The origin of the unprimed coordinate system is located at the center of the supports and the origin of the primed coordinate system is located at the center of the mode shape, which we permit to be displaced with respect to each other by an arbitrary small distance $\tilde{d}$ in an arbitrary direction θ. Hence, $\tilde{d}$ has arbitrary components d_1 and d_3. The static solution for the biasing state is referred to the unprimed coordinate system only.

As a consequence of the geometry of the supporting sidewalls, we now take the displacement w_s in the tangential direction of the supporting sidewalls to vanish. On account of the stubbiness of the cross-section of a sidewall, we model its resistance to the normal displacement w_n of the center-plane of the plate as a simple variable shear strip, as in earlier work[6], which yields

$$K_{(n)(n)} = \mp k w_n , \tag{3.8}$$

where in this work we have[7]

$$k = 2\mu t_f / h^2 , \tag{3.9}$$

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, in accordance with the discussion concerning the tangential displacement along the sidewall we have

$$w_s = 0 . \tag{3.10}$$

at the sidewall.

4. Extensional Biasing Deformation

The extensional problem has been defined in Section 3 with the aid of Fig.1. Clearly the equations of anisotropic plane stress, i.e., Eqs.(3.1) and (3.2), apply. From Eqs.(3.8) and (3.10) we have the boundary conditions

$$w_3 = 0, \; K_{11} \pm k_1 w_1 = 0 \;\text{ at }\; X_1 = \pm a, \; |X_3| \le b,$$
$$w_1 = 0, \; K_{33} \pm k_3 w_3 = 0 \;\text{ at }\; X_3 = \pm b, \; |X_3| \le a . \tag{4.1}$$

We now transform the inhomogeneities from the differential equations, which consist of (3.1) with (3.2), into the boundary conditions (4.1) by writing

$$w_A = \hat{w}_A + \overset{*}{w}_A , \tag{4.2}$$

where

$$\overset{*}{w}_1 = A_1 (X_3^2 - b^2), \quad \overset{*}{w}_3 = A_3 (X_3^2 - b^2) - \rho a_3 b / k_3 . \tag{4.3}$$

The substitution of (4.2) with (4.3) into (3.1) with (3.2) yields

$$\gamma_{11}\hat{w}_{1,11} + 2\gamma_{15}\hat{w}_{1,31} + \gamma_{55}\hat{w}_{1,33} + \gamma_{35}\hat{w}_{3,33} + (\gamma_{13} + \gamma_{55})\hat{w}_{3,31} + \gamma_{15}\hat{w}_{3,11} = 0$$
$$\gamma_{15}\hat{w}_{1,11} + (\gamma_{13} + \gamma_{55})\hat{w}_{1,31} + \gamma_{35}\hat{w}_{1,33} + \gamma_{33}\hat{w}_{3,33} + 2\gamma_{35}\hat{w}_{3,31} + \gamma_{55}\hat{w}_{3,11} = 0, \tag{4.4}$$

since A_1 and A_3 are selected as

$$A_1 = \frac{\rho}{2} \frac{(\gamma_{33} a_1 - \gamma_{35} a_3)}{(\gamma_{33}\gamma_{55} - \gamma_{35}^2)} ,$$
$$A_3 = - \frac{\rho}{2} \frac{(\gamma_{35} a_1 - \gamma_{55} a_3)}{(\gamma_{33}\gamma_{55} - \gamma_{35}^2)} . \tag{4.5}$$

The further substitution of (4.2) with (4.3) and (3.2) into (4.1) yields the transformed edge conditions

$$\hat{w}_3 = -\overset{*}{w}_3,\ \hat{K}_{11} \pm k_1\hat{w}_1 = -\overset{*}{K}_{11} \mp k_1\overset{*}{w}_1 \quad \text{at } X_1 = \pm a,\ -b < X_3 < b,$$

$$\hat{w}_1 = 0,\ \hat{K}_{33} \pm k_3\hat{w}_3 = 0 \text{ at } X_3 = \pm b,\ -a < X_1 < a, \qquad (4.6)$$

where

$$\overset{*}{K}_{11} = 2(\gamma_{13}A_3 + \gamma_{15}A_1)X_3, \qquad (4.7)$$

and we note that

$$\overset{*}{K}_{33} \pm k_3\overset{*}{w}_3 = 0 \text{ at } X_3 = \pm b. \qquad (4.8)$$

As a solution of (4.4) we write

$$\hat{w}_A = \tilde{w}_A + \sum_m \beta_{Am} e^{i\eta_m X_3} e^{i\nu_m X_1}, \qquad (4.9)$$

where the $\tilde{w}_A$ are polynomials in X_1 and X_3, which satisfy (4.4) and certain boundary conditions and are discussed a little later on for convenience, and $\nu_m = m\pi/2\sigma a$ and σ is chosen to be an irrational number (here 2π) in order that neither $\cos \nu_m a$ or $\sin \nu_m a$ vanish. The substitution of (4.9) into (4.4) yields

$$[\gamma_{11}\nu_m^2 + 2\gamma_{15}\nu_m\eta_m + \gamma_{55}\eta_m^2]\ \beta_{1m} + [\gamma_{35}\eta_m^2 + (\gamma_{13}+\gamma_{55})\eta_m\nu_m + \gamma_{15}\nu_m^2]\beta_{3m} = 0$$

$$[\gamma_{15}\nu_m^2 + (\gamma_{13}+\gamma_{55})\eta_m\nu_m + \gamma_{35}\eta_m^2]\ \beta_{1m} + [\gamma_{33}\eta_m^2 + 2\gamma_{35}\eta_m\nu_m + \gamma_{55}\nu_m^2]\ \beta_{3m} = 0, \qquad (4.10)$$

for each m. Equations (4.10) constitute a system of two linear homogeneous algebraic equations in β_{1m} and β_{3m}, which yields nontrivial solutions when the determinant of the coefficients vanishes. The vanishing of the determinant yields four roots $\hat{\eta}_m^{(p)}$ $(p = 1,2,3,4)$ for each ν_m with four sets of amplitude ratios $\beta_{1m}^{(p)} : \beta_{3m}^{(p)}$. By scaling this is independent of m. Hence, as a solution of the boundary value problem we may write

$$\hat{w}_A = \tilde{w}_A + \sum_m \sum_{p=1}^{4} D_p^{(m)} \beta_A^{(p)} e^{i\eta_m^{(p)} X_3} e^{i\nu_m X_1}, \qquad (4.11)$$

where the $D_p^{(m)}$ are amplitude coefficients still to be determined. By satisfying $(4.6)_3$ at $X_3 = b$ and $(4.6)_4$ at $X_3 = \pm b$ for each m, we can reduce the number of unknowns from 4 to 1 for each m. Accordingly, we obtain

$$\sum_{p=1}^{4} \beta_1^{(p)} D_p^{(m)} e^{i\eta_m^{(p)} b} = 0,$$

$$\sum_{p=1}^{4} (i\nu_m k_3^{(p)} + k_3\beta_3^{(p)}) D_p^{(m)} e^{i\eta_m^{(p)} b} = 0,$$

$$\sum_{p=1}^{4} (i\nu_m k_3^{(p)} - k_3\beta_3^{(p)}) D_p^{(m)} e^{-i\eta_m^{(p)} b} = 0, \qquad (4.12)$$

where

$$k_3^{(p)} = \gamma_{13}\beta_1^{(p)} + \gamma_{33}\beta_3 h_p + \gamma_{35}(\beta_1^{(p)} h_p + \beta_3^{(p)}), \qquad (4.13)$$

and $h_{(p)} = \eta_m^{(p)}/\nu_m$ is independent of m.

Equations (4.12) constitute three equations in four unknowns, the satisfaction of which enables us to write

$$D_p^{(m)} = g_p^{(m)} D^{(m)}, \qquad (4.14)$$

which reduces the number of unknown coefficients from 4 to 1 for each m. The polynomials $\tilde{w}_A$ are taken in the form[8]

$$\tilde{w}_1 = D_o(X_3 - b) - \bar{D}_o(X_3 - b)[C_o C_1 X_1 + C_2(X_3 + b)],$$

$$\tilde{w}_3 = -\frac{\gamma_{35}}{\gamma_{33}+k_3 b} D_o X_3 + \bar{D}_o\Big[C_o\Big(C_3 + X_3\Big(\frac{\gamma_{35}}{\gamma_{55}} X_1 - \frac{\gamma_{13}}{\gamma_{55}} b\Big)\Big) + C_4(X_3^2 - b^2)\Big], \qquad (4.15)$$

which satisfy the same three homogeneous edge conditions in (4.6) as well as (4.4). The C_r in (4.15) are given by

$$C_o = 2(\gamma_{55}\gamma_{33} - \gamma_{35}^2)/\gamma_{55}^2,\ C_1 = (\gamma_{33} + k_3 b)/\gamma_{55}$$

$$C_2 = \gamma_{35}(\gamma_{13}\gamma_{33} + \gamma_{33}\gamma_{55} - 2\gamma_{35}^2)/\gamma_{55}^3 + C_1(\gamma_{13}\gamma_{35} + \gamma_{35}\gamma_{55} - 2\gamma_{15}\gamma_{33})/\gamma_{55}^2,$$

$$C_3 = -b^2\Big[1 + \frac{\gamma_{55} C_o}{2k_3 b}\Big]$$

$$C_4 = \gamma_{35}^2(\gamma_{13} - \gamma_{35})/\gamma_{55}^3 + C_1(\gamma_{13}\gamma_{55} + \gamma_{55}^2 - 2\gamma_{15}\gamma_{35})/\gamma_{55}^2, \qquad (4.16)$$

and D_o and $\bar{D}_o$ are amplitude coefficients still to be determined. The substitution of (4.14) into (4.11) enables us to write

$$\hat{w}_A = \tilde{w}_A + \sum_m \Big[\sum_{p=1}^{4} \beta_A^{(p)} g_p^{(m)} e^{i\eta_m^{(p)} X_3}\Big] D^{(m)} e^{i\nu_m X_1}, \qquad (4.17)$$

as the approximate solution.

The solution functions in (4.17) satisfy the differential equations (4.4) and the above-mentioned three transformed homogeneous edge conditions in (4.6), but they do not satisfy the remaining boundary conditions. We satisfy these latter conditions by means of a variational approximation procedure. Since the solution functions satisfy the differential equations (3.1) with (3.2) exactly and one edge condition is of constraint type and the other is of natural type, all that remains of the variational principle in which all conditions are unconstrained[9] is

$$\int_{S_C} w_{(s)} \delta K_{(n)(s)} dS - \int_{S_N} (k_{(n)} w_{(n)} \pm \pm K_{(n)(n)})\ \delta w_{(n)}\, dS = 0. \qquad (4.18)$$

Substituting from (4.2) with (4.3), (4.17) and (4.15) and employing the homogeneous equations in (4.6), which have been satisfied exactly by each term in the solution sum, into (4.18), we obtain

$$-\int_{-a}^{a} \hat{w}_1 \delta\hat{K}_{13}\Big|_{X_3=-b} dX_1 + \int_{-b}^{b} (\hat{w}_3 + \overset{*}{w}_3)\delta\hat{K}_{13}\Big|_{X_1=-a}^{X_1=a} dX_3 - \int_{-b}^{b} (k_1\hat{w}_1 + \hat{K}_{11} + k_1\overset{*}{w}_1 + \overset{*}{K}_{11})\delta\hat{w}_1\Big|_{X_1=a} dX_3 - \int_{-b}^{b} (k_1\hat{w}_1 - \hat{K}_{11} + k_1\overset{*}{w}_1 - \overset{*}{K}_{11})\delta\hat{w}_1\Big|_{X_1=-a} dX_3 = 0, \qquad (4.19)$$

where the notation $\Big[\ \Big]_{X_1=-a}^{X_1=a}$ means the quantity evaluated at $+a$ minus the quantity evaluated at $-a$ and the $\delta\hat{K}_{AB}$ are obtained from (3.2) with $\hat{w}_A$ instead of w_A since the A_B are fixed and the variation of prescribed quantities vanishes. Substituting from (4.2), (4.3), (4.15) and (4.17) into (4.19) and performing the integrations, we obtain

$$\sum_m \sum_n a_{mn} D^{(m)} \delta D^{(n)} + \sum_n b_n \delta D^{(n)} = 0,$$

$$m,n = 0, \bar{0}, \pm 1, \pm 2, \ldots \pm N, \qquad (4.20)$$

where 0 and $\bar{0}$ refer to the amplitudes D_o and $\bar{D}_o$, respectively. The expressions for a_{mn} and b_n are very lengthy, provide no additional understanding and were left in the computer where they were generated. Since the variations $\delta D^{(n)}$ are arbitrary, we obtain

$$\sum_n a_{mn} D^{(m)} = -b_n, \quad m,n = 0, \bar{0}, \pm 1, \pm 2, \ldots \pm N, \qquad (4.21)$$

which constitute $2N+2$ inhomogeneous linear algebraic equations in the $2N+2$ unknowns $D^{(m)}$, the inversion of which gives the approximate solution. We consistently obtain good convergence for $N < 14$.

5. The Mode Shape in the Contoured Resonator

A schematic diagram of the contoured crystal resonator along with the associated coordinate system is shown in Fig.2. 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[5]

$$u_{1n} = \sin \frac{n\pi X_2}{2h}\, u_{noo} e^{i\omega_{noo} t}, \quad n = 1,3,5, \ldots, \qquad (5.1)$$

where

$$u_{noo} = e^{-\alpha_n(\hat{X}_1^2/2) - \beta_n(\hat{X}_3^2/2)}, \qquad (5.2)$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2 \hat{c}^{(1)}}{8Rh_o^3 M'_n}, \quad \beta_n^2 = \frac{n^2\pi^2 \hat{c}^{(1)}}{8Rh_o^3 P'_n}, \qquad (5.3)$$

and M'_n and P'_n are involved lengthy expressions that are defined in Sec.II of Ref.5 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.5 in this work we have

$$\hat{X}_A = R_{AB}(X_B - d_B), \qquad (5.4)$$

where, from Eq.(98)_1 of Ref.5

$$R_{AB} = \begin{Vmatrix} \cos\hat{\beta}_n & -\sin\hat{\beta} \\ \sin\hat{\beta}_n & \cos\hat{\beta}_n \end{Vmatrix}, \qquad (5.5)$$

and $\hat{\beta}$ is defined in Eq.(98)_2 of Ref.5. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2 \hat{c}^{(1)}}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \times \left(\sqrt{\frac{M'_n}{\hat{c}^{(1)}}} + \sqrt{\frac{P'_n}{\hat{c}^{(1)}}}\right)\right], \qquad (5.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.5.

In the in-plane case being considered in this work only the dominant displacement field in (5.1) is needed. Consequently, the equivalent trapped energy mode employed in Refs.1 and 2 is not required here.

6. Small Offset of Centers

In this section we obtain a very useful result for the limiting case of small offset of centers. From Fig.1 it is clear that we may write the transformation

$$X_A = d_A + X'_A, \qquad (6.1)$$

and from (4.2), (4.3), (4.15) and (4.17), we may write

$$w_A = \overset{*}{w}_A + \tilde{w}_A + \sum_m D^m e^{i\nu_m X_1} \sum_{p=1}^{4} \beta_A^{(p)} g_p^{(m)} e^{i\eta_m^{(p)} X_3}, \qquad (6.2)$$

where $\overset{*}{w}_A$ and $\tilde{w}_A$ are polynomials in X_1 and X_3. The substitution of (6.1) in (6.2) enables us to write

$$w_A = \overset{*}{w}_A(d_B, X'_B) + \tilde{w}_A(d_B, X'_B) + \sum_m \hat{D}^{(m)} e^{i\nu_m X'_1} \sum_{p=1}^{4} \beta_A^{(p)} \tilde{g}_p^{(m)} e^{i\eta_m^{(p)} X'_3}, \qquad (6.3)$$

where

$$\hat{D}^{(m)} = e^{i\nu_m d_1} D^{(m)}, \quad \hat{g}_p^{(m)} = e^{i\eta_m^{(p)} d_3} g_p^{(m)}. \qquad (6.4)$$

Since the perturbation integral in (2.9) vanishes for $d_A = 0$, the substitution of (6.3) with (6.4) into (2.9) and expansion to first order in d_A yields, with the aid of (2.1),

$$\Delta\omega/\omega)^D = \Gamma^D = C_A^D d_A, \qquad (6.5)$$

where the superscript $D(=1,3)$ represents the 1 and 3 components of the in-plane acceleration sensitivity vector. The C_A^D are complicated expressions which are not terribly revealing and have been calculated.

Since it is of interest to know the maximum value of the magnitude of the in-plane acceleration sensitivity $\Gamma_{(p)}$ for a given $|\underset{\sim}{d}|$, we write

$$\Gamma_{(p)}^2 = \Gamma^D\Gamma^D = C_A^D d_A C_B^D d_B = C_A^D C_B^D Q_A Q_B d^2, \qquad (6.6)$$

where

$$Q_1 = \cos\theta, \quad Q_3 = \sin\theta, \qquad (6.7)$$

and θ is the angle between the X_1-axis and $\underset{\sim}{d}$. Since the maximum value of the magnitude of the in-plane acceleration sensitivity per unit error d is the quantity of actual interest here, we define $\gamma_{(p)}$ by

$$\gamma_{(p)}^2 = \Gamma_{(p)}^2/d^2, \qquad (6.8)$$

which may be maximized by

$$\partial\gamma^2_{(p)}/\partial\theta = 0, \quad \partial^2\gamma^2_{(p)}/\partial\theta^2 < 0 . \tag{6.9}$$

The result is given by

$$\gamma^2_{(p)}\Big)_{max} = \frac{1}{2}\left[C^D_A C^D_A + \Delta_{(p)}\right], \tag{6.10}$$

where

$$\Delta_{(p)} = \sqrt{(C^D_1 C^D_1 - C^D_3 C^D_3)^2 + 4(C^D_1 C^D_3)^2}. \tag{6.11}$$

The foregoing includes the influence of d on the in-plane acceleration sensitivity only. Since the influence of d on the resultant acceleration sensitivity including the normal acceleration sensitivity is of primary practical importance, we extend the treatment to include the influence of d on the normal acceleration sensitivity which was calculated in earlier work[1,2]. This is accomplished simply by writing

$$\Delta\omega/\omega)^L = \Gamma^L = C^L_A d_A , \tag{6.12}$$

in place of (6.5) because D takes the values 1 and 3 only and L takes the values 1, 2 and 3. We then define the magnitude of the resultant acceleration sensitivity $\Gamma_{(r)}$ for a given $|\hat{d}|$ by

$$\Gamma^2_{(r)} = \Gamma^L\Gamma^L = C^L_A d_A C^L_B d_B = C^L_A C^L_B Q_A Q_B d^2. \tag{6.13}$$

As before, we define the magnitude of the resultant acceleration sensitivity per unit error d by $\gamma_{(r)}$ thus

$$\gamma^2_{(r)} = \Gamma^2_{(r)}/d^2 . \tag{6.14}$$

Clearly then $\gamma^2_{(r)})_{max}$ is given by

$$\gamma^2_{(r)})_{max} = \frac{1}{2}\left[C^L_A C^L_A + \Delta_{(r)}\right], \tag{6.15}$$

where

$$\Delta_{(r)} = \sqrt{(C^L_1 C^L_1 - C^L_3 C^L_3)^2 + 4(C^L_1 C^L_3)^2} . \tag{6.16}$$

Calculations of $\gamma_{(p)})_{max}$ and $\gamma_{(r)})_{max}$ are presented and discussed in the next section.

7. Calculated Results

From Sec.5 we now know g^{μ}_{γ} and from Secs.3 and 4 we know $\hat{c}_{L\gamma M\alpha}$ for in-plane acceleration with symmetric rectangular sidewall supports. Hence, we can now evaluate H_{μ} in Eq.(2.9) for nonzero d_A. However, we now note that before the calculations can be performed the scalar equation for H_{μ} in (2.9) must be transformed to the $\hat{X}_A$ coordinate system defined in Eq.(5.4). Such calculations have been performed using the known values of the second order[10] and third order[11] elastic constants of quartz.

The calculated in-plane sensitivities for acceleration in the X_1-direction of an SC-cut plate with an aspect ratio of a/b = 2 and a given support orientation are plotted in Fig.3 as a function of offset in different directions. It can be seen from the figure that all curves go through zero at $d_A = 0$, as they should. For reference the a_1 distance to the point of inflection of the Gaussian in the $\hat{X}_1$-direction is shown in the figure. Since the direction of the relative displacement $\hat{d}$ is unknown, only the most rapidly varying curve is significant. It can be seen that as $|\hat{d}|$ gets small, the curves all become straight lines in accordance with Section 6. The figure shows that the sensitivity increases very rapidly with small offset of centers, which clearly demonstrates the importance of care in fabrication[12,13]. Figure 4 shows the curves for the maximum values of the in-plane acceleration sensitivity as a function of $|\hat{d}|$ for acceleration in the X_1- and X_3-directions, respectively, for the same cut and geometry. This employs the work of Sec.6 for one component of Γ^A only. In addition, the curve for the normal acceleration sensitivity from Refs.1 and 2 is plotted in the figure along with the magnitude of the resultant acceleration sensitivity. The figure shows that the in-plane sensitivity is nearly an order of magnitude lower than the normal sensitivity. Clearly, all curves go through zero at $d_A = 0$, as they should.

Since the linear increase in the in-plane acceleration sensitivity with $|\hat{d}|$ tends to be about an order of magnitude smaller than the linear increase in the normal case, and in earlier work[1,2] an orientation of the support configuration for which the linear increase in the normal case exhibits a strong minimum was found, we calculate the linear increase in the in-plane acceleration sensitivity with $|\hat{d}|$ for the particular aspect ratio (a/b = 2) involved here. Accordingly, the calculated values of C^1_1 and C^1_3 are plotted in Fig.5 as a function of support orientation[14] for an SC-cut plate with an aspect ratio a/b = 2 for N = 1, 3 and 5, which turn out to be nearly the same. Similarly, the calculated values of C^3_1 and C^3_3 are plotted in Fig.6, which also shows that the values are very nearly the same for N = 1, 3 and 5. Figure 7 shows the calculated values of $\gamma_{(p)}$ defined in Eq.(6.10) as a function of support orientation β as the lowest slightly varying large dashed curve in the figure for N = 1, 3 and 5. It also shows $|C^2|$ for the case of normal acceleration for N = 1, $\hat{3}$ and 5 obtained from Refs.1 and 2 referred to the aspect ratio and orientational conventions of this work. These are the small dashed curves in the figure. In addition, Fig.7 also shows $\gamma_{(r)}$ for N = 1, 3, 5 obtained from Eq.(6.15) as a function of β, as the solid curves. These solid curves are the really important curves in the figure because they represent the linear increase in resultant acceleration sensitivity per unit error d. Since these curves all exhibit a strong minimum around $\beta = -50^\circ$, this is a practical orientation for the rectangular support configuration for SC-cut contoured quartz resonators for the reduction of the influence of the type of fabrication imperfection considered in this work.

Acknowledgements

We wish to thank Dr. J. Vig for suggesting that we investigate the influence of a relative displacement of the centers of the mode shape and support rectangle on the acceleration sensitivity.

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 Influence of a Fabrication Imperfection on the Normal 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, 452 (1990).

2. 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," to be published in the Journal of Applied Physics.

3. 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).

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 and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).

6. 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).

7. Since in this work the shear strip is supported symmetrically on both sides, we obtain $k = 2\mu t_f/h^2$ in place of $2\mu t_f/3h^2$, which was obtained when the shear strip was supported on one side only.

8. The inclusion of the polynomials in the solution function increases the rate of convergence by a factor of 2.

9. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4, Eq.(6.44).

10. R. Bechman, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 100. 1060 (1958).

11. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

12. R. Besson, J.J. Gagnepain, D. Janiaud and M. Valdois, "Design of Bulk Wave Quartz Resonator," Proceedings of the 33rd Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, 337 (1979).

13. E.P. Eer Nisse and R.W. Ward, "Crystal Resonator with Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Pat.4,837, 475, June 6, 1989.

14. 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 2, the same aspect ratio is denoted a/b = 2 in this work. This occurs because the direction in which X_1 points is always taken in the long direction and when the wave numbers 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 2 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 2 differs from the orientation β in this work by $\pi/2$.

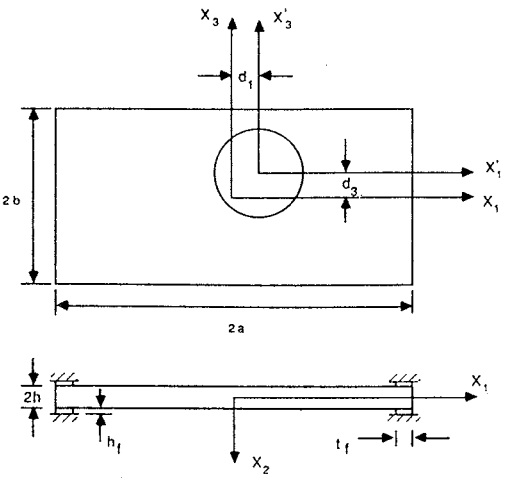


Figure 1 Plan View and Cross Section of Rectangular Quartz Plate Showing Relative Displacement of the Mode Shape and Support Rectangle

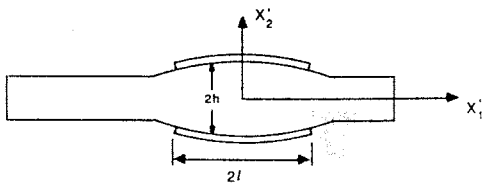


Figure 2 Cross-Section of Bi-Convex Resonator Displaced with Respect to the Supports

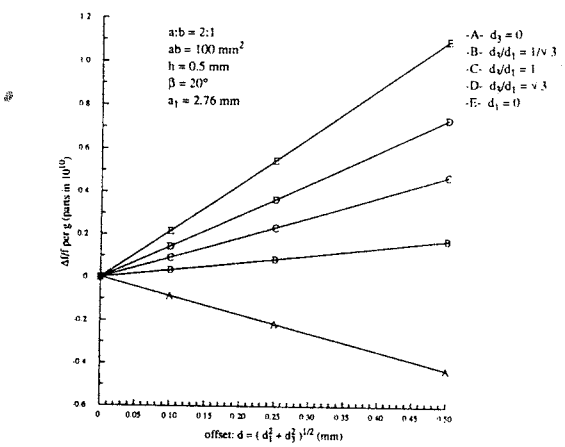


Figure 3 Sensitivity Versus Offset in Different Directions for X_1-Acceleration for the Fundamental Mode of Rectangular SC-Cut Plate

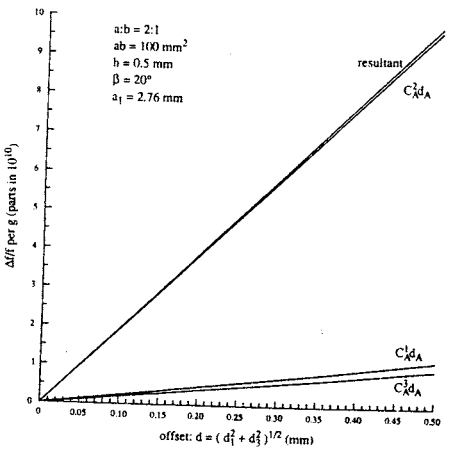


Figure 4 Maximum Values of Each Component of In-Plane Sensitivity, the Normal Sensitivity and the Resultant Versus Offset for Rectangular SC-Cut Plate

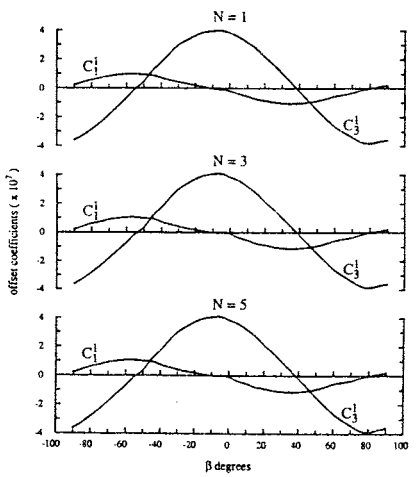


Figure 5 Components of Influence of Offset Vector for X_1-Acceleration Versus Orientation of Support Rectangle of SC-Cut Plate

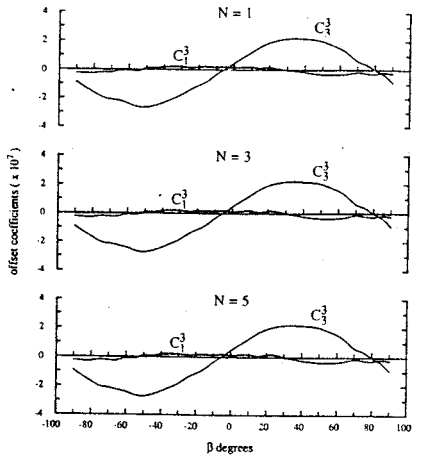


Figure 6 Components of Influence of Offset Vector for X_3-Acceleration Versus Orientation of Support Rectangle of SC-Cut Plate

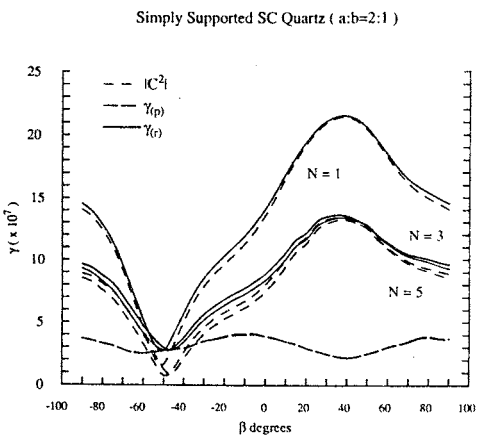


Figure 7 Maximum Value of the Magnitude of the Sensitivity per Unit Error Versus Orientation of Support Rectangle of SC-Cut Plate

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 5A Number 28

A GENERAL MODEL FOR QUASI THICKNESS VIBRATIONS IN CONTOURED QUARTZ PLATES

B. DULMET

Laboratoire de Chronométrie, Electronique et Piezoélectricité,
Ecole Nationale Supérieure de Mécanique et des Microtechniques

La Bouloie, Route de Gray, 25030 BESANCON CEDEX - FRANCE -

ABSTRACT

A convenient framework for establishing approximate solutions of vibration problems in monoclinic crystal plates is given by modal decomposition methods :

① a first step consists of solving partial derivative equations and boundary conditions, where a limited subset of material coefficients is retained,

② the whole set of coefficients is taken into account in the second step : the complete equations are then solved bydecomposing the solutions over the basis of previously obtained simple solutions.

In this way, the whole set of partial derivative equations and associated boundary conditions are eventually turned into an infinite matrix eigenvalues/eigenvectors problem. All equations are mapped onto the natural state of coordinates, thereby eliminating problems due to complex geometric changes that affect the shape of the resonator under the effect of thermal expansion. Solving the eigenvalues problem by the reverse-iteration method, with a basis including up to one thousand modes, we achieve drastic improvements from an earlier work based on a simple perturbation iteration method.

INTRODUCTION

In this paper, we study quasi-thickness modes of quartz resonators submitted to some slowly-varying, homogeneous, temperature change. The shape of the resonator slightly evolves during this change, as a consequence of anisotropic thermal expansion. In the case of a plano-convex geometry, this fact makes difficult a writing of the boundary conditions in the so-called initial (intermediate) state defined as the quasi-static equilibrium, at current temperature, upon which the vibration is superimposed. This is a very good reason to adopt a lagrangian description, mapping all equations to a known set of coordinates, defined at some fixed temperature, usually 25°C. Thanks to this description, we can forget anything about subsequent finite changes in geometry under the effect of thermal expansion. This obvious advantage is associated with a drawback, indeed : the effective elastic, piezoelectric and dielectric coefficients appearing in equations of motion have a weaker symmetry than usual coefficients encountered in linear Piezolectricity. For instance, effective elastic coefficients reduce into a 9×9 matrix, instead of a 6×6 matrix. All notations of this paper remain consistent with conventions used in a previous paper [1], where numerical values of the thermal derivatives of effective stiffnesses were deduced from the well-known BBL set of experimental data [2].

Although it should be known for a correct mastering of the formalism, the theoretical background of which may be found at Refs [3,4], the weaker symmetry of lagrangian effective stiffnesses does not include any additional difficulty to solve partial derivatives equations of motion.

Only studying quasi-thickness modes, we found very convenient to decompose the unknown mechanical displacement on the triad of eigensolutions of Christoffel's problem for the desired plate orientation [5]. Transformed equations are thus obtained. Neglecting all mixed derivatives as well as intercomponents coupling terms, we get a simplified problem. Its mechanical displacement solutions are colinear to either one of the three eigendisplacements that would propagate along the thickness of an infinite plate instead of the finite studied resonator. All solutions of the simplified problem provide us with the needed initial modal basis. This first model is little more than a generalization of Wilson's model [6]. It should be considered as inaccurate to describe properties of guided waves in anisotropic thin plates, but is sufficient to establish a basis upon which more accurate solutions are constructed.

Previously ignored partial derivatives terms are taken into account in the second step of the method, as perturbation terms. It can be shown [7] that both Green's tensor and variational methods turn then the partial derivatives equations into the same infinite matrix eigenvalues problem. Eigenvectors components are the decomposition coefficients of the actual modes over the initial basis of simplified solutions. Our model gives access to all three components of actual modes. Each component is actually generated by the decomposition over a different subset of the modal basis, constituted by either thickness extensional, or slow shear, or fast shear modes.

Since decomposition coefficients depend on temperature, the model is efficient to study activity-dip phenomena which occur when some frequencies of the modal basis cross during temperature variations [8].

In addition, our model predicts a rotation of the mode patterns around the plate normal of doubly rotated cuts, as did the asymptotic model previously proposed by Stevens and Tiersten [5]. Despite different predicted values for this rotation, both models tend to agree about modal identification.

I. Basic Equations

We define the natural state of the crystal as the stress free static equilibrium state at reference temperature 25°C. The set of X_L material coordinates of the body in this state defines the so-called reference configuration, or Lagrange configuration. Any intermediate, stress free, static state, at a different temperature, defines a corresponding intermediate set of material coordinates denoted by ξ_α. The actual state of the vibrating body defines the final configuration, i.e. a new set of coordinates denoted by y_j. Thus, the mechanical displacement can be decomposed in two parts : the static part, induced by thermal expansion, and the dynamic incremental part, denoted by u_α :

$$y_j = (\xi_\alpha + u_\alpha)\delta_{j\alpha} \qquad (1)$$

The electric potential could be decomposed in the same manner, but the static part $\bar{\phi}$ can be ignored in this paper, since we assume that no static field is exerted on the crystal. ϕ^- denotes the dynamic potential of our interest. Using the theory of infinitesimal dynamic fields superposed on a bias [4], it can be shown that incremental equations of motion, mapped onto the X_L coordinates may be written as follows :

$$\left.\begin{aligned} G_{L\alpha M\beta}u_{\beta,ML} + R_{A\alpha L}\phi^-_{,AL} &= \varrho_0 d^2u_\alpha/dt^2 \\ R_{A\alpha L}u_{\alpha,LA} - N_{AB}\phi^-_{,AB} &= 0 \end{aligned}\right\} \qquad (2)$$

Because of the mapping onto the reference coordinates, all partial derivatives appearing in any equation of this paper are unambiguously taken with respect to X coordinates, thereby facilitating the use of the comma convention. In the preceeding equations, the mass density ϱ_0 is taken in the natural state. Thus, it remains constant with respect to temperature variations. The G, R, N constants are respectively called effective elastic, piezoelectric and dielectric constants referred to the reference (natural) configuration. They are defined as the coefficients of linear expansion of the incremental mechanical stress tensor $\mathcal{L}$ and the incremental electric displacement $\mathcal{D}$ versus mechanical displacement and electric potential gradients :

$$\left.\begin{aligned} \mathcal{L}^-_{Lj} &= \delta_{j\alpha}[G_{L\alpha M\beta}u_{\beta,M} + R_{A\alpha L}\phi^-_{,A}] \\ \mathcal{D}^-_{A} &= R_{A\alpha L}u_{\alpha,L} - N_{AB}\phi^-_{,B} \end{aligned}\right\}$$

Because Piola's stress tensor $\mathcal{L}$ is asymmetric, effective coefficients exhibit restricted symmetries by comparison with corresponding c, e, ε tensors in linear Piezoelectricity :

$$\left.\begin{aligned} &G_{L\gamma M\beta} \equiv G_{M\beta L\gamma} \quad \text{but} \quad G_{L\gamma M\beta} \neq G_{L\beta M\gamma} \\ &R_{A\alpha L} \neq R_{AL\alpha} \\ &N_{AB} \equiv N_{BA} \end{aligned}\right\} \qquad (4)$$

As a consequence, the required index compression for matricial representation is made according to the following rules, which result into $G_{9\times9}$, $R_{3\times9}$, $N_{3\times3}$ matrixes :

$L\gamma$	11	22	33	23	31	12	32	13	21
λ	1	2	3	4	5	6	7	8	9

(5)

At reference temperature, the effective coefficients present the usual symmetries and are *actually* identical to c, e, ε coefficients. This is no more true at another temperature. Thus, using them implies specific thermal derivatives. Such tables for $T^{(n)}G_{\lambda\mu}$ can be found in [1] and [7]. Taking advantage of small piezoelectric coupling in quartz, we neglect the thermal derivatives of R and N coefficients. The reader is compelled to notice that all preceeding formulas would remain valid without this approximation.

We can assume homogeneous boundary conditions to find the natural frequencies of electrodeless resonators. Furthermore, trapped modes are essentially determined by the conditions on the major surfaces of the thin plates :

$$\left.\begin{aligned} \mathcal{L}^-_{2\alpha} &= 0 \\ \mathcal{D}^-_{2} &= 0 \end{aligned}\right\} \text{ at } X_2 = \pm h_0(X_1, X_3) \qquad (6)$$

Because of the mapping on the natural set of coordinates, we have not to bother with the variations of the actual boundary surface when temperature varies.

The effective coefficients rotate according usual tensorial rules. In the followings, all equations are assumed to be written in a doubly rotated frame in such a way that the plate normal lies along X_2 axis. All orientations are consistently defined at 25°C. Christoffel's problem for plane waves is written as follows :

$$[G_{2\alpha2\beta} + R_{2\alpha2}R_{2\beta2}/N_{22}]u^\mu_\beta = \bar{G}^\mu u^\mu_\alpha \qquad (7)$$

$\bar{G}^\mu/\varrho_0$ is the square of the plane wave velocity at current temperature. The μ superscript indicates the propagation mode : A-thickness extensional, B-fast shear, C-slow shear. Putting normalized solutions of (7) in a square matrix, the mechanical displacement u is projected onto the triad of eigenvectors, thereby defining new mechanical unknown v :

$$v_\mu = u^\mu_\alpha u_\alpha \qquad (8)$$

Using orthogonality of this transformation while dropping the u→v variable substitution in (2) results into the following transformed equations :

$$\left.\begin{aligned} g_{L\delta\gamma M}\, v_{\gamma,ML} + r_{AL\delta}\, \bar{\phi}_{,AL} &= \rho_0 d^2 v_\alpha / dt^2 \\ r_{AL\beta}\, v_{\beta,AL} - N_{AB}\, \bar{\phi}_{,AB} &= 0 \end{aligned}\right\} \qquad (9)$$

where :

$$\left.\begin{aligned} g_{L\delta\gamma M} &= Q_{\delta\alpha}\, G_{L\alpha M\beta}\, Q_{\gamma\beta} \\ r_{AL\delta} &= Q_{\delta\alpha}\, R_{A\alpha L} \end{aligned}\right\} \qquad (10)$$

We did all conventional definitions in such a way that **g** coefficients have the same symmetries as transformed **c** coefficients appearing in Stevens' and Tiersten's paper [5] :

$$g_{L\delta\gamma M} \equiv g_{M\gamma\delta L} \qquad (11)$$

In this paper, we adhere to Tiersten's quasi thickness approximation. Hence, we only retain pure thickness variations of any terms figuring in Poisson's equation. Thus, all transverse variations attached to piezoelectric coefficients in Eq.(9a) are ignored, and Poisson's equation is easily eliminated. As an example, we give here the developped equation obtained for Eq.(9a), with δ=1 :

$$\begin{aligned} &\{ \bar{G}^1 v_{1,22} + g_{11} v_{1,11} + g_{58} v_{1,33} \} \\ &+ 2g_{16} v_{1,12} + 2g_{56} v_{1,23} + 2g_{18} v_{1,13} \\ &+ (g_{12}+g_{66}) v_{2,12} + (g_{28}+g_{76}) v_{2,23} \\ &+ (g_{17}+g_{86}) v_{3,12} + (g_{36}+g_{48}) v_{3,23} \\ &+ g_{19} v_{2,11} + (g_{14}+g_{59}) v_{2,13} + g_{54} v_{2,33} \\ &+ g_{15} v_{3,11} + (g_{13}+g_{55}) v_{3,13} + g_{38} v_{3,33} \\ &= \rho_0 d^2 v_1 / dt^2 \end{aligned} \qquad (12)$$

Because of Eq.(7), $g_{2\delta\gamma 2}$ coefficients with δ≠γ disappear from the developped equations.

Linearity and orthogonality of u→v transformation turn the homogeneous boundary conditions (6) into the following transformed set, which also remains homogeneous :

$$\left.\begin{aligned} \bar{l}_{2\alpha} &= 0 \\ \bar{\mathcal{D}}_2 &= 0 \end{aligned}\right\} \text{at } X_2 = \pm h_0(X_1, X_3) \qquad (13)$$

with :

$$\bar{l}_{2\alpha} = g_{2\alpha\gamma M} v_{\gamma,M} + r_{A2\alpha} \bar{\phi}_{,A}$$

II. Modal Basis

The modal basis consists of the solutions of a simplified problem neglecting all coupling terms between different components v_α and v_β in Eqs.(9), as well as all cross derivative terms. For instance, only the terms appearing between the pair of braces are retained in the left hand member of Eq.(12). In this way, Eqs.(9a) give three simplified uncoupled equations :

$$\bar{G}^\alpha v^\lambda_{\alpha,22} + g_{1\alpha\alpha1} v^\lambda_{\alpha,11} + g_{3\alpha\alpha3} v^\lambda_{\alpha,33} = -\rho_0\, \omega_\lambda^2 v^\lambda_\alpha \qquad (\alpha = 1,2,3) \qquad (14)$$

We adopt the following notation to indicate the various propagation families :

$$\left.\begin{aligned} &\alpha=1 \quad \text{slow shear C mode} \\ &\alpha=2 \quad \text{thickness extensional A mode} \\ &\alpha=3 \quad \text{fast shear B mode} \end{aligned}\right\} \qquad (15)$$

According to Eq.(14) and boundary conditions (13) on the major surfaces of the plate, the thickness dependency of the solution can be obtained separately from in-plane variations :

$$\left.\begin{aligned} v^\lambda_\alpha &= V^\lambda(X_1, X_3) \sin(n\pi X_2/2h_0)\, e^{j\omega t} \quad n \text{ odd} \\ v^\lambda_\alpha &= V^\lambda(X_1, X_3) \cos(n\pi X_2/2h_0)\, e^{j\omega t} \quad n \text{ even} \end{aligned}\right\} \qquad (16)$$

Both symmetric and antisymmetric modes should appear in the modal basis, although we are only interested in final results exhibiting a main componentantisymmetric along the thickness, according to usual piezoelectric driving conditions. Substituting (16) into (14) and expanding the thickness of the spherically contoured resonator up to the second order in $O(X_\alpha/R)$, where R is the radius of contour, we get the following equations :

$$\left.\begin{aligned} &g_{1\alpha\alpha1} V^\lambda_{,11} + g_{3\alpha\alpha3} V^\lambda_{,33} \\ &+ [\rho_0\omega_\lambda^2 - \bar{G}^\alpha \frac{n^2\pi^2}{4h_0^2}(1 + \frac{X_1^2+X_3^2}{2h_0 R})] V^\lambda = 0 \end{aligned}\right\} \qquad (17)$$

We adopt the variable change earlier proposed by Vangheluwe [9] :

$$\left.\begin{aligned} x &= K X_1 \quad K = (\sqrt{n\pi})(\bar{G}^\alpha / 2g_{1\alpha\alpha1} R h_0^3)^{1/4} \\ z &= L X_3 \quad L = (\sqrt{n\pi})(\bar{G}^\alpha / 2g_{3\alpha\alpha3} R h_0^3)^{1/4} \end{aligned}\right\} \qquad (18)$$

The V^λ amplitude variations along the dimensionless coordinates x and z may readily be separated in Eq.(17) which is then turned into the folowing set :

$$\left.\begin{aligned} V^\lambda(x,z) &= f^\lambda(x) g^\lambda(z) \\ d^2 f^\lambda / dx^2 &- [\,(1/4)x^2 + a\,] f^\lambda = 0 \\ d^2 g^\lambda / dz^2 &- [\,(1/4)z^2 + b\,] g^\lambda = 0 \\ g_{1\alpha\alpha1} K^2 a &+ g_{3\alpha\alpha3} L^2 b = \bar{G}^\alpha n^2\pi^2/4h_0^2 - \rho_0\omega_\lambda^2 \end{aligned}\right\} \qquad (19)$$

Finite solutions at infinity are obtained for special values of the dimensionless parameters a and b :

$$\left.\begin{aligned} a &= -(2m+1)/2 \\ b &= -(2p+1)/2 \end{aligned}\right\} \quad (20)$$

Symmetric and antisymmetric solutions respectively correspond to even and odd values of m and p. They are Hermite functions and provide us with a suitable basis to construct more accurate trapped solutions. See Ref.[9] for more general solutions of (19) than Hermite functions. Modal basis eigenfrequencies are easily obtained by substituting (20) into (19). In the followings, modal basis solutions are denoted by v^{λ} and ω^{λ}, where λ is a symbolic index for the four numbers (α,n,m,p) respectively determining the thickness propagation group, the overtone number, and anharmonic ranks along X_1 and X_3. Metrological modes of common use correspond to zero values of m and p, giving a gaussian amplitude distribution along in-plane axes.

III. Eigenvalues Problem

We adopt the convention to denote by a '0' superscript any quantity 'generated' by a material coefficient previously taken into account in Sect.II, and by a '+' superscript any quantity 'generated' by previously ignored material coefficients . Then, the full set of partial derivatives equations may be put in the form :

$$\left.\begin{aligned} \overset{-0}{l}_{L\alpha,L} + \varrho_0 \omega^2 v_\alpha &= -\overset{-+}{l}_{L\alpha,L} \\ \overset{-0}{\mathcal{D}}_{L,L} &= -\overset{-+}{\mathcal{D}}_{L,L} \end{aligned}\right\} \quad (21)$$

It has been shown that, using Green's tensor formalism, those equations, together with boundary conditions, are mathematically equivalent to the following integral equation :

$$\mathcal{F}_\gamma(P) = \int_{S0} \overset{0}{n}_L \left[\overset{0\gamma}{l}_{L\alpha} v_\alpha - \overset{-}{l}_{L\alpha} \Gamma^\gamma_\alpha + \overset{0\gamma}{\mathcal{D}}_L \overset{-}{\phi} - \overset{-}{\mathcal{D}}_L \Gamma^\gamma_4 \right] dS_0$$

$$+ \int_{V0} \left[\overset{-+}{l}_{L\alpha} \Gamma^\gamma_{\alpha,L} + \overset{-+}{\mathcal{D}}_L \Gamma^\gamma_{4,L} \right] dV_0(Q)$$

$$(\gamma = 1,2,3,4) \quad (22)$$

$\mathcal{F}$ is the complete piezoelectric solution, put in a vector of dimension four :

$$\left.\begin{aligned} \mathcal{F}_\gamma(P) &\equiv v_\alpha(P) \quad \forall \alpha = 1,2,3 \\ \mathcal{F}_4(P) &\equiv \overset{-}{\phi}(P) \end{aligned}\right\} \quad (23)$$

while $\Gamma^\gamma_\alpha(P,Q)$ is the Green's tensor. P indicates a field-point, and Q a source-point, i.e. a point in the integration domain V_0 of (22). $\overset{0\gamma}{l}_{L\alpha}$ and $\overset{0\gamma}{\mathcal{D}}_L$ are Green's respective equivalent of previously defined $\overset{-0}{l}_{L\alpha}$ and $\overset{-0}{\mathcal{D}}_L$ ordinary tensors :

$$\left.\begin{aligned} \overset{0\gamma}{l}_{L\alpha} &= \overset{0}{g}_{L\alpha\beta M} \Gamma^\gamma_{\beta M} + \overset{0}{r}_{AL\alpha} \Gamma^\gamma_{4A} \\ \overset{0\gamma}{\mathcal{D}}_A &= \overset{0}{r}_{AL\beta} \Gamma^\gamma_{\beta L} - \overset{0}{N}_{AB} \Gamma^\gamma_{4B} \end{aligned}\right\} \quad (24)$$

In agreement with previous works [11,12], we find that the components Γ^γ_β (γ=1,2,3) may be expanded as series in terms of the elements of the modal basis :

$$\left.\begin{aligned} \Gamma^\gamma_\alpha(P,Q) &= \sum_\lambda \frac{v^\lambda_\gamma(P)}{\omega^2 - \omega^2_\lambda} v^\lambda_\alpha(Q) \\ (\alpha &= 1,2,3) \\ \Gamma^\gamma_4(P,Q) &= \sum_\lambda \frac{v^\lambda_\gamma(P)}{\omega^2 - \omega^2_\lambda} \overset{-}{\phi}_\lambda(Q) \end{aligned}\right\} \quad (25)$$

This result is attained through orthogonality of v^{λ} initial solutions. In the same way, the complete solution $\mathcal{F}$ itself is searched as a decomposition over the modal basis :

$$\mathcal{F} = \sum_\lambda C_\lambda \mathcal{F}^\lambda \quad (26)$$

Putting both (25) and (26) into the integral equation (22) allows us to turn it into an algebraic system, after some calculations to lengthy to be reproduced here [7] :

$$\left.\begin{aligned} C_\mu v^\mu_\alpha &= \frac{v^\lambda_\alpha}{\omega^2 - \omega^2_\lambda} q_{\lambda\upsilon} C_\upsilon \\ q_{\lambda\upsilon} &= \int_{V0} \left[\overset{-+\upsilon}{l}_{L\alpha} v^\lambda_{\alpha,L} + \overset{-+\upsilon}{\mathcal{D}}_L \overset{-\lambda}{\phi}_{,L} \right] dV_0 \end{aligned}\right\} \quad (26)$$

$\overset{-+\nu}{l}$ and $\overset{-+\nu}{\mathcal{D}}$ are defined as the particular values of $\overset{-+}{l}$ and $\overset{-+}{\mathcal{D}}$, defined in Eq.(23), for the modal basis elements $\mathcal{F}^\lambda$ in place of the general solution $\mathcal{F}$. It is worth pointing out that Eqs.(26) are valid only with associated homogeneous boundary conditions for both the simplified initial problem of Sect.II and the complete problem. Using orthogonality of the modal basis, Eqs.(29) are quickly turned into an eigenvalues problem :

$$\left.\begin{aligned} K_{\lambda\mu} C_\mu &= \omega^2 C_\lambda \\ K_{\lambda\mu} &= q_{\lambda\mu} + \omega^2 \delta_{\lambda\mu} \end{aligned}\right\} \quad (27)$$

As explained near the end of Sect. I, we stick to quasi-thickness approximation. Hence, $\overset{-+}{\mathcal{D}}$ is neglected in this paper. Its influence is supposedly small, because of the weak piezoelectric coupling in quartz. Nevertheless, Eqs.(29) and (30) leave open a possibility to include additional piezoelectric and dielectric terms in the here-represented model, in order to extend it for such driving conditions as parallel field techniques.

IV. Numerical Techniques

The dimensions of the eigenvalues/eigenvectors problem (27) are theoretically infinite. Although we are able to only examinate finite dimensions problems, their dimensions are not small. Hence, we have to carefully choose numerical techniques to solve the system.

The size of the K matrix is determined by the amount of anharmonic modes and the maximum overtone number of the initial solutions incorporated into the modal basis. A typical basis may include any solution described in Sect.II, limited by : $n \leq 8$, $m+p \leq 7$. This results into a symmetric 864×864 matrix, storing half of which requires almost 3 Megabytes of RAM, according to IEEE double precision floating point representation. Increasing the maximum overtone up to 12 results into a 1296×1296 matrix. The storage space amounts then to nearly 6.7 Megabytes of memory. The eigenvalues problem was solved on a SUN 4/110 RISC workstation, initially equipped with 8 Megabytes of memory. Recently upgrading to 20 Megabytes, we can easily solve 1296×1296 problems, and larger. Nevertheless, we think more reasonable sticking to numerical methods that avoid any auxiliary matrix as large as K itself.

A safe method to obtain an intermediate eigenvalue with its related eigenvector is the shifted reverse iteration method. A linear system is solved at every iteration, according to the following scheme :

$$\left.\begin{aligned} K'.X^{i+1} &= V^{i} \\ X^{i+1} &= (1/\Omega^{2}_{i+1})V^{i+1} \end{aligned}\right\} \quad (28)$$

where the $(i+1)^{th}$ approximation of Ω is given by the value of Rayleigh's ratio :

$$\Omega^{2}_{i+1} = \frac{V^{i+1}.K.V^{i}}{V^{i+1}.V^{i}} \quad (32)$$

Convergence is attained when the series Ω_i is stable. This method gives the smallest eigenvalue of the matrix K'. This is the reason explaining why it is necessary to replace the initial K matrix by K' , through a shift of the diagonal elements of the former :

$$\left.\begin{aligned} K' &= K - sI \\ \Omega^{2} &= \omega^{2} - s \end{aligned}\right\} \quad (33)$$

The value of the shift s should be close to the searched eigenvalue and smaller than it, in order that the shifted eigenvalue be the smallest positive eigenvalue of K' . This is the main drawback of the method : you always get convergence, but not necessarily towards the mode of your interest. To overcome this difficulty, we started from elaborate initial estimations of the frequency and eigenvector, obtained by solving a reduced subsystem that includes a hundred of elements of the modal basis. This subset retains the most important terms appearing in a second order perturbation expansion of the frequency, given by a previous simpler model [[12]. The 100×100 subsystem is readily solved by Jacobi's method. Although it is far from being straightforward, this approach shows perfect efficiency for contoured AT cut modes, but it may be necessary to adjust the shift s for some anharmonic modes in contoured SC cuts. Frequency convergence (10^{-8}, relatively), is attainedt hrough a number of iterations usually ranging from 5 to 20. Getting one eigenvalue/eigenvector pair for a 864×864 matrix requires roughly 120s on the above-mentionned computer system.

The series of linear systems (31) are solved by a gaussian direct factorization method (Choleski). The required computing time is reduced by a factor 2.2 with help of sparse matrixe techniques. Because of symmetry properties involved by expressions (26b) of $q_{\lambda\mu}$ matrix elements, only the third of these elements are non zero. The factorization process induces some fill-in of the K matrix, but this fill-in is reduced by a proper initial reordering of the modal basis, in such a way that initial simplified modes caracterized by the same symmetries are contiguously numbered. Thus, the ratio of non-zero matrix elements is held below 50% upon completion of the direct factorization.

As a refinement of simple reverse iteration method, the subspace method is largely used in various finite elements software packages. It provides with a means to avoid adjusting the initial shift s and reconstructing the factorized matrix for each solution : this is because this method consists of finding a subset of shifted lower frequencies solutions of K' in a single run, from a subset of initial estimations. But we found it not worth the additional programming effort, since finding several solutions at once was accompanied by a strong increase of the iterations number and some convergence difficulties. This is due to the fact that complete solutions ordering differs from modal basis ordering, possibly resulting in complete solutions that are orthogonal to the initial estimations subset. For this reason, all results presented in the following section have been obtained by the simple reverse iteration method.

V. Obtained Results

The major interest of this model is to provide with a detailed description of the mode shape, including all three components. In addition, predicted frequencies are extremely close to values predicted by other models [5,12]. This is true to such a point that a detailed model comparison based on the frequencies would not be very meaningful, since non negligible errors occur on some construction parameters of actual resonators : for instance, the maximum thickness is known up to 4 digits, at most....

Thus,we concentrate our interest to a presentation of some predicted mode patterns. For instance, Figs 1a-c show the

cut resonator. The v_3 component pattern is in agreement with experimental topograps performed in our laboratory with help of X rays diffraction apparatus. Fig.2a-d show predicted major components topograph simulations in an SC cut contoured resonator presenting a (C,3,0,0) mode at 5 MHz. The reference frame used for these figures is counterclockwise rotated by 9° around the plate normal of the usual doubly rotated frame. Because the modal basis is modified by such a rotation, the predicted frequencies and absolute orientations slightly depend on its value. We think that the most reliable results are achieved when the predicted modes shapes are aligned with respect to in-plane axes of the rotated frame : the complete solutions are then closer to simplified initial solutions, thus minimizing the number of iterations and probably limitating the inaccuracy of the numerical solution. Adopting this criterion, we respectively found third rotations values equal to 11°, 10°, 16° to align (C,3,m,p), (C,5,mp) and (B,3,m,p) SC-cut modes shapes in the working frame. Corresponding orientations predicted by Stevens and Tiersten model respectively amount to 21°, 29° and 32°. Both models generally tend to the same modal identification.

In addition, the here presented model confirmed all previous results established on studying strong activity-dip phenomena occuring in 7th overtone of C modes in contoured AT cut resonators[8]. Such phenomena arise because some (B,6,m,p) modes cross the frequency range of (C,7,r,s) modes when temperature varies. As another checking, we could run our model with the construction parameters appearing in [13]. Then, we observed strong variations of the theoretical v_3 component of (C,7,0,0) and (C,7,2,0) modes along temperature variations. At 60°C and 35°C (theoretical), the predicted modes shapes are in agreement with experimental topographs produced by Capelle *et al.* in [13]. Some corresponding simulations are produced on Figs 3 and 4 .

It should be understood that precise simulations of mode patterns rapidly evolving with temperature are especially subject to cumulative inaccuracies attached to both the vibration model and the thermal sensitivities of material coefficients [1].

References

[1] B. Dulmet, R. Bourquin, "Influence of the definition of material constants in the study of crystal plates thermal behavior", Proc. of 3rd European Frequency and Time Forum, pp 305-312, (1989).

[2] R. Bechmann, A. Ballato, T. Lukaszek, "High Order Temperature Coefficients of The Elastic Stiffnesses and Compliances of Quartz", P.I.R.E, pp 1812 sq. (1962).

[3] H. Tiersten, "On the Non Linear Equations of ThermoElectroElasticity", Int. J. Engng. Sci., Vol 9, pp 587-604, (1971).

[4] J. C. Baumhauer, H. F. Tiersten, "Non linear electroelastic equations for small fields superposed on a bias", J. Ac. Soc. Am., Vol 54, N°4, pp 1017-1034, (1973).

[5] D. S. Stevens, H. F. Tiersten, "An Analysis of Doubly Rotated Quartz Resonators utilizing Essentially Thickness Modes with Transverse Variations", J. Ac. Soc. Am., Vol 79, n°6, pp 1811-1826, (1986).

[6] C. Wilson, "Vibration Modes of At cut Convex Quartz Resonators", J. Phys. D: Appl. Phys., Vol 7, pp 2449 sq., (1974).

[7] B. Dulmet, "Contribution à l'Etude des vibrations à énergie piégee des résonateurs piézoélectriques à ondes de volume : comportement thermique statique et modes de vibration", Thèse Phys. Besançon, 1991.

[8] R. Bourquin, B. Dulmet, G. Genestier, "Mechanical couplings Involving Discontinuities of the frequency-temperature curves of contoured quartz resonators", Proc. of 39th Ann. Freq. Cont. Symp., pp 405-414, (1985).

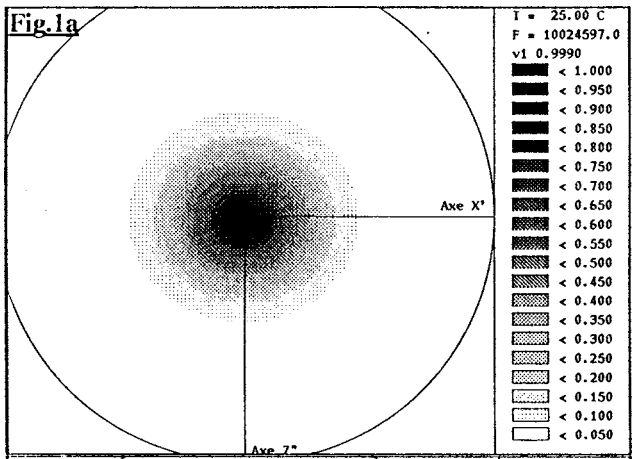
[9] D.C Vangheluwe, "The Frequency and Motional Capacitance of Partial Contoured crystal Resonators", Proc. 34th Freq. Cont. Symp., pp 412-418 (1980).

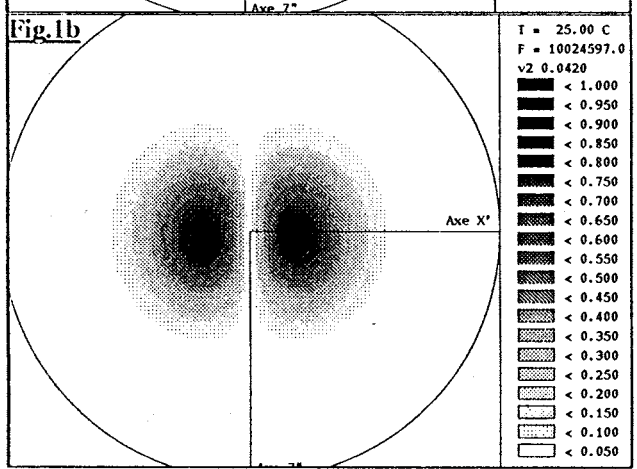
[10] B. Dulmet, H. F. Tiersten, "On The Effect of Edge Boundaries on Nearly Trapped Vibrations of Contoured Rectangular Quartz Resonators", Proc. of 1986 IEEE Ultrasonics Symposium, pp 355-360, (1986).

[11] H. F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias", J. Ac. Soc. Am., Vol 64, N°3, pp 832-837, (1978).

[12] B. Dulmet, "Application d'une méthode de perturbation à l'étude de résonateurs à quartz présentant des modes d'épaisseur à énergie piégée", Rev. Phys. Appl. Vol 19, pp 839-849, (1984).

[13] B. Capelle, Y. Zheng, A. Zarka, J. Detaint, J. Schwartzel, "Recent Results in Study of Piezoelectric Resonators by X-Ray Topography using Synchrotron Radiation", 4th European Frequency and Time Forum, pp 675-682, Neuchâtel CH (1990).





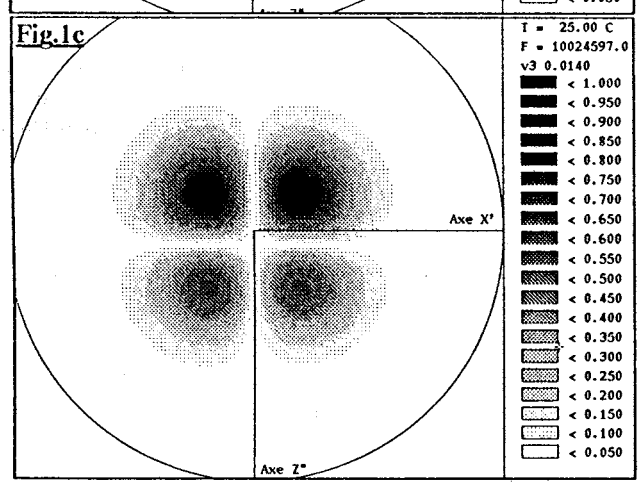
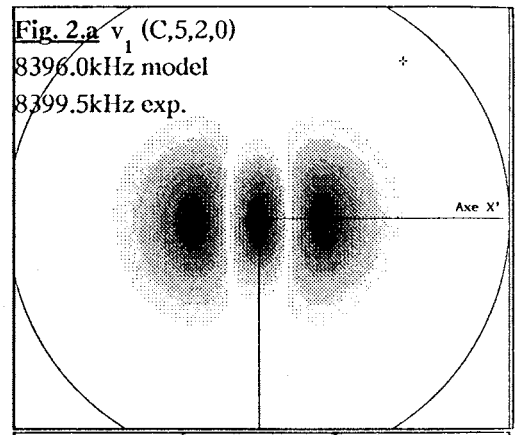


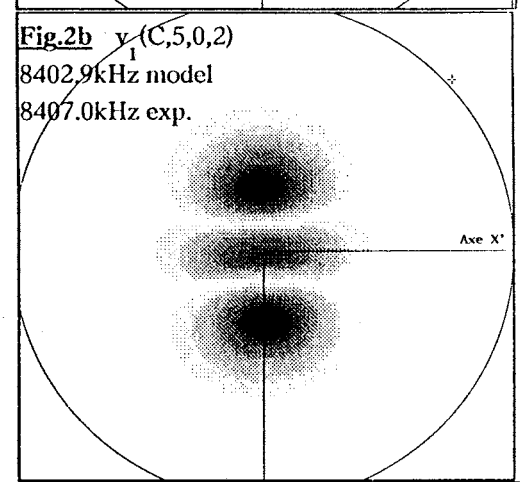
Fig. 1

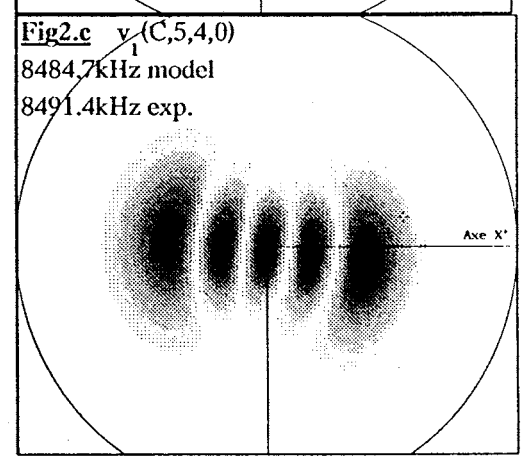
v_1 v_2 v_3 components of AT cut (C,5,0,0) mode at 25°C

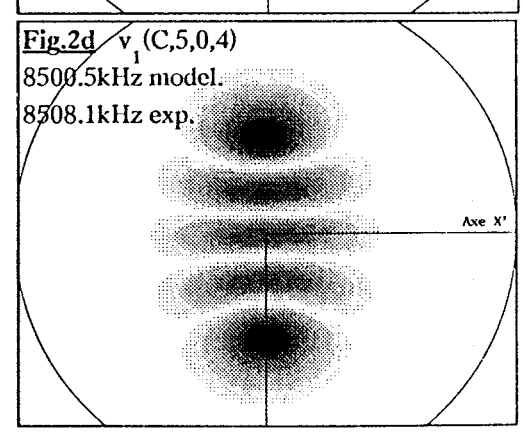
$\phi = 0°$ $\theta = 35°27'$

$2h_0 = 0.835$mm $R = 150$mm







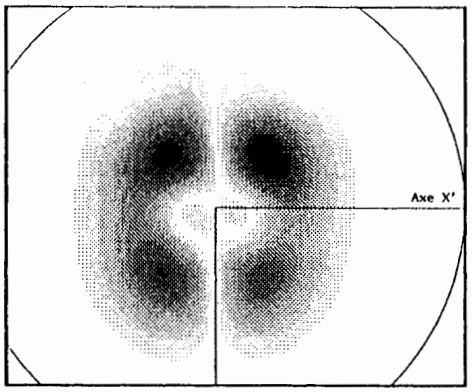


For all Figs.2 :

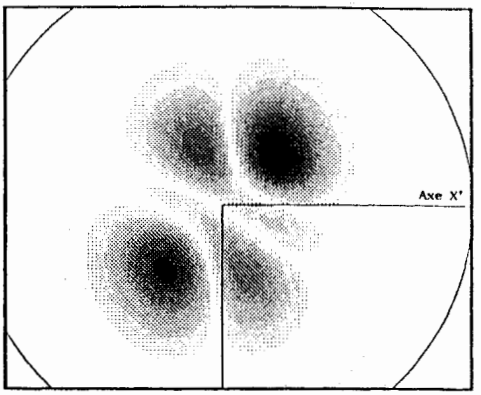
v_1 components in SC cut C modes at 25C

$\phi = -22°25'$ $\theta = 34°06'30"$

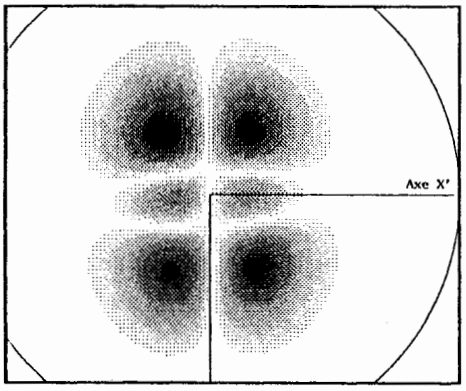
$2h_0 = 1.093$mm $R = 250$mm



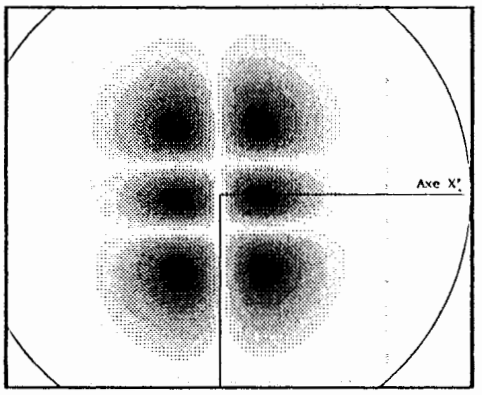
T = 0.00°C, F = 8814968.0 Hz
v1=0.999, v2= 0.017, v3 =0.042



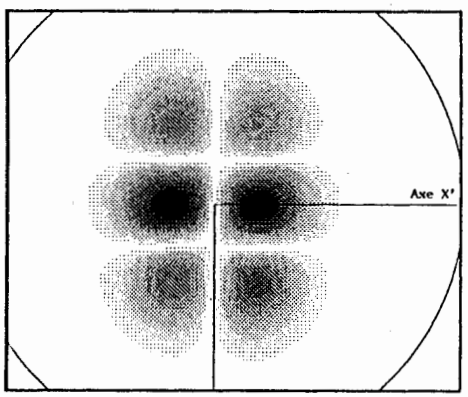
T = 20.00°C, F = 8814876.0 Hz
v1 =0.997, v2= 0.022, v3 =0.076



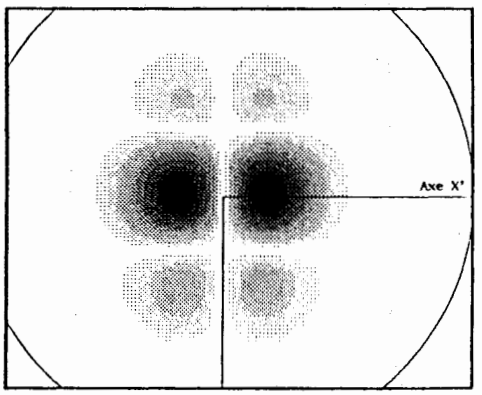
T = 40.00°C, F = 8814718.0 Hz,
v1= 0.995, v2 =0.028, v3 =0.101



T = 60.00°C, F = 8815358.0 Hz
v1= 0.904, v2= 0.099, v3= 0.416
strong couplin with (B,3,1,2)mode
agrees with Fig 9b Ref [13]



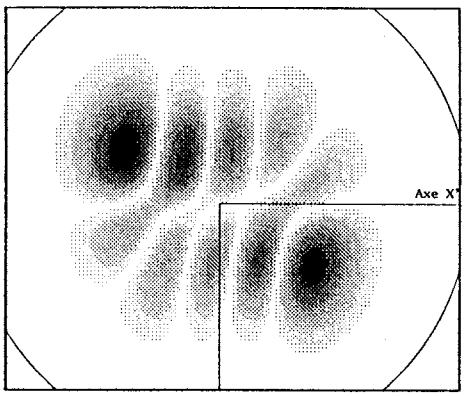
T = 100.00°C, F = 8814889.0 Hz,
v1= 0.998, v2= 0.022, v3= 0.066



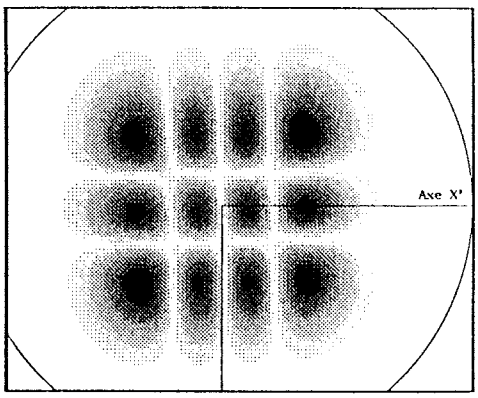
T = 200.00°C, F = 8818614.0 Hz,
v1= 0.999, v2 =0.018, v3= 0.041

Fig. 3

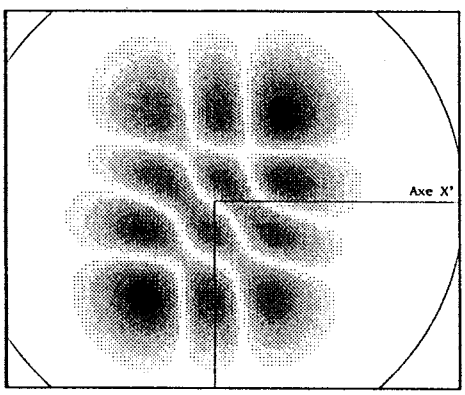
predicted v_3 component of (C,7,0,0) mode : AT cut : θ=35°27' $2h_0$=1.327mm R=200.0mm
v_1 component pattern has the usual shape all over the temperature range.



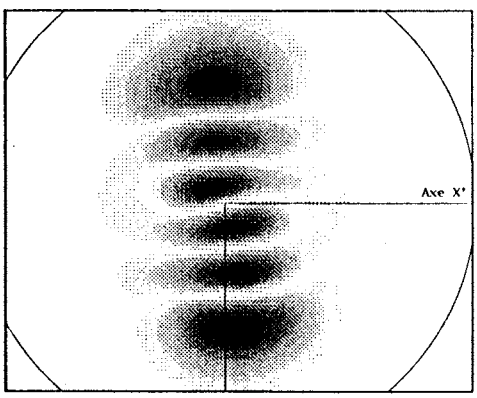
T = 0.00°C, F = 8921053.0 Hz
v1= 0.980, v2= 0.061, v3= 0.190



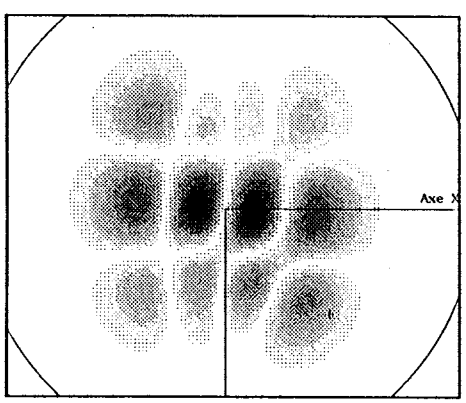
T = 35.00, F = 8922895.0 Hz
v1= 0.747, v2= 0.160, v3= 0.645
maximum coupling with (B,6,3,2) mode
agrees with Fig 10b) Ref[13]



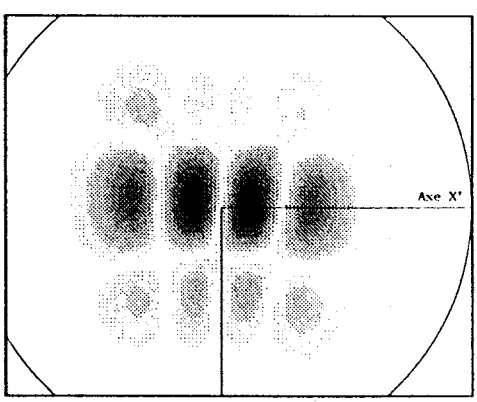
T = 65.00°C, F = 8921236.0 Hz
v1= 0.937, v2= 0.093, v3= 0.337



T = 100.00°C, F = 8921495.0 Hz
v1= 0.828, v2= 0.195, v3= 0.525
strong coupling with (B,6,0,5)



T = 120.00°C, F = 8921684.0 Hz,
v1= 0.991, v2= 0.059, v3= 0.117



T = 200.00°C, F = 8925247.0 Hz
v1= 0.994, v2= 0.056, v3= 0.097

Fig. 4

predicted v_3 component of (C,7,0,0) mode : AT cut : 0=35°27' $2h_0$=1.327mm R=200.0mm
v_1 component pattern has the usual shape all over the temperature range.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 5A Number 29

ENHANCED-SELECTIVITY, INTERMEDIATE-BANDWIDTH, CRYSTAL FILTERS

P. K. Wall

Peter Wall Electroacoustics U.K.

ABSTRACT

In existing intermediate bandwidth crystal filters, inductor-capacitor tank circuits tuned to the centre frequency have been added to narrow band crystal filter networks to absorb unwelcome crystal shunt capacitances.

An alternative type of design, which we shall call Method A, consists of embedding a narrow band filter (with elements unrealizable by using crystals), between suitable coil-capacitor networks. Part of these networks can be absorbed into the narrow band filter to produce a new network realizable with piezoelectric elements. The skirts of the filter can be improved by several decibels by this means. The design can be tailored to the "Q" and stability of the inductors which the designer wishes to use. Susceptibility to unwanted responses of the piezoelectric devices is reduced significantly.

A second design method, which we shall call Method B, enables steep selectivity crystal filters with transmission zeroes, to be designed in a simple manner from a low pass parent network. With care, these filters can be extended well into the intermediate bandwidth region.

Some computer generated examples are given.

INTRODUCTION

R.G.Kinsman in Ref.1 & 2 describes the addition of parallel tuned circuits, tuned at the centre of the filter, to provide sufficient circuit capacitances to represent the crystal shunt capacitances. These are used at the input and output of the filter and at the junctions between sections, each of two crystals. He also uses inductors to represent negative capacitances in "T" or"TT" impedance invertors. Some of these networks result in crystal frequencies which are symmetrical with respect to the centre of the filter.

Intermediate bandwidth crystal filters are defined as those whose bandwidth is greater than that which is realizable by use of crystals and capacitors only. This applies to a succession of filter sections which may be ladders or lattices according to the circumstances. Lattices are represented in practical filters by hybrid coils configurations to economize in crystals.

The bandwidth limitation is caused by the well known minimum capacitance ratio of the crystal elements.

Wideband techniques exist in which inductors are inherently part of the design. These nable filters with bandwidths up to about 5% to be designed. They necessiate accurate inductance values as well as accurate tuning. As the bandwidth decreases, the insertion loss of these filters increases due to the losses in those added inductors. This is the reason why alternative design methods are needed for filters of intermediate bandwidth.

The new design methods provide alternative intermediate bandwidth filter configurations in which the crystal series resonant frequencies are symmetrical relative to the centre frequency of the filter and typically close to the cutoffs. There are two basis benefits:

Component errors or circuit misalignments have little effect on the position of the cutoffs.

Some of the effects of tolerances are approximately symmetrical with respect to the centre frequency.

These result in better control of the characteristics of filters with steep transition bands.

DESIGN METHOD A

We start with a lattice network consisting of series resonant circuits having no shunt capacitances at all. This is of course unrealizable with piezoelectric elements. There are 2 basic ways to derive such lattices to meet specific requiremnts:

Directly in lattice form using Insertion Loss techniques.
(Zverev Ref.3 pp 483-486 reports on the synthesis technique by D.S. Humphreys.)

From Low Pass Networks

The example in Fig.1 corresponds to a 4th order Chebychev low pass filter for which there are two distinct types of passband ripple. Fig.2A shows the band pass form which is then transformed into 2 halves separated by an impedance invertor. This is a well-known procedure. This network turns out to be symmetrical and therefore realizable in lattice form. Chebychev and maximally flat filters conform, but some other types such as Gaussian and Legendre types do not become symmetrical in this way.

Using Bartlett's Bisection Theorem (Fig.2B), the network finally becomes that in Fig 3. The position of the crystal series resonant frequencies is shown in Fig. 4 for a 10 MHz filter with a 3 dB bandwidth of 105 kHz total. They have symmetry relative to the centre frequency of the filter.

It is possible to add L-C circuits tuned to the centre frequency at input and output in order to provide shunt capacitances across the series resonant arms and thereby make it possible to realize them as crystal elements. If the filter is quite wide, say 1% or 2%, the inductors can cause a small curling of the edges.

An alternative way is to use an half section of L-C band pass network at each end as shown in Fig. 5. The bandwidth of these half sections can be chosen by the designer but will certainly be several times the bandwidth of the crystal section. The advantage of this arrangement is that the impedance facing the crystal section can be made very closely equal to the terminating resistances even for low values of shunt end inductors. The series resonant arms of the end sections can now be absorbed into the centre crystal section, so these series elements have the same "Q"and stability as the crystals. The convenient place to do this operation is at the stage of transforming from Fig. 2A to Fig. 3.

Better impedance matching can be provided by the use of full bandpass sections as shown in Fig. 6 which also provides means for avoiding transformers with 3 windings.

Conversion into single sections of crystals of even order lowpass filters with finite positions of peaks must start from a lowpass characteristic with finite loss at infinity. (Ref. 4 deals with transformations from lowpass networks containing peaks to bandpass form but results in tandem sections of 2 crystals each.)

A comparison of the passband and stopband characteristics of the 10 MHz Chebychev filter of +/_ 52.5 kHz bandwidth with and without the end L-C sections is shown in Figs. 7 and 8. Fig. 7A and 7B are without any LC half sections and Fig.8A and 8B with half sections absorbed into the lattice. In the stopband, the same spurious response has been included in both cases to demonstrate that filter sensitivity to unwanted responses is reduced by the addition of the half sections. This spurious resonance was arbitrarily chosen to have 5 Henries inductance and to be at 10.2 MHz.

Inspection of the responses shows a negligible change to the passband. The new filter stopband has increased by about 12 dB at -0.5 MHz and the spurious response is 3 dB less severe. This benefit increases for spurious resonances further from the passband. The circuit elements with added L-C sections are shown in Fig. 9 and a comparison of the crystal series resonant frequencies without and with the half sections is given in Fig. 10. It will be seen that the frequencies have shrunk towards the centre of the filter.

It will be noted that none of the crystals are at the centre frequency. This is because we started with an even order lowpass. Fig. 11 shows a 5th order network having two identical finite peaks. One of the crystals in the final circuit is at the centre frequency. (fB2 = fc, the centre frequency.)

It can be rightly claimed that design with only one section has limited use due to the problems of various element tolerances and unwanted responses in the crystals. There are two basic ways to achieve a higher stopband attenuation from this approach:-

Use of 2 or more sections separated by small attenuator pads. The sections do not need to be identical.

Use of Design Method B, which will now be described.

DESIGN METHOD B

This starts with a suitable low pass filter in minimum inductor configuration. A 5th order network configuration with 2 identical finite peaks is chosen as an example. The low pass to band pass transformation is shown in Fig. 12. The network is broken into 2 halves, each of which can in fact be made symmetrical under certain practical conditions. This is true for certain insertion loss designs as well as for image parameter designs. This is explained later.

Assuming this, each half can be converted into a lattice with 3 series resonant circuits, one at the centre frequency and the other two just outside the upper and lower cutoffs of the filter.

The attenuation characteristics for this example, without any tank circuits, are shown in Figs. 13 and 14. The circuit including end circuits and a damped centre tuned circuit is shown in Fig.19. The end and centre tuned circuits are added in order provide static capacitances Co for the crystals. The characteristics under these conditions are those in Fig. 15 and 16.

If damping had not been added to the centre tuned circuit, the characteristics would have been as in Fig.17 and 18.

Very little difference exists in the passband between the 3 sets of characteristcics. A considerable increase in the stopband outside the peaks is offset by some loss on the slopes where the resonance in the undamped case occurs. This resonance is inevitable because the reactance of the centre tuned circuit drops and the reactance of opposite sign of the two filter sections facing it, increases steadily as the frequency becomes more remote from the centre.

These crystals have an inductance rather more than twice that of the one at centre frequency.

The lower the shunt inductance of the centre section, the closer the resonances come to the passband, and the more difficult they become to deal with. Practical values can be obtained as Fig. 19 shows.

Use of damping on the centre tuned circuit is one way to suppress the unwanted dips. Another method of suppression is to position the design attenuation peaks at the same place as the dips caused by the centre coil.

Another example of Design Method B is provided by a 7th order network. The characteristics of the filter before addition of shunt tuned circuits at input, output and the two internal junctions are shown in Fig. 20A, B and C. The characteristics with added tuned circuits are shown in Fig.21 A,B and C. The practical network configuration is shown in Fig. 22, derived by the means shown in Fig. 26.

DESIGN of SYMMETRICAL NETWORKS in SYMMETRICAL COMPONENT SECTIONS

Earlier in the paper, the possibility was mentioned of converting a symmetrical network into two tandem circuits which themselves are symmetrical and identical. This is of concern for 5th and 9th order networks but not for 7th order. Fig 23 shows what happens to the inductance ratio L3/L1 as the ripple changes on a 5th order LP filter. For two equal finite peaks, a suitable ripple can also be found in many of these cases.

In the case of the 7th order, the internal inductances do not need to be twice the end inductances because the join between the middle section and the end two sections can always be selected to make all 3 sub-networks symmetrical. However it is a convenience if the 2:1 ratio is achieved because this minimizes the number of different values of crystal inductance which are necessary with this design method. Fig. 24 shows what happens in the 7th order Chebychev case. The procedure for dealing with the 7th order transformation to a crystal filter configuration is given in Fig. 26.

It is sometimes thought that filters with very low ripple are inefficient in terms of rate of cutoff. If the filters are compared with the 3 dB points made the same instead of the points defining the end of the controlled passband ripple, the difference is much less, as demonstrated in Fig. 25. This also shows the benefit of using controlled peaks in the stopband.

The 9th order case does not appear to give any practical combinations of peak and ripple conditions for achieving the objective of a succession of symmetrical networks. Investigations are continuing.

SUMMARY

Methods A and B above provide alternative means for obtaining one or more crystal sections with crystal frequencies which are symmetrical relative to the centre of the filter. They provide alternative configurations which have value for specific intermediate bamndwidth filter requirements.

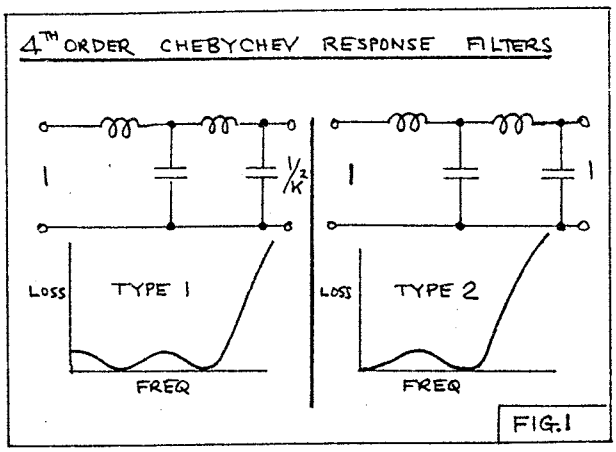
REFERENCES

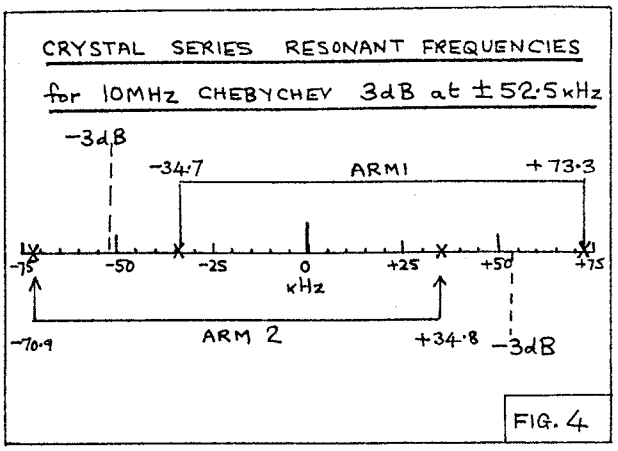
(1) Robert G. Kinsman, "Crystal Filters: Design, Manufacture & Application" John Wiley & Sons 1987.

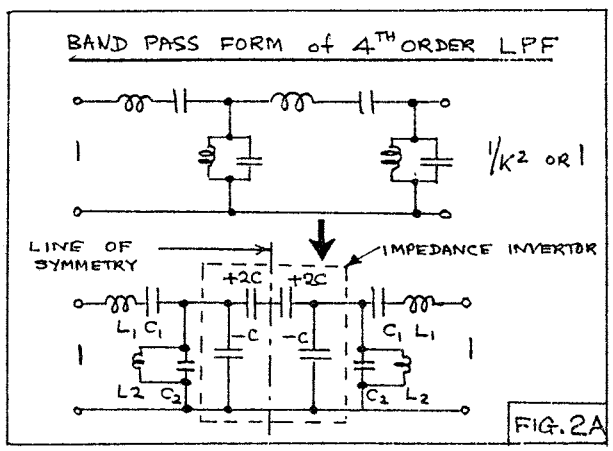
(2) Robert G. Kinsman, "Design of Crystal Filters with Transmission Zeroes", Proc. 5th Quartz Crystal Conference, August 1983.

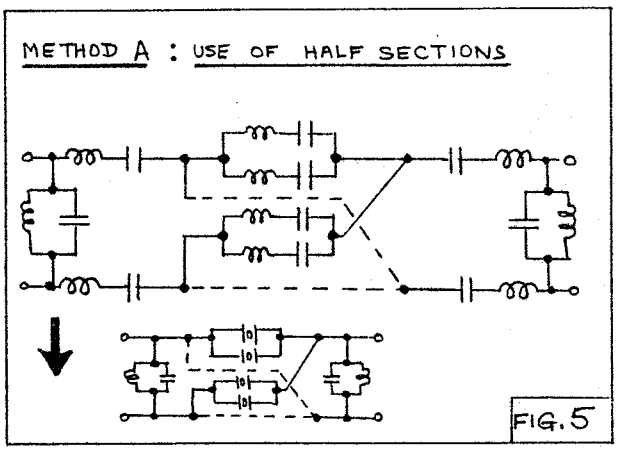
(3) Anatol I. Zverev, "Handbook of Filter Synthesis". John Wiley & Sons 1967

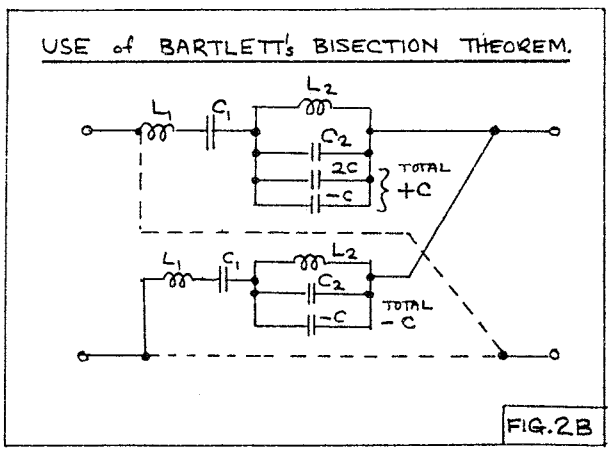
(4) Alan G.J. Holt & R.L. Gray "Bandpass Crystal Filters by Transformation of a LOWpass Ladder". IEEE Trans. on Circuit Theory Vol. CT-15 No. 4 Dec 1968 pp 492-494.

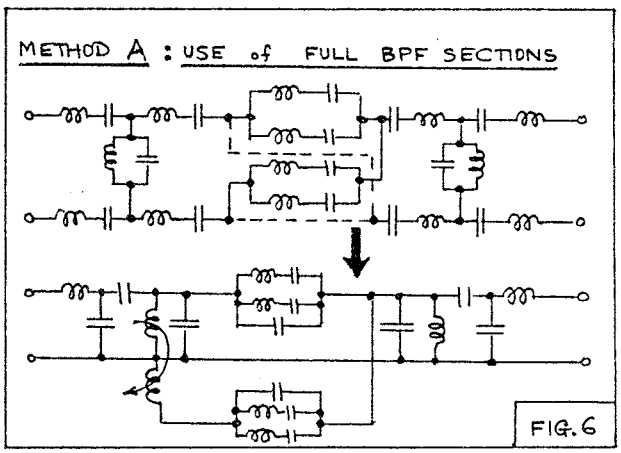

4TH ORDER CHEBYCHEV RESPONSE FILTERS
1
1/K
1/2
LOSS
TYPE 1
FREQ
LOSS
TYPE 2
FREQ
FIG.1

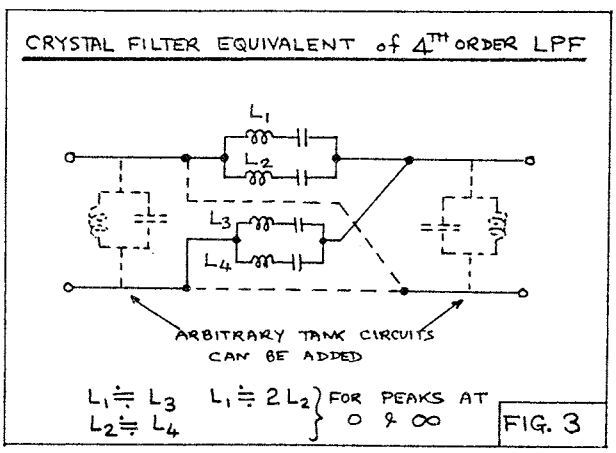

CRYSTAL SERIES RESONANT FREQUENCIES
for 10MHz CHEBYCHEV 3dB at ±52·5kHz
-3dB
-34·7
ARM1
+73·3
-75
-50
-25
0
+25
+50
+75
kHz
-70·9
ARM 2
+34·8
-3dB
FIG. 4


BAND PASS FORM of 4TH ORDER LPF
1/K2 OR 1
LINE OF SYMMETRY
IMPEDANCE INVERTOR
+2C
+2C
-C
-C
L1 C1
C1 L1
L2 C2
C2 L2
FIG.2A


METHOD A : USE OF HALF SECTIONS
FIG.5


USE of BARTLETT's BISECTION THEOREM.
L1
C1
L2
C2
2C
-C
TOTAL
+C
L1
C1
L2
C2
-C
TOTAL
-C
FIG.2B


METHOD A : USE of FULL BPF SECTIONS
FIG.6


CRYSTAL FILTER EQUIVALENT of 4TH ORDER LPF
L1
L2
L3
L4
ARBITRARY TANK CIRCUITS CAN BE ADDED
L1 ≒ L3
L2 ≒ L4
L1 ≒ 2L2
FOR PEAKS AT 0 & ∞
FIG. 3

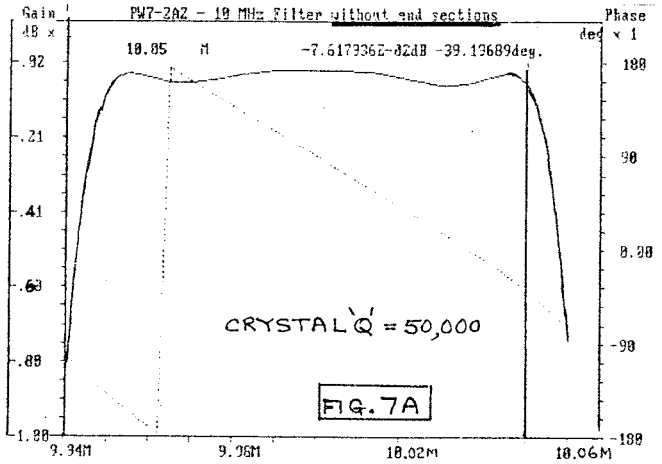


FIG. 7A

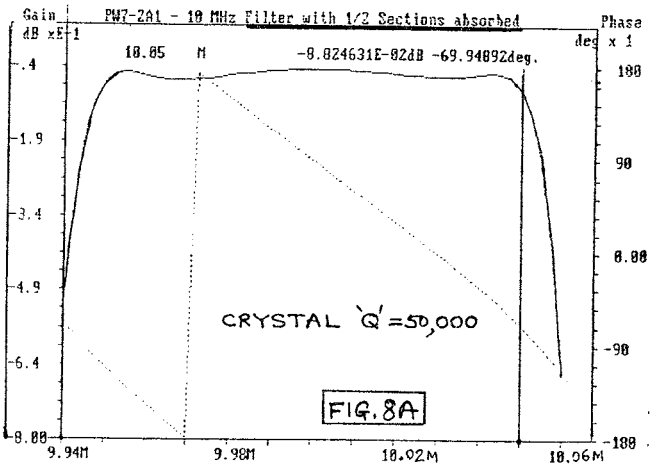


FIG. 8A

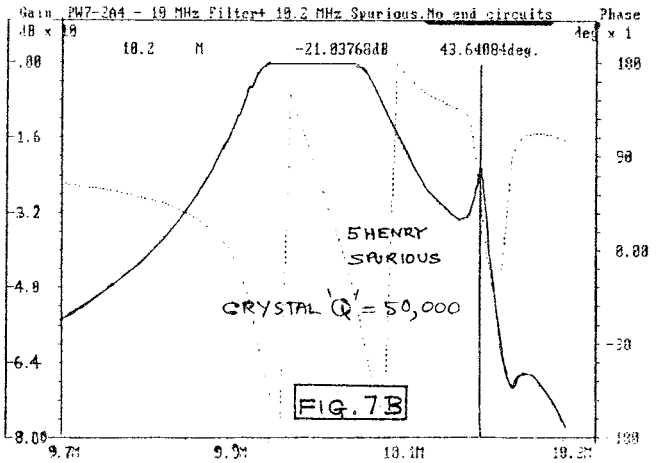


FIG. 7B

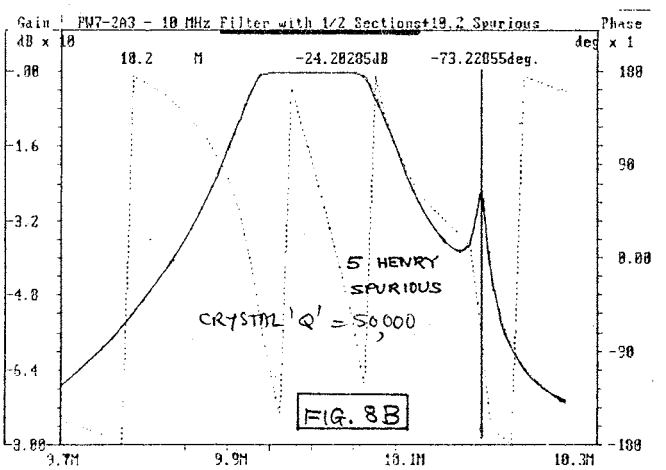
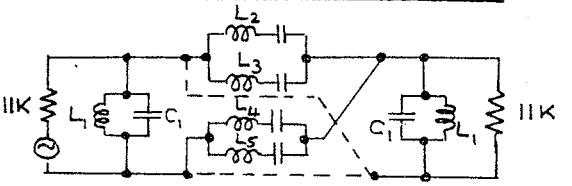


FIG. 8B

CIRCUIT DETAILS for 10MHz FILTER



L_1	6·33 µH	C_1	40 pF
L_2	56·68 µH	FREQ^Y	9·934 MHz
L_3	24·24 µH	"	10·029 "
L_4	56·32 µH	"	10·067 "
L_5	24·30 µH	"	9·971 "

FIG. 9

CRYSTAL SER. RES. FREQUENCIES of 10MHz FILTER

3dB ±52·5 kHz

FILTER ARM	CRYSTAL	FREQUENCIES kHz REF 10MHz (TO NEAREST kHz)	
		WITHOUT ANY END CIRCUITS	WITH 1/2 SECTIONS ABSORBED
1	1	+35	+29
	2	−71	−66
2	1	−35	−29
	2	+73	+67

FIG. 10.

TRANSFORMATION of 5TH ORDER—IDENTICAL PEAKS

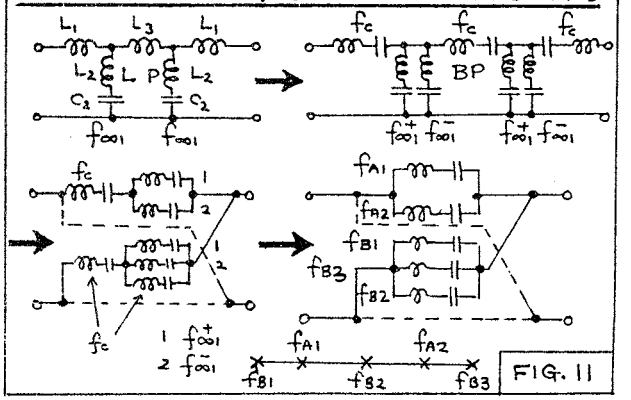


FIG. 11

METHOD B : 5TH ORDER NETWORK

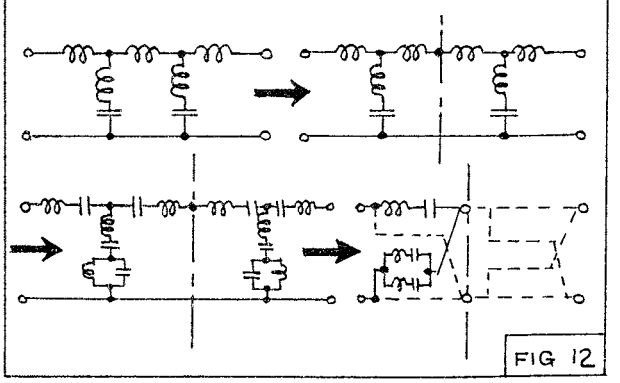
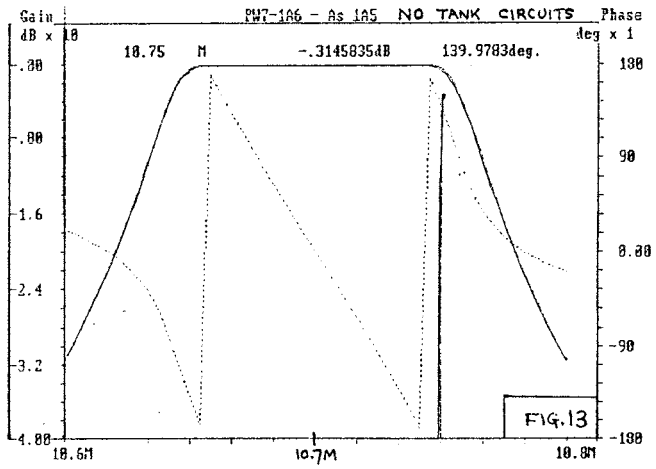
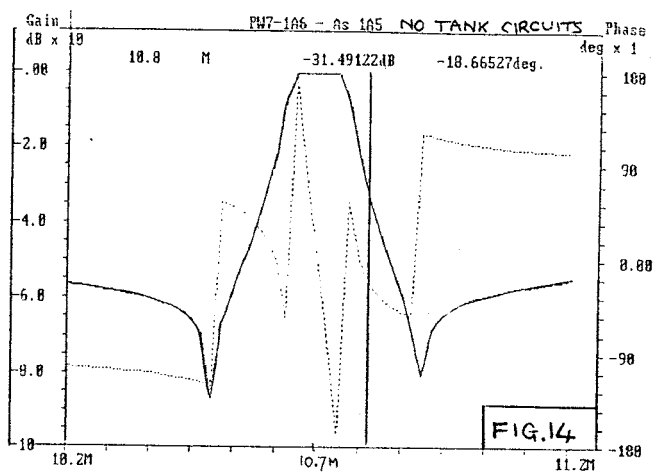
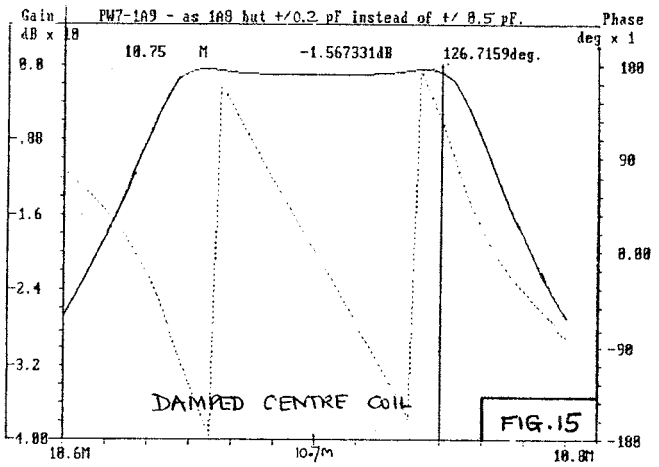
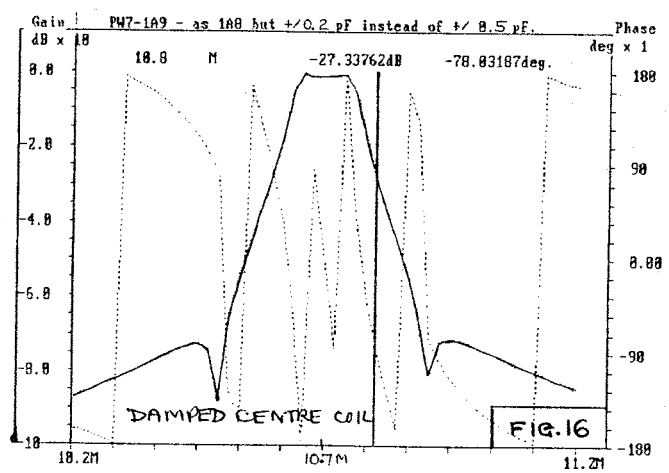


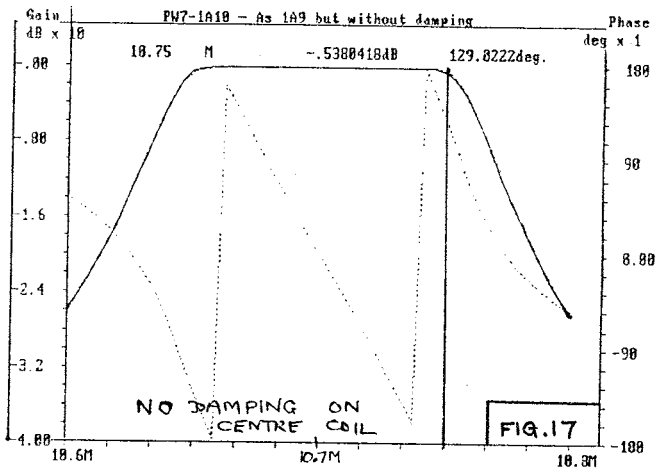
FIG 12

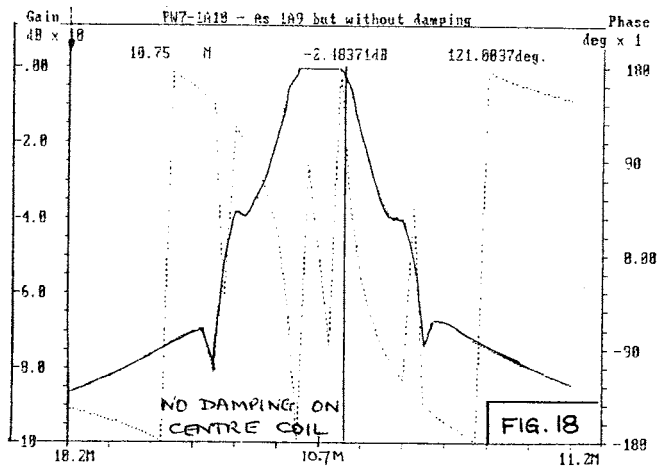

PW7-1A6 - As 1A5 NO TANK CIRCUITS
Gain dB x 10
Phase deg x 1
10.75 M
-.3145835dB
139.9783deg.
10.6M
10.7M
10.8M
FIG.13

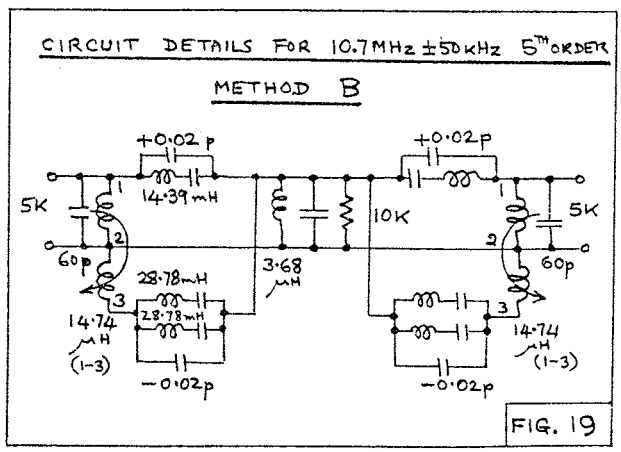

PW7-1A6 - As 1A5 NO TANK CIRCUITS
Gain dB x 10
Phase deg x 1
10.8 M
-31.49122dB
-18.66527deg.
10.2M
10.7M
11.2M
FIG.14


PW7-1A9 - as 1A8 but +/0.2 pF instead of +/ 0.5 pF.
Gain dB x 10
Phase deg x 1
10.75 M
-1.567331dB
126.7159deg.
DAMPED CENTRE COIL
10.6M
10.7M
10.8M
FIG.15


PW7-1A9 - as 1A8 but +/0.2 pF instead of +/ 0.5 pF.
Gain dB x 10
Phase deg x 1
10.8 M
-27.33762dB
-78.83187deg.
DAMPED CENTRE COIL
10.2M
10.7M
11.2M
FIG.16


PW7-1A10 - As 1A9 but without damping
Gain dB x 10
Phase deg x 1
10.75 M
-.5380418dB
129.8222deg.
NO DAMPING ON CENTRE COIL
10.6M
10.7M
10.8M
FIG.17


PW7-1A10 - As 1A9 but without damping
Gain dB x 10
Phase deg x 1
10.75 M
-2.183714B
121.8837deg.
NO DAMPING ON CENTRE COIL
10.2M
10.7M
11.2M
FIG. 18


CIRCUIT DETAILS FOR 10.7MHz ±50kHz 5TH ORDER
METHOD B
+0·02p
+0·02p
5K
14·39mH
10K
5K
60p
60p
28·78mH
3·68 mH
14·74 mH (1-3)
28·78mH
14·74 mH (1-3)
-0·02p
-0·02p
FIG. 19

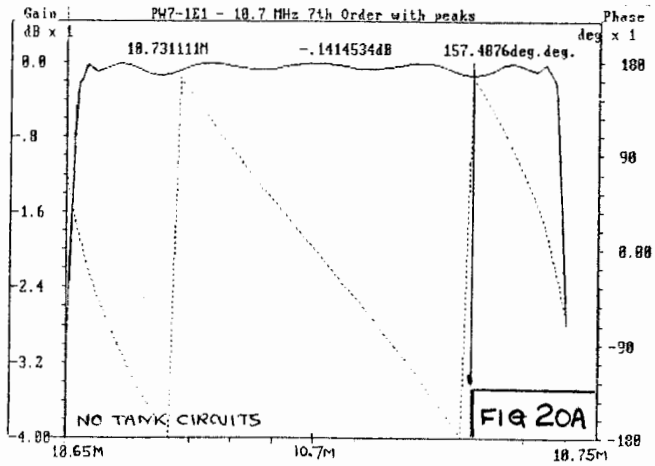


FIG 20A

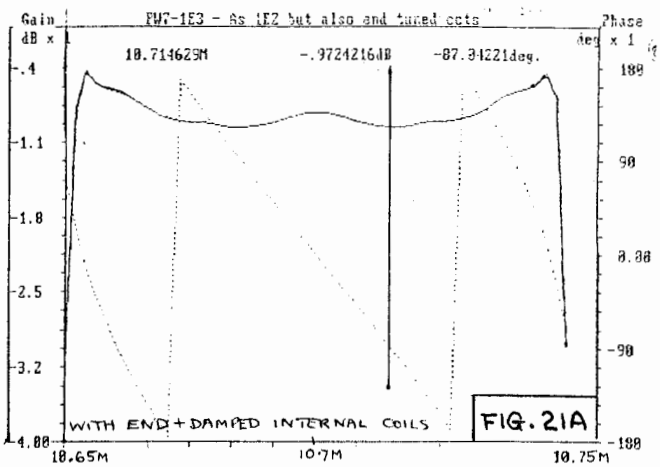


FIG. 21A

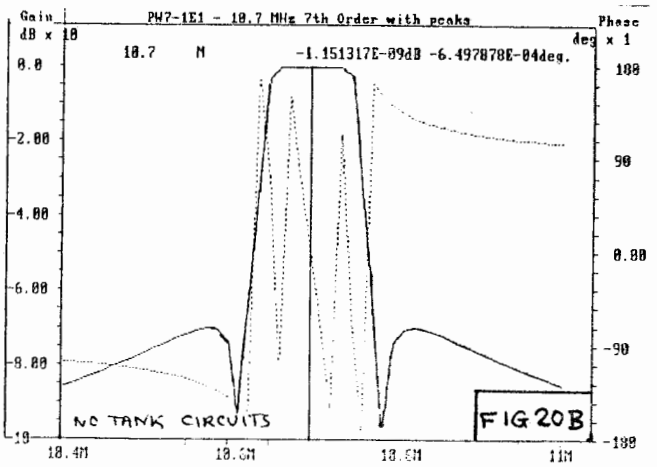


FIG 20B

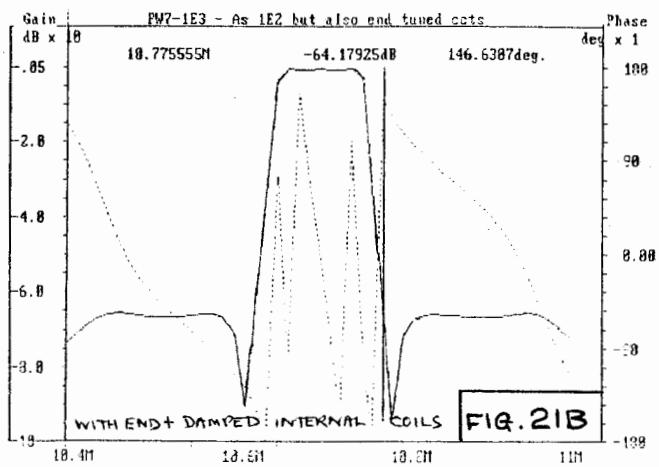


FIG. 21B

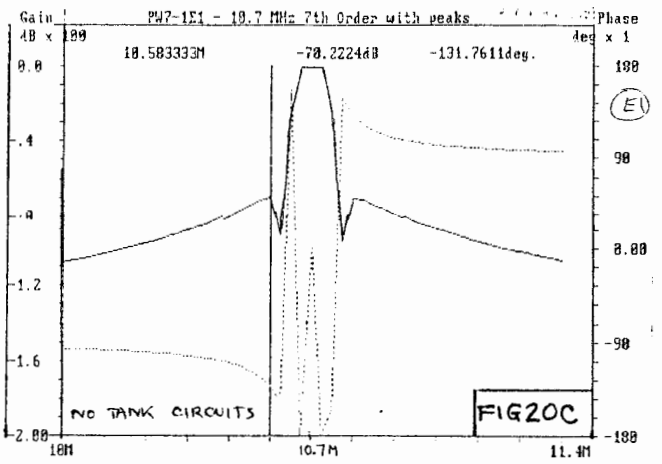


FIG20C

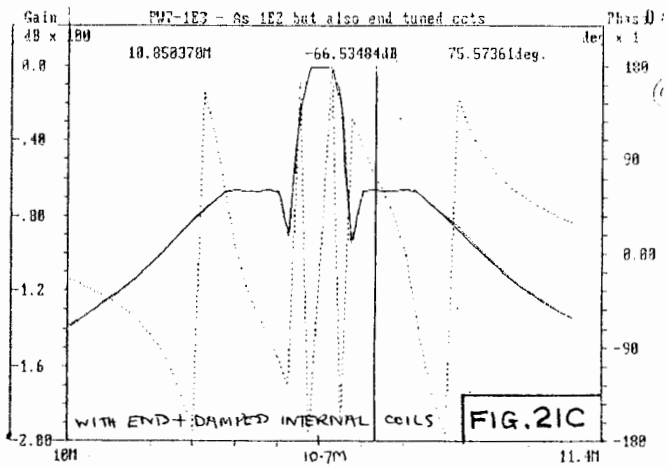


FIG. 21C

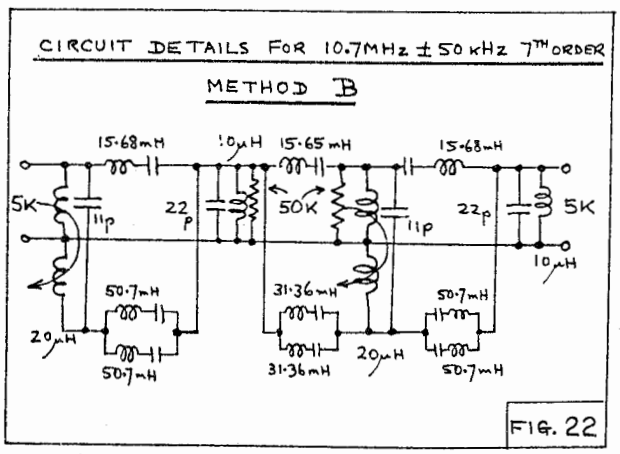


FIG. 22

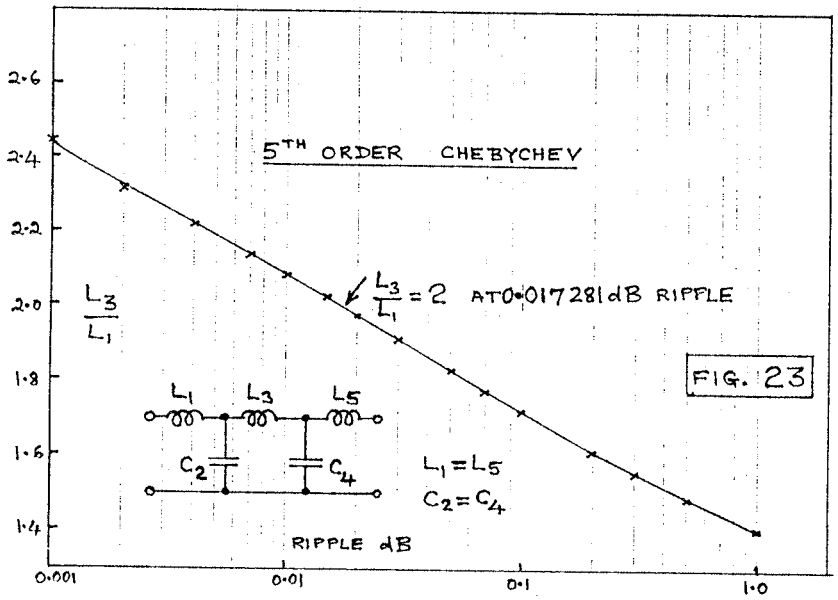


FIG. 23

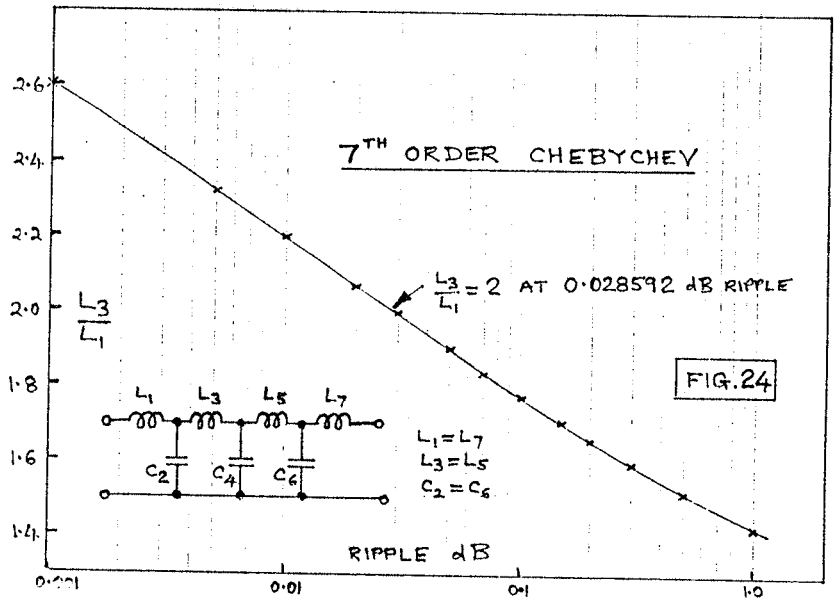


FIG. 24

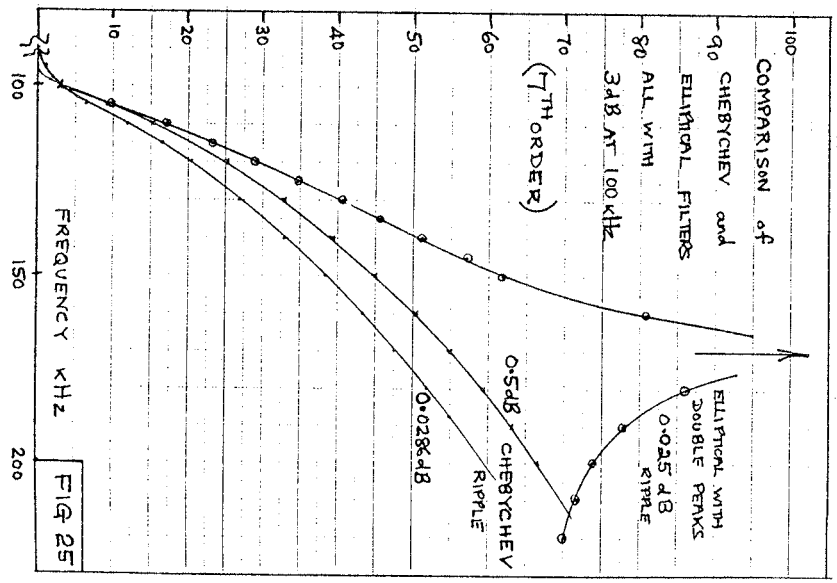


FIG 25

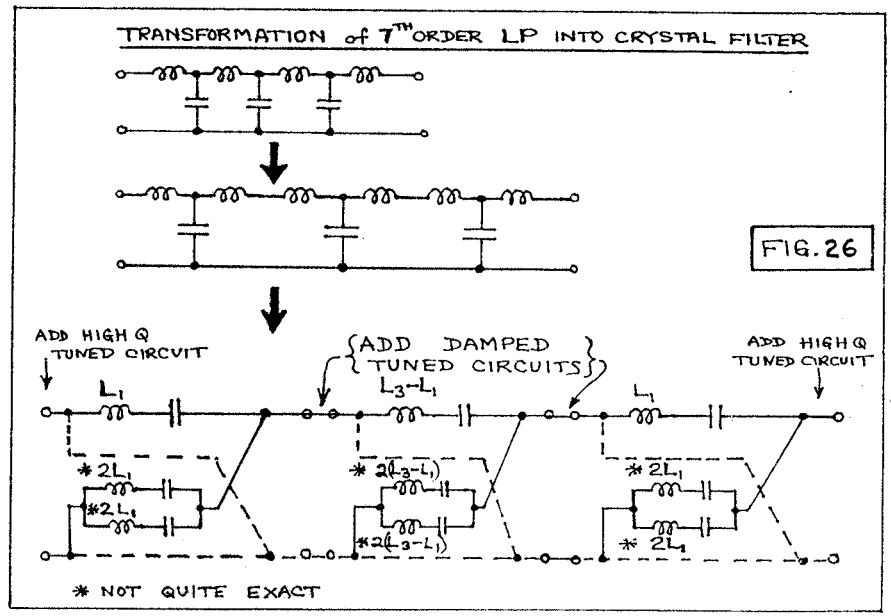


FIG. 26

SESSION 4B

ATOMIC CLOCKS II

Chairman : N. KOSHELYAEVSKY

Session 4B Number 30

Incoherent and coherent optical pumping in Rubidium standards

Philippe Willemin, Giovanni Busca
Observatoire de Neuchâtel, CH-2000 Neuchâtel/Switzerland

Jin Ting Liu[1], Pierre Thomann[2]
Oscilloquartz SA, CH-2000 Neuchâtel/Switzerland

Abstract

Experiments in a classical rubidium standard using several different isotopic mixtures in the lamp and different mixtures of buffer gas in the cell have suggested a phenomenological model of the distributed light shift which is able to predict the cell temperature for zero light shift operation and for zero thermal coefficient. This model reveals a basic limitation of the conventional optical pumping.

Preliminary studies of laser pumping in rubidium have shown improvement in the control of the light shift but have also shown some new problems associated with frequency and amplitude noise of the laser itself. A measured short-term stability of $2\cdot10^{-11}\ \tau^{-1/2}$ is reported in the range τ=1s to 100s.

1. Conventional Rubidium Standards

1.1. Phenomenological model of the distributed light shift

Models for predicting signal, noise and linewidth are available [1]. However, only crude calculations exist concerning the output frequency [2]. The basis of this calculation is a linearization of the equation of the isotopic-filter-induced optical displacement $\Delta\nu_{op}$ given in ref [3] and a linearization of the equation for light absorption.

The light-shift $\Delta\nu_{LS}$ experienced by the first layer of atoms in the absorption cell can be expressed, with small changes from ref [2], as

$$\Delta\nu_{LS} = I_{in}\,[1-K_0(\theta-\theta_0)]\,K_1(\theta-\theta_0) \qquad (1)$$

where θ is the filter temperature
I_{in} is the light intensity at the input of the filter,
$I_{in}\,[1-K_0\,(\theta-\theta_0)]$ the light intensity at the output of the filter and entering the rubidium cell,
K_1 a light shift parameter depending upon the spectral shape of the pumping line and
θ_0 the filter temperature for which the light shift goes to zero.

With the same meaning of the parameters, considering a mixture of isotopes in the lamp and, for the integrated filter approach, a mixture of isotopes in the cell, we have the general equation giving the frequency vs light intensity and filter cell temperature:

$$\nu = \nu_0 - I_{in}^- \left[1-K_0^-(\theta-\theta_0^-)\right] K_1^- (\theta-\theta_0^-) + I_{in}^+ \left[1-K_0^+ (\theta-\theta_0^-)\right] K_1^+ (\theta+\theta_0^+) + \delta_m(\theta-\theta_0^-) + \gamma_m(\theta-\theta_0^-)^2 \qquad (2)$$

where
ν_0 is the unperturbed frequency (independent from the temperature and light intensity)
the index - refers to Rb^{87} isotope
the index + refers to Rb^{85} isotope
θ is the filter-cell temperature
δ_m and γ_m are respectively the linear and quadratic temperature dependence given by the buffer gas mixture [2].

Equation (2) expresses the fact that Rb^{87} causes, for normally used cell temperatures > 65°C, a negative light shift; Rb^{85} causes a positive light shift which never goes to zero.

The parabolic behaviour given in eq. (2) has a minimum for a temperature $\theta = \theta_{min}$ given by

$$\theta_{min} - \theta_0^- = \frac{1 + \dfrac{K_0^+(\theta_0^+-\theta_0^-)}{P_{in}-1} - \dfrac{\delta_m}{I_{in}[\alpha_{in}K_1^- - (1-\alpha_{in})K_1^+]}}{\dfrac{2K_0^-}{P_{in}-1}\left(P_{in}-\dfrac{K_0^+}{K_0^-}\right) + \dfrac{2\gamma_m}{I_{in}[\alpha_{in}K_1^- - (1-\alpha_{in})K_1^+]}} \qquad (3)$$

where

$$I_{in} = I_{in}^- + I_{in}^+$$
$$\alpha_{in} = I_{in}^-/I_{in}$$
$$(1-\alpha_{in}) = I_{in}^+/I_{in} \qquad (4)$$
$$P_{in} = \alpha_{in}K_1^-/(1-\alpha_{in})K_1^+$$

P_{in} represents the ratio (absolute value) of the negative to the positive light shift, the index "in" meaning that the ratio is taken at the input of the filter-cell.

Equation (3) shows that θ_{min} depends upon the total light intensity I_{in} emitted by the lamp if δ_m and γ_m are different from zero.

Equation (2) can be expressed in terms of a measurable output parameter, i.e. light intensity I_{out} transmitted by the cell.

$$\nu = \nu_0 - \alpha_{out}\,I_{out}\,K_1^-(\theta-\theta_0^-) + (1-\alpha_{out})\,I_{out}\,K_1^+\,(\theta+\theta_0^+) + \delta_m(\theta-\theta_0^-) + \gamma_m\,(\theta-\theta_0^-)^2 \qquad (5)$$

where

$$I^-_{in}\left[1 - K^-_0\,(\theta-\theta^-_0)\right] = I^-_{out}$$

$$I^+_{in}\left[1 - K^+_0\,(\theta-\theta^-_0)\right] = I^+_{out}$$

$$I_{out} = I^-_{out} + I^+_{out} \qquad (6)$$

$$I^-_{out} = \alpha_{out}\, I_{out}$$

$$I^+_{out} = (1-\alpha_{out})\, I_{out}$$

The index "out" refers to transmitted intensities. From eq. (6), the value of I^+_{in} or I^-_{in} are just the value of the transmitted light intensities I^+_{out} or I^-_{out} for a cell temperature of θ^-_0.

Equation (5) states that the frequency is a linear function of the total transmitted light, in agreement with experiments [2].

By differentiating eq. (5) with respect to the transmitted light intensity I_{out} one obtains that $\partial\nu/\partial I_{out} = 0$ for $\theta = \theta_{LS=0}$ given by:

$$\theta_{LS=0} - \theta^-_0 = \frac{\theta^+_0 + \theta^-_0}{p_{out}^{-1}} \qquad (7)$$

where

$$p_{out} = \frac{\alpha_{out} K^-_1}{(1-\alpha_{out}) K^+_1} \qquad (8)$$

p_{out} has the same meaning as p_{in} but referred now to the transmitted total light intensity.

For obtaining good standard performances

$$\theta_{min} = \theta_{LS=0} \qquad (9)$$

this relation can be realised by choosing the isotopic ratio in the lamp and in the cell and in addition the buffer gas mixture.

For one commercial rubidium standard this occurs at a filter-cell temperature of 75°C ([2], page 1318).

The model shows that the main limitation of the conventional rubidium standard is that zero temperature coefficient and zero light shift occurs respectively for a unique light intensity and a unique filter-cell temperature.

In addition, the previous model can enlight the mechanism of the rubidium drift.

1.2. Typical experimental values of the constants

From experiments in our laboratory we found that, for a natural rubidium cell, $K^-_0 = K^+_0 = K_0$. In this condition the experimental values of the optical parameters are:

$$\theta^-_0 = 68°C \qquad \theta^+_0 = 45°C$$
$$K_0 = 0.025/K \qquad (10)$$
$$K^+_1 = 0.045\ Hz/\mu A\cdot K \qquad K^-_1 = 0.07\ Hz/\mu A\cdot K$$

The parameter α_{out} depends upon the lamp isotopic composition. For a lamp with a mixture of 50% ^{87}Rb and 50% ^{nat}Rb, $\alpha_{out} = 0,87$.

These values which are certainly influenced to some degree by the particular experimental set-up used, can be applied successfully to other experimental situations as shown below.

1.3. Application of the theory to experimental data

We consider the experimental data of ref [2], page 1318, referring to the following experimental situation: integrated filter approach, lamp with 50% of ^{87}Rb and 50% of ^{nat}Rb, natural rubidium in the cell and pure nitrogen buffer gas at a pressure of 9.5 Torr.

1.3.1. Temperature coefficient of the light-shift

From the data of ref [2], we can plot the function $\frac{\partial\nu}{\partial I_{out}}$ (light-shift) versus the filter cell temperature θ.

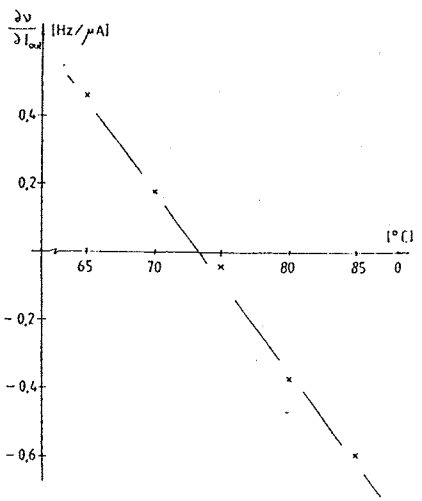


Fig.1

From fig. 1 the value of the slope $\frac{\partial}{\partial\theta}\left(\frac{\partial\nu}{\partial I_{out}}\right)$ is:

$$\left[\frac{\partial}{\partial\theta}\left(\frac{\partial\nu}{\partial I_{out}}\right)\right]_{exp.} = -0.053\ \text{Hz}/\mu\text{A}\cdot\text{K} \qquad (11)$$

On the other hand from eq. (5) of the model we have the theoretical value

$$\left[\frac{\partial}{\partial\theta}\left(\frac{\partial\nu}{\partial I_{out}}\right)\right]_{th.} = (1-\alpha_{out})\,K_1^+ - \alpha_{out}\,K_1^- \qquad (12)$$

Inserting into eq. (12) the values of the relevant parameters given in eq. (10) and the value of $\alpha_{out} = 0.87$ corresponding to the case of interest here, we obtain:

$$\left[\frac{\partial}{\partial\theta}\left(\frac{\partial\nu}{\partial I_{out}}\right)\right]_{th.} = -0.055\ \text{Hz}/\mu\text{A}\cdot\text{K} \qquad (13)$$

Which is in good agreement with the experimental value.

1.3.2. Filter-cell temperature for zero light-shift

We compare the experimental and theoretical values of $\theta_{LS=0}$.

From fig. 1 we have:

$$\left[\theta_{LS=0}\right]_{exp.} = 74^\circ\text{C} \qquad (14)$$

Equation (7), using the typical values given in eq. (10) and $\alpha_{out} = 0.87$, gives

$$\left[\theta_{LS=0}\right]_{th.} = 80^\circ\text{C}. \qquad (15)$$

Considering the number of parameters involved in eq. (7) and the fact that the light spectrum depends upon the particular experimental set-up used, the comparison between the experimental and the theoretical value of $\theta_{LS=0}$ can be considered satisfactory.

1.3.3. Filter-cell temperature for zero temperature coefficient

In order to compare the theoretical and the experimental values of θ_{min}, we need the precise knowledge of the parameter $K_0 = K_0^- = K_0^+$ in the particular case considered here. From eq. (6), with the assumption $K_0 = K_0^- = K_0^+$ we get:

$$\left(I_{in}^- + I_{in}^+\right)\left[1 - K_0\,(\theta-\theta_0^-)\right] = I_{out}^- + I_{out}^+ \qquad (16)$$

or

$$I_{in}\left[1 - K_0\,(\theta-\theta_0^-)\right] = I_{out} \qquad (17)$$

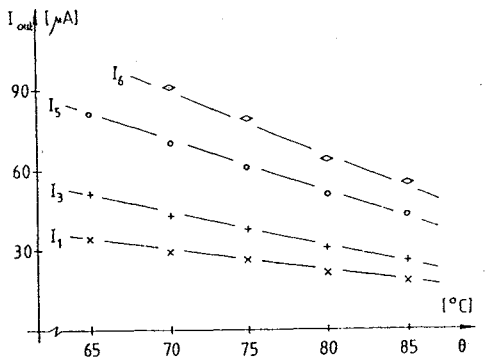


Fig 2

From eq. (17) the experimental value of K_0 is the slope $\frac{\partial I_{out}}{\partial\theta}$ divided by I_{in}, which is the value of I_{out} for $\theta=\theta_0^-$. From data of ref [2], let us plot the graph I_{out} vs θ for the different light intensities I_j impinging on the filter. The mean value of K_0 obtained from the four different light intensities is

$$\left[K_0\right]_{exp.} = 0.026/\text{K}. \qquad (18)$$

This value is comparable to the typical experimental value measured in our laboratory and reported in eq. (10).

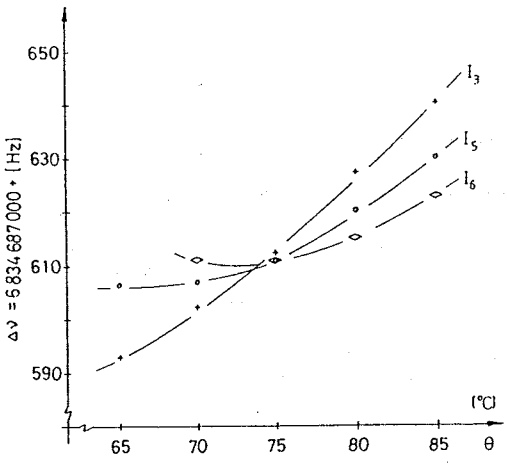


Fig 3

Let us consider now the experimental values of θ_{min}. From fig. 3 we have for the different light intensities:

$$\begin{aligned} I_3 &: \left[\theta_{min}\right]_{exp.} \ll 65^\circ\text{C} \\ I_5 &: \left[\theta_{min}\right]_{exp.} \simeq 65^\circ\text{C} \\ I_6 &: \left[\theta_{min}\right]_{exp.} \simeq 72^\circ\text{C} \end{aligned} \qquad (19)$$

Considering now the theoretical value of θ_{min}. From eq. (3) with the assumption that $K_0 = K_0^- = K_0^+$ and $\gamma_m = 0$, we have:

$$\theta_{min} = \frac{\theta_0^- + \theta_{LS=0}}{2} + \frac{1}{2K_0} - \frac{K_1^-(\theta_{LS=0} - \theta_0^-) + K_1^+(\theta_{LS=0} + \theta_0^+)}{2K_0K_1^-K_1^+(\theta_0^+ + \theta_0^-)} \cdot \frac{\delta_m}{I_{in}} \quad (20)$$

Inserting in eq. (20) the typical parameters given in eq. (10), a value of δ_m of 6.8 Hz/K corresponding to 9.5 Torr of pure nitrogen, we have for:

I_3 : $[\theta_{min}]_{th.} = 45°C$

I_5 : $[\theta_{min}]_{th.} = 62°C$

I_6 : $[\theta_{min}]_{th.} = 68°C$

1.4 Conclusion

As a conclusion we can state that the phenomenological model used here is able, in spite of its simplicity, to predict the cell temperature for zero light shift and the cell temperature for zero cell-temperature coefficient. This ability to predict $\theta_{LS=0}$ and θ_{min} gives the possibility to choose the relative isotopic mixture of rubidium in the lamp, the light intensity impinging on the filter cell and the mixture of buffer gas and their pressure in the cell in order to reach $\theta_{LS=0} = \theta_{min}$.

2. Laser pumped Rubidium Standards

2.1. Impact of laser pumping on performance

Semiconductor lasers are very promising light sources for optical pumping of rubidium standards. Their very high efficiency and small size make them particularly attractive for miniature rubidium standards where power consumption, warm-up time and overall size are important parameters. In this perspective, such requirements as low cost and simplicity of the optical set-up preclude the use of elaborate techniques to enhance the spectral and noise properties of the laser (e.g. electrical and/or optical feedback using an external cavity). Measurements performed previously in our laboratory [7] on a Rb standard with a cell containing pure Rb^{87}, pumped by an off-the-shelf laser (Hitachi HL7801), show that

a) the light-shift coefficient (microwave frequency vs laser frequency) is $5 \cdot 10^{-11}$/MHz. A rubidium standard with $10^{-11}\ \tau^{-1/2}$ stability requires a laser stability of $5 \cdot 10^{-10}\ \tau^{-1/2}$

b) owing to the saturation effect of the light shift vs laser power [7], sensitivity of the microwave frequency to laser power changes is much smaller than in a conventional device but is still close to $3 \cdot 10^{-12}$/%

c) the laser frequency, when locked to the Rb^{87} absorption cell ($5S_{1/2}$, F=1 to $5P_{3/2}$ transition) shifts by -9MHz/K with cell temperature

d) this optical frequency shift converts into a microwave shift through the light-shift effect (a) and adds a buffer-gas independent contribution to the overall temperature coefficient of the standard. This contribution is ~ $4.5 \cdot 10^{-10}$/K (cell temperature)

e) in the relevant Fourier frequency range (100Hz to 1kHz) the laser intensity noise is > $10^{-6}/Hz^{1/2}$, more than an order of magnitude over the ideal shot noise level normally associated with incoherent pumping ($<10^{-7}/Hz^{1/2}$)

f) Aging of the laser parameters is commonly observed. When the laser is frequency locked as in a Rb standard, aging may substantially affect the long-term stability and also reduce the operational lifetime because of eventual mode jumps.

2.2. Influence of reflected light

Figure 4 shows an experimental set-up of the laser diode pumped ^{87}Rb frequency standard. A collimating lens is used to increase the power density passed through an isotopic ^{87}Rb cell contained in the cavity. The focal distance of the lens is 20mm. The light transmitted by the absorption cell is used to lock the laser frequency to the D_2 resonance of ^{87}Rb $5S_{1/2}$ -$5P_{3/2}$ by an electronic servo system. The oscilloscope is used to observe the absorption curve and monitor if the laser diode is locked. PM6652C counter, HP9826 calculator and SR530 lock-in amplifier are used to measure noise, signal, beat frequency and short term stability. Figure 5 shows the slow quasiperiodic changes observed in the beat frequency, between the laser diode pumped ^{87}Rb clock and a reference Cs clock (at 5MHz) having $\sigma_y(\tau) < 10^{-12}$ for $\tau > 1s$. This change is caused by the reflected light of the laser diode from the lens. In fact the reflected light can be easily injected into the laser diode from the lens. When the reflected light is injected into the laser cavity, the frequency of the laser diode will be pulled [4], [5], [6] according to the length of the external cavity. Indeed, the frequency of the quasi-periodic variations of the beat frequency was observed to be strongly correlated with the time-derivative of ambiant

temperature changes. In order to reduce injection of reflected light, we did not use an optical isolator because of its high cost. Instead, the lens was removed and the laser diode was fitted onto the physical package, as close to the ^{87}Rb absorption cell as possible in order to allow the whole laser beam to pass through the cell. In our experiment, the distance between laser diode and ^{87}Rb cell is 30mm. A glass plate was introduced in front of the cell, its unpolished side facing the laser diode, in order to further reduce backscattering. If the frequency instability of the ^{87}Rb clock is mainly caused by the reflected light of the laser diode in previous experiment, then in this new condition, it must be better. As was expected the experimental result is better than before because fig. 6 shows a small and slow frequency fluctuation of the Rb clock.

2.3. Stability

The following parameters were chosen in order to get the largest rf resonance signal and the smallest light shift coefficient and temperature coefficient [7]: power density of the laser diode 270 $\mu w/cm^2$, cavity temperature 65°C and Rb^{87} isotopic absorption cell with a mixture gas (CH_4/N_2 = 1, total pressure 10mbar). Preliminary measurements of the short term stability of the laser diode pumped ^{87}Rb frequency standard (fig 7) are as good as that of a small conventional Rb frequency standard ($2 \cdot 10^{-11}\ \tau^{-1/2}$). Improvement of the stability can be expected by reducing the influence of reflected light and noise from laser diode.

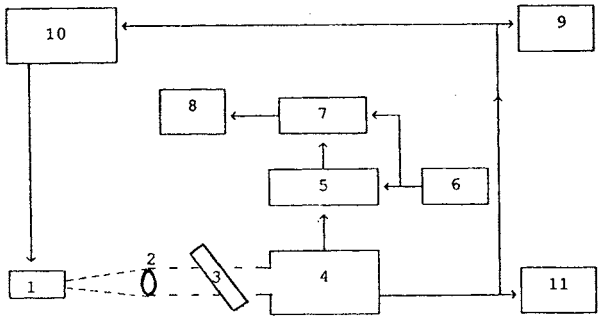


Fig. 4 Experimental set-up diagram of the laser diode pumped ^{87}Rb frequency standard

1. laser diode
2. lens
3. filter
4. Rb clock without the lamp
5. mixer
6. reference frequency (5MHz)
7. PM6652C counter
8. HP9826 calculator
9. SR530 lock in amplifier (S/N measurement)
10. servo system for laser diode locking
11. ocilloscope

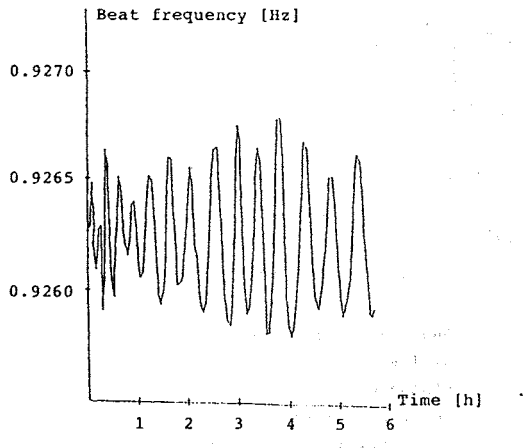


Fig. 5 Frequency fluctuation of the laser diode pumped ^{87}Rb frequency standard (with optical feedback from lens and cell: DC level from photodetector, 0.270V)

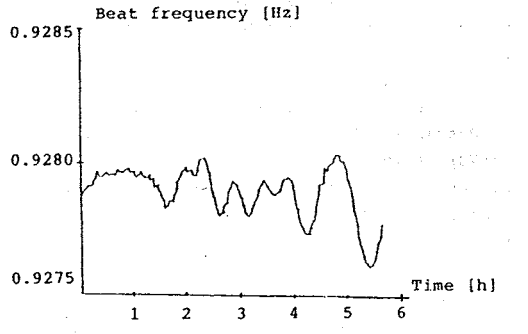


Fig 6 Frequency fluctuation of the laser diode pumped ^{87}Rb frequency standard (feedback reduced, no lens, groundglass plate; DC level from photodetector, 0.210V)

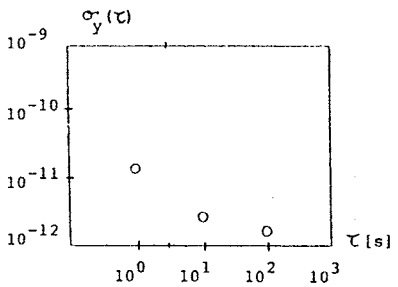


Fig 7 Short term frequency stability of laser pumped ^{87}Rb frequency standard (average of 10 runs; reference oscillator stability is better than 10^{-12} in the relevant range)

References:

[1] J.C. Camparo: Proc. of the 4th Symposium on Frequency Standards and Metrology, Ancona, 1988, p. 62.

[2] J. Vanier and C. Audoin: The Quantum Physics of Atomic Frequency Standards (A. Hilger, 1989).

[3] G. Busca, M. Têtu, J. Vanier: Can. J. Phys. 51, (1973), 1379.

[4] R. Lang, and K. Kobayashi, IEEE J. Quantum Electron QE-16, p. 347 (1980)

[5] R.F. Brown, E. Mohn, C. Risch and R. Salater, 1970, IEEE J. Quantum Electron QE-6, 328 (1970)

[6] K. Kobayashi, Trans. IECE Jpn. E, 59, 8 (1976)

[7] J.T. Liu, P. Thomann, L. Zhang and G. Busca, Proceedings of the 4th European Frequency and Time Forum, EFTF90, p.157.

(1) Present address:
Wuhan Institute of Physics, P.O. Box 241, Wuhan, Hubei, The People's Republic of China

(2) Present address:
Observatoire de Neuchâtel, CH - 2000 Neuchâtel, Switzerland

Acknowledgements:

We thank the Swiss Commission CERS, for supporting the present work.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4B Number 31

EXPERIMENTAL STUDY OF TWO-LASER OPTICAL PUMPING IN A CESIUM CLOCK

N. DIMARCQ, V. GIORDANO, C. JACQUES*, G. THEOBALD, P. CEREZ

Laboratoire de l'Horloge Atomique
Unité Propre de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud - 91405 ORSAY Cedex - France

ABSTRACT

Two-laser optical pumping shows great promise for cesium preparation since it can accumulate all the atoms in a single $m_F = 0$ ground state sublevel. Therefore a large enhancement of the clock signal to noise ratio should result. Moreover the Rabi pulling effect has no reason to be present anymore.

We have focused our attention on the most promising pumping scheme which was shown to be the $F = 3 \rightarrow F' = 4$ π (D_2) ; $F = 4 \rightarrow F' = 4$ π (D_2) scheme. We have used powerful commercial laser diodes for the experimental study. Full pumping efficiency was almost obtained. However no improvement of the clock S/N ratio was observed. This seems to be due to the presence of unpumped atoms on the one hand and to laser frequency noise on the other hand.

I. INTRODUCTION

Preparation of a cesium atomic beam by laser optical pumping has been proposed to increase the performances of cesium beam frequency standards which currently operate with magnetic state selection. Different pumping schemes have been introduced using one or two lasers.

In the case of one-laser pumping a series of experiments [1-5] has been of great importance in verifying the performance of OPCsB frequency standards operated with either cesium D_1 or D_2 lines.

Two-laser pumping shows great promise for cesium preparation since it can realize a nearly perfect state selector by accumulating all the atoms in a single $m_F = 0$ ground state sublevel. Therefore a large enhancement of the clock signal amplitude results. The vanishing of the neighbouring field dependent microwave transitions eliminates the Rabi pulling effect[1]. However, restrictions imposed by the fundamental properties of cesium atoms and by experimental conditions prevent the achievement of perfect pumping. Earlier experiments [6,3] pointed out the difficulty of attaining a large population difference between the two clock sublevels. Moreover, a poor clock S/N ratio was obtained due to the occurrence of an additional noise in the optical detection process. Consequently, the performance of an optically pumped cesium beam resonator operated with two-laser pumping preparation was not yet comparable to that obtained when a single laser is used both for optical preparation and detection.

The importance of effects such as the different kinds of atomic coherences, the finite laser linewidths and the optical leakages has been emphasized by recent models of optical pumping [7-8]. They have demonstrated that atomic coherences lead to the reduction of the pumping efficiency when one of the two lasers is σ-polarized. To obtain a higher optical pumping efficiency, it would be better to avoid the σ polarization of the laser light. The models have also pointed out that use of narrow linewidth lasers creates hyperfine coherences, which results in less efficient pumping. Finally, broad linewidth lasers excite neighbouring levels not desired in the pumping scheme, which also leads to a smaller efficiency. Guided by these conclusions derived from the theoretical models, we have focused our attention on the most promising pumping scheme which was shown to be the $F = 3 \rightarrow F' = 4$ π (D_2) ; $F = 4 \rightarrow F' = 4$ π (D_2) scheme. We have used powerful laser diodes at 852 nm with linewidth around 6 MHz for the experimental study. This paper summarizes the results obtained for the pumping efficiency and the clock signal to noise ratio.

II. EXPERIMENTAL SET-UP

The experimental set-up used the cesium beam resonator Cs 3. A detailed description can be found in a previous article [9]. The optical pumping was performed in the first region by means of one laser tuned to the $F = 3 \rightarrow F' = 4$ transition of the cesium D_2 line and of a second laser tuned to the $F = 4 \rightarrow F' = 4$ transition of the cesium D_2 line. Both were π polarized. In the central region, the atoms undergo the microwave transition at 9 192 MHz. The optical detection was performed in the third region by a fraction of the light delivered by the laser tuned to the $F = 3 \rightarrow F' = 4$ transition. The laser diodes STC, Model LT50A, are commercially available. They deliver an optical power up to 50 mW and their linewidths have been measured in the range 5 - 7 MHz.

* Institute for National Measurement Standards, National Research Council, OTTAWA K1A0R6, Canada

III. EXPERIMENTAL RESULTS

When the microwave frequency was swept around 9 192 MHz, the resonator delivered a microwave spectrum corresponding to the seven transitions $\Delta F = 1$, $\Delta m_F = 0$. The central line of this spectrum is the well known clock transition betwen the $m_F = 0$ sublevels.

In order to characterize the efficiency of two-laser pumping, we evaluated the difference Δn in population of these two levels and the clock signal to noise ratio. The amplitude of the central peak at resonance defines the signal. The noise was measured at the top of this peak in a one Hertz bandwidth. Δn and S/N have been determined for two different methods of frequency stabilization of each laser.

a) Conventional laser frequency stabilization set-up

In this case, the frequency of each diode was modulated at a different modulation frequency lying in the range 20 to 50 kHz. The modulation amplitude was equal to 30 MHz for each laser. The whole fluorescence signal delivered in the pumping region was then detected by a photodiode and demodulated by means of two synchronous detectors. A servo-loop was used to lock the frequency of each laser to the top of the relevant fluorescence line. The servo bandwidth was about 1 kHz.

The microwave spectrum recorded with this experimental set-up is depicted in Fig. 1a. In Fig. 1b, the microwave spectrum obtained with one-laser pumping is recorded for purpose of comparison. One sees on the one hand, the large enhancement of the central line, and on the other hand, the vanishing of the field-dependent lines except the weak lines with $m_F = \pm 1$ which are still present. It is worth noting that the background I_b due to the unpumped atoms in the F = 3 level is not negligible in Fig. 1a whereas it vanishes in Fig. 1b.

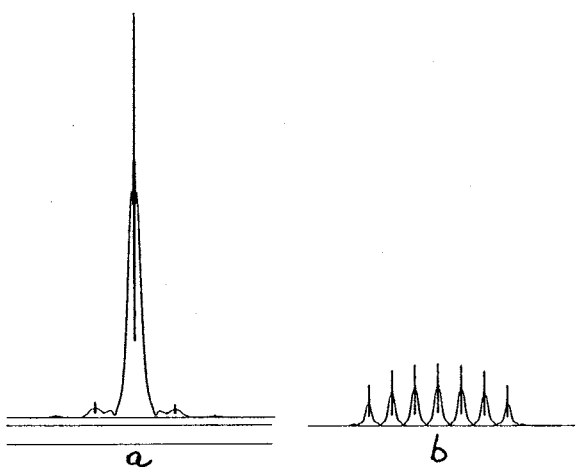


Fig. 1. Microwave spectrum. (a) with two-laser pumping. (b) with one-laser pumping. I_b is the fluorescence background due to unpumped atoms. I_c is the background due to strong light. The values for Δn and the S/N ratio are
(a) $\Delta n = 91$ % ; S/N = 4 100 in a 1 Hz BW
(b) $\Delta n = 13.7$ % ; S/N = 15 000 in a 1 Hz BW.

The difference Δn in population of the two levels of the clock transition was measured by comparison between the amplitude of the central peaks of Fig. 1a and 1b. Since in Fig. 1b the central peak corresponds to $\Delta n = 13,7$ % [6], we deduced that $\Delta n = 91$ % in Fig. 1a. A theoretical estimate of this quantity calculated according to [8] is $\Delta n = 95$ %. In Fig. 1a, the observed clock S/N ratio is 4100 while it is 15000 in Fig. 1b. The S/N ratios were measured in a 1 Hz bandwidth.

As a conclusion, a good signal amplitude was obtained but a rather large noise appeared at the detection which was detrimental to the S/N ratio. This noise strongly depends on the features of the modulation of the pumping lasers such as the frequencies and the amplitudes and depends on the beam injection conditions too. This led us to be suspicious of the method of stabilization of the mean frequency of the lasers. As a matter of fact, the error signal was derived from a large noisy fluorescence signal which reflects the frequency noise of the two lasers. Consequently, another method of laser frequency stabilization was introduced.

b) Fabry Perot prestabilization of the laser frequency

Its consisted in prestabilizing each laser frequency with a confocal Fabry Perot resonator. The F.P. resonators had a spacing length equal to 5 cm, their free spectral range was 1 500 MHz and their finesse was 30. The optical discriminator bandwidth was equal to 50 MHz and the servo bandwidth was about 20 kHz.

The population difference Δn was the same as previously obtained but the clock S/N ratio was then 10 000 in a 1 Hz bandwidth.

c) Summary

Although a considerably improved Δn was obtained with two-laser pumping, the S/N ratio remained less or of the order of that attained with the resonator operated with a single laser diode. The S/N ratio is limited by a strong additional noise which is likely due to the unpumped atoms and the frequency noise of the laser. To illustrate this large amount of noise, Fig. 2 reports the spectrum of noise recorded in the detection region.

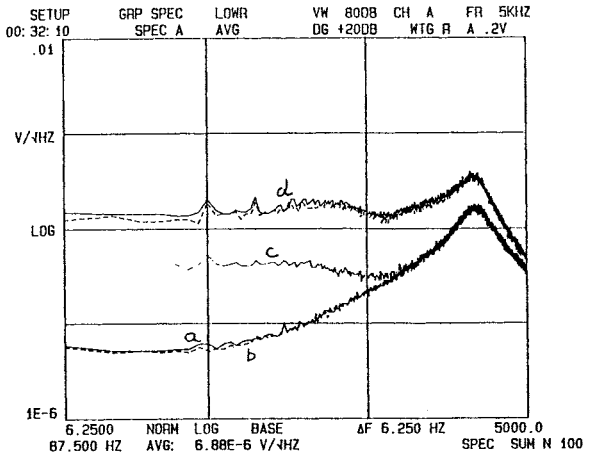


Fig. 2. Noise spectrum in the detection region.
(a) stray light (b) detector noise
(c) unpumped atoms without microwave excitation
(d) noise at the top of the central peak of the microwave spectrum.

IV. ENCOUNTERED DIFFICULTIES

In resonator Cs 3, the atom-light interaction time is about 20 μs. This value is comparable to the pumping time in our two-laser pumping scheme. As a consequence, many experimental conditions become restrictive. They are for example the finite laser linewidths which involve leakage effects, and variations in the polarization and the alignment of the laser beams.

Experimentally we established that :

1. It needs a stringent π polarization set-up. For example a weak proportion of $4 \rightarrow 4$ σ light creates leakages which lengthen the pumping time.

2. Serious drawbacks due to the anisotropy of the quartz window in the Cs 3 apparatus occur and they have to be compensated.

3. The two laser beams must be strictly superposed.

4. An important problem is related to the laser frequency stabilization, which can induce an additional noise.

V. CONCLUSION

Full optical pumping efficiency with two powerful commercial laser diodes at 852 nm was almost obtained. However, a non-negligible amount of unpumped atoms is still present which results in a noise increase so that no improvement of the clock S/N ratio has been obtained compared to the case where the resonator is operated with a single laser. Further experimental and theoretical investigations concerning the noise sources are required. However, these results let us predict that a large clock S/N ratio should be obtained as soon as the number of unpumped atoms can be reduced.

REFERENCES

(1) V. CANDELIER, V. GIORDANO, A. HAMEL, G. THEOBALD, P. CEREZ and C. AUDOIN. Appl. Phys. B 49, 365 (1989)

(2) N. DIMARCQ, V. GIORDANO, G. THEOBALD and P. CEREZ. J. Appl. Phys. 69, 1158 (1991)

(3) S. OHSHIMA, Y. NAKADAN and Y. KOGA. IEEE Trans. Instrum. Meas. IM-37, 409-413 (1988)

(4) D.H. YANG and Y.R. WANG. Optics Comm. 73, 285 (1989)

(5) S.I. OHSHIMA, Y. NAKADAN, T. IKEGAMA, Y. KOGA, R. DRULLINGER and L. HOLLBERG. IEEE Trans. Instrum. Meas. IM-38 (2), 533 (1989)

(6) G. AVILA, V. GIORDANO, V. CANDELIER, E. de CLERCQ, G. THEOBALD and P. CEREZ. Phys. Rev. A 36, 3719 (1987)

(7) E. de CLERCQ and P. MANGIN in Proc. 3rd European Time and Frequency Forum, Besançon, France (1989) pp. 281-292

(8) P. TREMBLAY and C. JACQUES, Phys. Rev. A 41, 4989 (1990)

(9) V. GIORDANO, A. Hamel, P. PETIT, G. THEOBALD, N. DIMARCQ, P. CEREZ and C. AUDOIN. IEEE Trans. Ultrasonics, Ferroelectrics, Freq. Control. To be puplished (July 1991).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4B Number 32

Laser cooling of Cs atomic beam at NRLM

Takeshi Ikegami, Shin-ichi Ohshima, and Yasuhiro Nakadan

National Research Laboratory of Metrology, 1-1-4, Umezono, Tsukuba-shi, Ibaraki-ken 305, Japan

Abstract

Present status of the laser cooling experiments of Cs atomic beam at NRLM are presented. In this year, the laser cooling experiment was started at NRLM to search for the feasibility of the atomic fountain clock. As a first step, we performed the Cs atomic beam deceleration experiment. Atomic-beam cesium atoms were decelerated by counterpropagating radiation from a semiconductor laser diode. The laser frequency was chirped to keep it in resonance with the decerelated atoms. The velocity distribution was almost Maxwellian when the cooling laser was off, and slow atoms were significantly increased when the cooling light was introduced. Furthermore, some of the atoms were slowed to zero or negative velocity by tuning the frequency of the cooling laser. We are now continuing to work on improving the production efficiency of slow atoms, and are pursuiting applications of slow atoms to trapping.

1. Introduction

Laser cooling is expected to be a promising technique to realize more accurate and stable atomic clocks[1]. Recently, many experimental results were reported concerning the trapping of sodium and cesium atoms[1]. By utilizing the multilevel structure of the atoms, the kinetic energy of the laser-cooled and trapped atoms was reduced to the order of microkelvin, which is the order of the recoil temperature[2]. Because of this ultra low temperature, the atomic fountain clock using Cs was recognized as a realizable one although it was abandoned a few tens of years ago because of the low S/N. Because the loss of atoms due to diffusion in flight is negligible, it is considered that the problem of the S/N can be cleared[3], and therefore some laboratories began the preliminary experiment[4], and the theoretical estimation[5] of the atomic fountain clock.

In NRLM, the laser cooling experiment was started this year. As a first step, we tried the deceleration of the cesium atomic beam using laser diodes(LDs). There are two methods to compensate the frequency change due to the first order Doppler-effect during the decerelation process, and we adopted the frequency chirping scheme this time. Although the obtained velocity width of the decelerated atomic beam was rather broad, we are now trying to improve the efficiency of the deceleration by narrowing the spectral linewidth of the LD and by the use of more accurate frequency control technique such as offset-locking. Furthermore, as a second step, we are also pursuing the trapping.

2. Experimental setup

Experimental setup is shown in Fig.1. Cesium atoms were effused from an oven which was heated to a temperature of about 130°C. The oven had an aperture equipped with a glass capillary array(GCA). The diameter of each capillary was 50μm, the length was 2mm, and the total diameter of the aperture was 2mm. By this GCA, the beam could be collimated to an angular divergence of 3° FWHM. Because of the small dimension of the GCA, the broadening of the angular divergence due to the atomic collision inside the GCA was negligible. Another aperture of 2mm diameter was placed between the oven and the detection region, whose distance from the oven was 130cm. The atoms which was deflected during the deceleration process were eliminated by this aperture. Therefore, only the atoms in the center line of the atomic beam can reach the detection region. The total length of the decerelation region was about 1.45m. The vacuum chamber was typically 1x10^{-5}Pa. All the optical windows of the chamber are with anti-reflection coating and their transmittion coefficient is more than 99%. Magnetic shield was not used.

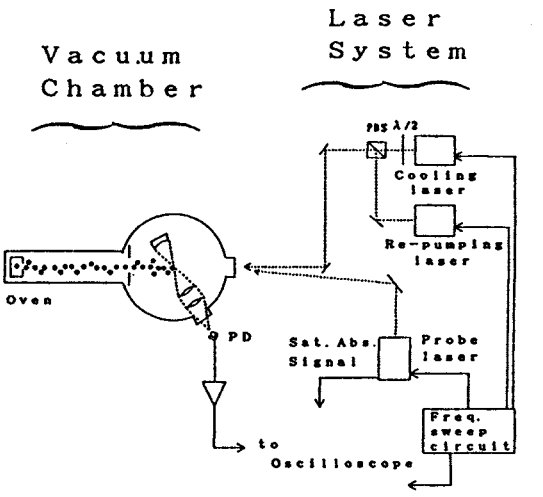


Fig. 1. Experimental setup

Three semiconductor laser diodes (Hitachi HL8314E, λ=852nm, P=10mW) were used for cooling, re-pumping, and detection of atoms. All the lasers were used under free running condition, and only the temperature was stabilized[6]. The spectral linewidth of each LD is typically of the order of 10MHz. Light beam from the LD were collimated to a

parallel beam with a ellipsoidal cross section of 2x4mm by a collimator lense. Available power of the each laser diode was 6mW, 2mW, 1.5mW at the entrance to the vacuum chamber, respectively.

Because the frequency of the laser diode can be changed almost linearly by changing the injection current to the LD(-3GHz/mA), we chirped the frequency of the LD by changing this injection current.

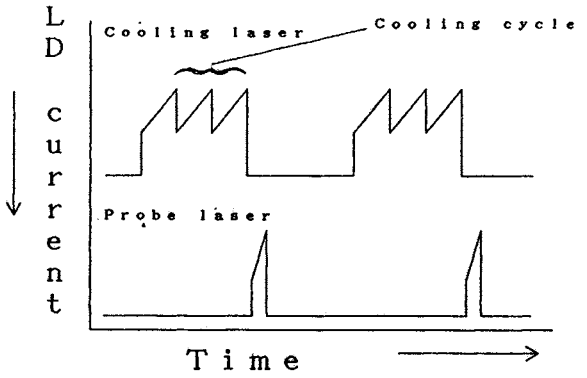


Fig. 2. Timing of frequency chirping

As shown in Fig.2, the frequency of the cooling laser was chirped to keep it in resonance with the decelerated atom. The dulation of one cooling cycle and the sweep width was tuned so as to maximize the decerelation efficiency monitored by the fluorescence signal. After the decerelating cycle was completed, the frequency of the cooling laser was shifted by a few hundred MHz.

The light from the optical pumping laser was overlapped with the cooling light by a polarization beam splitter and both light co-propagate. The optical pumping laser was always on. The frequency of the pumping laser was modulated with 12.5kHz with the modulation width of about a few 100MHz. Although the fluorescence signal could be detected even if the frequency of the pumping laser was chirped simultaniously with that of the cooling laser, the detected signal was small (almost equal to noise level) in this case because of the optical pumping during the detection process, which is stated in section 3. In our scheme, the optical pumping laser works to re-pump the atoms not only in the decerelation process, but in the detection process. We adopted this scheme to enhance the signal level in the detection process.

The third laser was used for velocity distribution measurement of the decerelated atoms. As shown in Fig.2, the laser frequency was quickly chirped after cooling cycle was completed. Utilizing the Doppler-effect, we can detect the velocity distribution by monitoring the fluorescence from the atom in the detection region as a function of the injection current to the LD. The angle between the atomic beam and the probe laser beam was about 4 degree. As already stated, because of the pumping laser, the transition becomes closed cycle. The solid angle subtended by the fruorescence light collection optics was about 1 steradian($\Omega/4\pi$ =10%). Photodetector was placed just outside the vacuum chamber and the fluorescence signal was sampled and accumulated by a digitizing oscilloscope. To scale the frequency axis, the saturated absorption spectrum of the cesium D_2 line was simultaniously taken[7]. These data was taken by the personal computer and analyzed later.

3. Results

When the cooling laser was off, the velocity distribution was almost Maxwellian as shown in Fig.3 if we consider these fluorescence was from F=4→ 5 transition. From the figure, we can see that the observed signal was proportional to $v^3\exp(-mv^2/2kT)$ when only the probe laser was introduced. Whereas, it was proportional to $v^2\exp(-mv^2/2kT)$ when the pumping laser was also introduced. This is due to the fact that when the pumping laser was not introduced, the atoms are pumped to the F=3 level due to the tail of the probe laser spectrum. When the pumping laser was introduced, F=4→ 5 becomes an closed cycle.

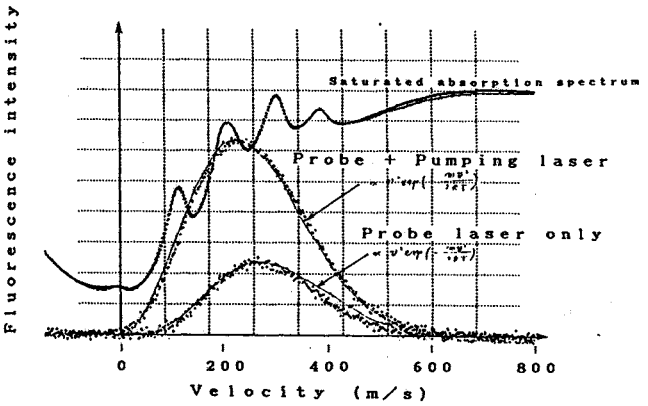


Fig. 3. Velocity distribution (cooling laser off)

Next, we introduced the cooling laser. Fig.4 shows the cooling laser power dependence of the velocity distribution at the fixed cooling laser frequency. We can see that the width of the valley becomes broader according to the increase of the laser power.

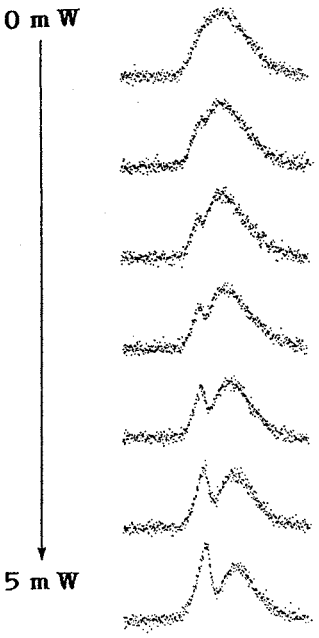


Fig. 4. Cooling laser power dependence

Next we chirped the frequency of the cooling laser. In this case the slow atoms was increased like Fig.5(b) from Fig.5(a). By adjusting the frequency of the cooling laser, we could bring the average velocity of the decerelated atoms to zero like Fig.5(c)., and further to negative velocities (Fig.5(d)). The frequency of the cooling cycle and the sweep width was adjusted so as to maximize the decerelation efficiency. The optimum frequency was about 75Hz and the width was about 300MHz.

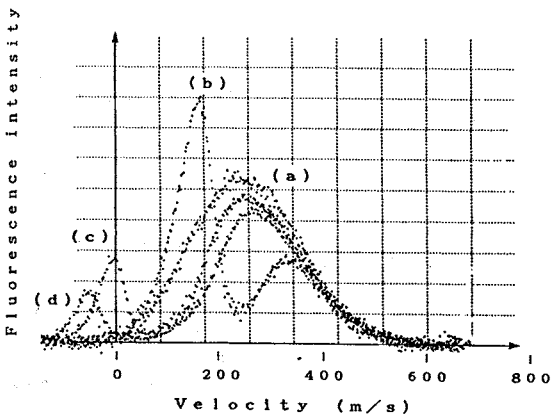


Fig. 5. Velocity distribution (cooling laser on)

4. Discussions

In our experiment, obtained velocity distribution of the decerelated atoms was much broader than the spectral linewidth of the laser diode. To consider this, we simulated the velocity distribution of the atoms. Fig.6 shows the simulated velocity distribution when the frequency of the cooling laser is fixed assuming the LD spectral linewidth of 10MHz. In the figure, G represents a saturation parameter $G=2(dE_0/2(h/2\pi))^2/\Gamma\gamma$ [8], where d is the matrix element of the atomic dipole moment, E_0 is the amplitude of the electric field of the laser light, h is the Planck constant, $2\gamma=1/\tau$ is the natural linewidth, and β in $\Gamma=\gamma+\beta$ is half of the spectral linewidth of the LD. Interaction time was assumed to be 5ms, which corresponds to the atomic velocity of 300m/s and the decerelation length of 1.45m. The effect of the diffusion of atoms was not taken into account and only the average motion of the atom was included. Comparing this figure with the bottom of the Fig.4, the saturation parameter can be estimated at the order of 1. The saturation parameter G can also be calculated directly by substituting the numerical value in the above formula of G. This again gives G the value of order 1, which is consistent with the above estimation.

If the saturation parameter is 1, the atoms with the most probable velocity (270m/s) can be decerelated to 0 m/s in a distance of about 1.4m during the time of 12ms, which corresponds to the frequency of $(12ms)^{-1}=83Hz$. and the atom with the velocity of 300m/s can be decerelated to 150m/s in the same distance and the same time. This value almost agrees with our experimental result of Fig.5.

Although the observed velocity width of the decerelated atoms was broad compared with this simulation, this is due to the finite bandwidth of the detector. In fact, we could narrow the velocity width by increasing the bandwidth of the electronic system. However, in the present, we cannot increase the bandwidth any more because the signal itself is very small. Although we can increase the integration time to increase the S/N, the laser frequency drifts during the integration and the observed velocity distribution broadens.

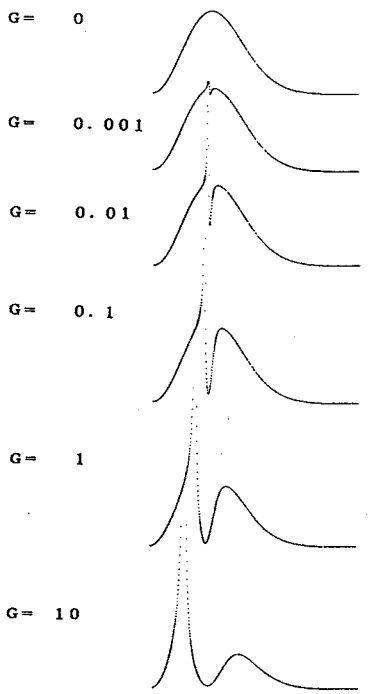


Fig. 6. Simulation

To proceed the experiment further, we must increase the S/N ratio and increase the bandwidth of the electronic system. The frequency of the LD must be controlled more precisely to supress the frequency drift.

These are resolved by increasing the atomic flux without degrading the vacuum and by increasing the fluorescence collection efficiency. Furthermore, we must collimate the atomic beam. Next, we must narrow the spectral lineidth of the laser diodes to increase effectively the saturation parameter.

We are now continuing to improve the efficiency and are also pursuiting the experiment for trapping these decerelated atoms.

If we can further increase the laser power, we can utilize another effective methods such as stimulated cooling[9]. By this method, we can shorten the deceleration length to the order of 10cm. Necessary power is estimated to be about $100mW/cm^2$[9] and this will be attained by a high power laser such as a titanium sapphire laser.

References

[1] The Special Issue of J.Opt.Soc.Am.B2 (1985), J.Opt.Soc.Am.B4(1989).
[2] C.Salomon, J.Dalibard, W.D.Phillips, A.Clairon and S.Guellati, Europhys.Lett. 12,683(1990).
[3] W.D.Phillips, Japan-United States seminar on Quantum Electronic Manipulation of Atoms and Fields, ed.F.Shimizu, W.D.Phillips, and C.I.Westbrook, p.317.
[4] M.Kasevich, E.Riis, and S.Chu, Phys.Rev.Lett.63,612(1989).
[5] A.De Marchi, Metrologia 18,103(1982);

R.G.Beausoleil and T.W.Haensch, Phys.Rev.A33, 1661(1986).
[6] T.Ikegami, S.-I.Ohshima, Y.Nakadan, and Y.Koga, Rev.Sci.Instrum.61,3719(1990).
[7] S.Ohshima, Y.Nakadan, and Y.Koga,IEEE Trans.Quantum Electron.QE-23,473(1987).
[8] V.G.Minogin and V.S.Letokhov, Laser light pressure on atoms, Gordon and Breach, New York, 1987, p.131.
[9] A.Aspect, J.Dalibard, A.Heidmann, C.Salomon, and C.Cohen-Tannoudji, Phys.Rev.Lett. 57,1688(1986).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 4B Number 33

Frequency Synthesis in Primary Cesium Clocks

R. Schröder

Physikalisch-Technische Bundesanstalt
3300 Braunschweig, Germany

Abstract

A frequency synthesis and control system has been realized to be used with the PTB's primary cesium clocks. It consists of all elements necessary for the generation of the signal for probing the atoms and locking a 5 MHz VCXO to the atomic transition line. Square wave frequency modulation is used for determining the line centre. The system will be described in some detail.

The mechanisms have been examined by which the electronically determined line centre may deviate from the "true" line centre at the given experimental conditions. Transients and phase-modulations in the synthesized frequency have been identified and will be explained in detail.

1. Introduction

In a cesium atomic clock a 5 MHz quartz oscillator is under control such that its output frequency is held in a well known relation to the centre frequency f_c of the (4,0)-(3,0) clock transition of a ^{133}Cs atom. This frequency f_c has the value $f_c=f_0+\Delta f$ where f_0 is defined as $f_0=9192631770$ Hz /1/ and Δf are the frequency shifts due to various physical effects occuring in the beam tube, like Zeeman effect and 2nd order Doppler effect. The transition is induced in the Ramsey scheme /2/ where the atoms are irradiated by two phase coherent microwave fields separated in time by an interval T. In case of the PTB's primary clocks CS1 and CS2, T is 8 ms and a linewidth of the central Ramsey fringe of W = 60 Hz is the result. In the recent constructions, CS3 and CS4, W is 44 Hz.

The subject of this paper is the electronic system to be used commonly in all PTB's clocks with a particular emphasis on the new design of the frequency synthesis system and possible sources of errors. In section 2 the basic principle of the electronic system is illustrated. A description of the frequency synthesis electronics follows in section 3. In two more sections the two principal electronic effects in the frequency synthesis are considered which may lead to frequency errors:
1) To determine the line centre the frequency f_s of the probing microwavae signal has to contain some modulation. Square wave frequency modulation is used here and it is discussed that transients in the signal at f_s can occur and may lead to frequency errors. They can, however, be suppressed to become smaller than 10^{-15} (in relative units). This is achieved by use of a so-called blanking interval: The beam signal is suppressed during the time when the transients occur.
2) Phase sidebands of the probing signal originate from the mechanism of frequency synthesis. It is shown that these phase modulations lead to a zero net effect only when a proper choice of the modulation period and the duration of the blanking interval is made. This is confirmed by experiments.

2. Principle of the operation

The clock control loop with the overall time constant of about 30 s contains the elements cesium beam tube, signal processing unit, VCXO and frequency synthesis unit. The system is illustrated in Fig. 1a. The sequence of some relevant signals is shown in Fig. 1b. The signal s_1 coming from the electrometer will contain an amplitude modulation if the mean frequency of the probing signal, $f_s°$, deviates from the line centre frequency f_c. By phase-sensitive detection, filtering, transient suppression and integration the control voltage to be applied to the VCXO is generated. The probing signal is then generated in the frequency synthesis unit. Its output frequency f_s is switched between $f_s = f_s° - f_H/2$ and $f_s = f_s° + f_H/2$ with the period T_m where f_H is close to W or 3 W. By changing the modulation width, an information about the symmetry of the Ramsey line can be obtained. The signal realized in the frequency synthesis unit is finally fed into the beam tube. Some features of the control loop shall be pointed out. All signals used for triggering the frequency switching, the blanking interval and the phase sensitive detection of the signal s_2 are derived from the VCXO. Signal b determines the blanking interval T_b. Signal p, which determines the phase sensitive processing of signal s_2, is delayed by T_b with respect to the frequency switching, taking into account the delay of the signal s_2 caused by the time constant of the filter amplifier.

The controller is a I-PI combination whereby the linear frequency drift of the free-running VCXO is eliminated without zero offset effects on the realized frequency f_s. Offset voltages in the order of 5 μV in the phase sensitive detector and the integrator have to be considered since they lead to frequency errors. The gain in the filter amplifier has to be chosen as high as possible to reduce the effect of these offset-voltages on the frequency f_s to values lower than 10^{-15} (in relative units). Therefore the suppression of the large transient impulses by use of the blanking unit may be necessary yet from this point of view.

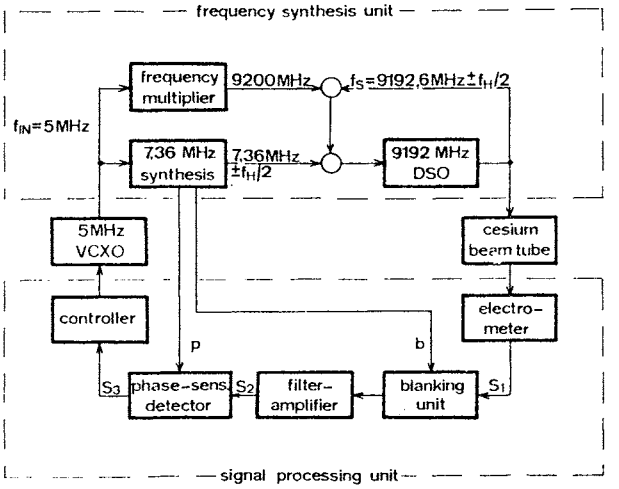


Fig. 1a: Control loop of cesium clock

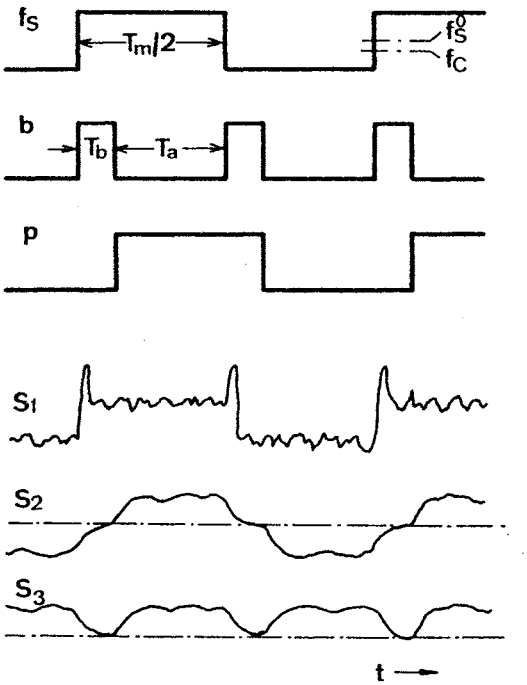


Fig. 1b: Relevant signals of control loop

3. Frequency synthesis unit

The frequency f_{IN} = 5 MHz of the VCXO is multiplied up to the microwave frequency range in four steps. The first three stages (5•2•2) are realized using low phase noise transistor multipliers. The last step (x92), done by a step recovery device, yields a signal at 9200 MHz. When it is mixed with the signal from the microwave source (DSO-oscillator) at 9192,6 MHz, a signal at the difference frequency 7,36.. MHz is the result. It is compared with the signal from the 7,36 MHz-synthesis (frequency f_D = 7,36.. MHz) in a second mixer, whose output delivers then the control voltage applied to the DSO. Its frequency follows the relation

$$f_s = 1840\ f_{IN} - f_D \qquad (1)$$

Fig. 2 depicts the circuit used for the synthesis of f_D = 7,36.. MHz. Two phaselock loops are included. With the primary 7,36 MHz-phaselock loop the signal at f_D

$$f_D = 7{,}\bar{3}\ \text{MHz} + f_2 \pm f_H/2$$

is realized. The signal at $f_2 \pm f_H/2$ is delivered by the 29,6 MHz phaselock loop. So finally the relation holds

$$f_D = 7{,}\bar{3}\ \text{MHz} + \frac{20\ \text{MHz} \cdot m}{n_1 n_2} \qquad (2)$$

or

$$f_D = \left(\frac{22}{15} + \frac{4m}{n_1 n_2}\right) f_{IN} \qquad (3)$$

The frequency modulation $\pm f_H/2$ is generated by switching the m-divider with the modulation-period T_m. The value m can be changed by $\Delta m=1$ or $\Delta m=3$ so that f_H is equal to $f_{H1} \approx W$ or $f_{H3} = 3f_{H1} \approx 3W$. Rewriting equation (2) leads to an expression which clearly shows the switching of f_D around a mean value f_D^0.

$$f_D = f_D^0 \pm f_H/2$$

where

$$f_D^0 = 7{,}\bar{3}\ \text{MHz} + 20\ \text{MHz}\ \frac{m_0 + \frac{1}{2}}{n_1 n_2} \qquad (4)$$

and

$$f_H = \frac{20\ \text{MHz} \cdot \Delta m}{n_1 n_2}$$

The choice of the integer n_1, n_2 and m_0 makes the system adaptable to the different clocks CS1, CS2, CS3 and CS4. The relevant values are given in table 1.

Resulting from the selected values of n_1, n_2 and m_0, f_D^0 differs slightly in both cases. This is counteracted by a proper setting of the C-field whereat simultaneously the other known frequency shifts due to second order doppler effect and the cavity phase difference can be compensated by this provision.

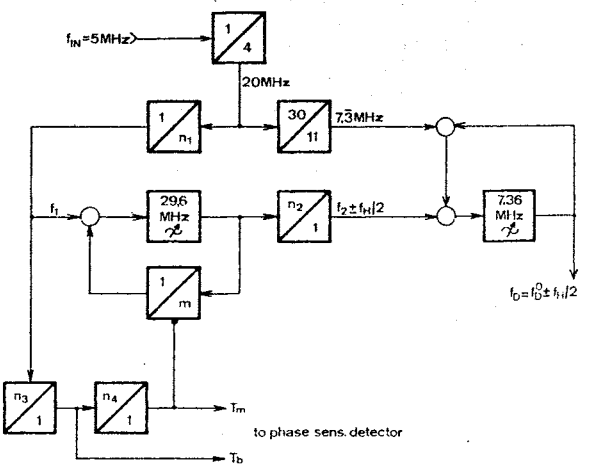


Fig. 2: Principle of 7,36 MHz-synthesis

clock designation	n_1	n_2	m_0	f_1[Hz]	f_2[Hz]	f_0°[Hz]	$\frac{f_H}{2}$[Hz]	requested Zeeman frequency [kHz]
CS1, CS2	382	850	566	52356	34893	7368227.08	30.8	≈58
CS3, CS4	500	850	741	40000	34894	7368227.45	23.5	≈54

Parameter of the 7,36 MHz-synthesis suitable for use with the PTB's primary clocks CS1-CS4.

Table 1

4. Discussion of transients

4.1 Transient behaviour of the frequency synthesis

The transient time of the output frequency f_s is dominated by that of the 7,36 MHz phase-lock loop. A PIT_1 controller is used here and the servo-loop is designed following the principle of symmetric optimum /4/. Therefore after switching from $f_s^\circ - f_H/2$ to $f_s^\circ + f_H/2$ the deviation $\Delta f_s(t)$ between the instantaneous frequency $f_s(t)$ and the stationary value $f_s^\circ + f_H/2$ follows the relation

$$\Delta f_s = f_H\left(\frac{D}{1-D}\exp(-\omega_T t) - \frac{\exp(-D\omega_T t)}{1-D}\cos(\sqrt{1-D^2}\,\omega_T t)\right)$$

D was chosen as $1/\sqrt{2}$, so that

$$\frac{\Delta f_s}{f_H} = 2{,}41\exp(-\omega_T t) - 3{,}41\exp(-\omega_T t/\sqrt{2})\cos(\omega_T t/\sqrt{2}) \qquad (5)$$

In fig. 3 the transient process of $\Delta f_s(t)$ is illustrated after switching the frequency. The transient frequency $f_T = \omega_T/2\pi$ has been chosen to f_T=300 Hz.

4.2 Transient in the atomic beam signal

At first the occurence of transients in the beam signal shall be illustrated in the case of instantaneous switching of the frequency f_s. A monokinetic beam is assumed and the flight time through the irradiation sections of the resonator is neglected. Then, from the atoms being within the cavity at the moment of switching f_s an transient impulse is formed which can be interpreted as generated by a linear sweep of f_s with sweep rate 1/T over the centre of the Ramsey curve, from $f_c + f_H/2$ to $f_c - f_H/2$ /3/. The impulse is delayed by T_1 referred to the moment of switching, where T_1 is the time the atoms need for the passage from the resonator to the detector. In fig. 4 this idealized transient signal is illustrated in the case f_H=W.

Due to the transient response of the generated frequency f_s, the beam transient is modified as indicated in fig. 4. Here the rf-transient from fig. 3 has been assumed.

Individual devices in the phaselock loops exhibit non linear behaviour (mixer, varacter-diodes as tuning elements of the VCO's). The rf-transient process will therefore be different for both directions of frequency switching. Thus, the two transient impulses of the beam signal occuring per modulation period T_m will also differ from each other and frequency errors are the consequence. Frequency errors of $2{,}5\cdot10^{-13}$ could be measured for the system in use with f_H=3W. They are caused by non-linear behaviour of the f_2-Mixer in the 7,36 MHz phaselock loop (s. fig. 2).

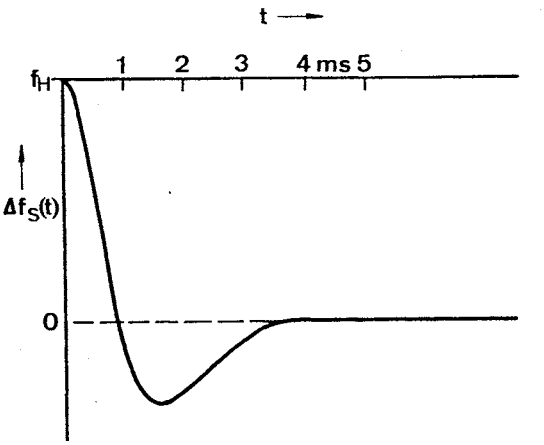


Fig. 3: Transient response $\Delta f_s(t) = f_s(t) - f_s(\infty)$ after switching f_s from $f_s^\circ + f_H/2$ to $f_s^\circ - f_H/2$

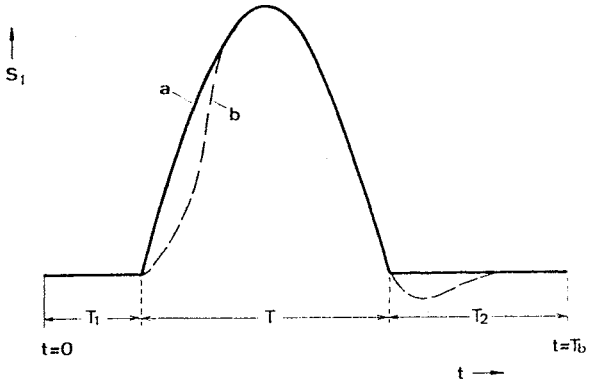


Fig. 4: Transient impulse of beam signal s_1 with frequency f_s switched from $f_c - W/2$ to $f_c + W/2$ in case of
a) instantaneous rf-switching
b) rf-switching with transient behaviour of Fig. 3

By use of a blanking interval of duration T_b,

$$T_b = T + T_1 + T_2,$$

the transients in the signal s_1 (s. fig. 1a, b) are suppressed. Here T_2, explained in fig. 4, covers the effects of the rf-transients and will be calculated in the following:

The atoms which arrive at the detector at time t during the effective measurement interval T_a (it lasts from $t=T_b$ to $t=T_b+T_a$) contain an information corresponding to an error $\Delta f(t)$ which is equal to the mean frequency difference between $f_s(t)$ and the stationary value $f_c \pm f_H/2$

$$\Delta f(t) = \frac{1}{T} \int_{t-(T+T_1)}^{t-T_1} \Delta f_s(t')dt'$$

which is present during their flight through the resonator.

The resultant error $\langle\Delta f\rangle$ in the determination of the true line centre is found as the average of $\Delta f(t)$ over T_a

$$\langle\Delta f\rangle = \frac{1}{T_a} \int_{T_b}^{T_b+T_a} \left(\frac{1}{T} \int_{t-(T+T_1)}^{t-T_1} \Delta f_2(t')dt'\right)dt \qquad (6)$$

where $\Delta f_s(t')$ is inserted from (5).

The calculation of (6) is tedious but straightforward. The solution is

$$\langle\Delta f\rangle/f_H = 2,41\exp(-\omega_T T_2)(1-\exp(-\omega_T T))(1-\exp(-\omega_T T_a))/(\omega_T^2 T T_a) -$$

$$-\sqrt{2}\exp(-\omega_T(T+T_2)/\sqrt{2})(\sin(\omega_T(T_a+T+T_2)/\sqrt{2})-\exp(-\omega_T T_a/\sqrt{2})\sin(\omega_T(T+T_2)/\sqrt{2}))/\omega_T T_a +$$

$$+\sqrt{2}\exp(-\omega_T T_2/\sqrt{2})(\sin(\omega_T(T_a+T_2)/\sqrt{2})-\exp(-\omega_T T_a/\sqrt{2})\sin(\omega_T T_2/\sqrt{2}))/\omega_T T_a$$

Considering that $\omega_T T_a \approx 180$ and $\omega_T T$, $\omega_T T_2 \approx 10$ and $\omega_T = 2\pi \cdot 300$ Hz some approximations are allowed. The final useful result is

$$\langle\Delta f\rangle/f_H \approx \sqrt{2}\exp(-\omega_T T_2/\sqrt{2})/(\omega_T T_a). \quad (7)$$

With $f_H = 3W$, $T_a \approx 100$ ms and the demand $\langle\Delta f\rangle/f_0 < 10^{-15}$ the relation $\omega_T T_2 > 16,9$ has to be fulfilled. It is $\omega_T = 2\pi \cdot 300$ Hz so that T_2 must be chosen as $T_2 > 9$ ms. In the case of CS1 and CS2 where $T = 9$ ms and $T_1 = 4,5$ ms the duration of the blanking interval finally has to be

$$T_b > 22,5 \text{ ms}$$

If the time constant of the detector and the electrometer are additionally considered, T_b has to be prolonged to about 30 ms. It is worth to note, however, that the calculated demand for T_b is derived for the single switching process. It is neglected that the effects due to the two consecutive transients from the two opposite switching processes per modulation period will cancel out to some extent.

Finally it is interesting to note that the signal to noise ratio is effectively not reduced although the measurement time is reduced from $T_m/2$ to $T_a = (T_m/2)-T_b$. The reason is as follows: During the time interval when the transient impuls occurs (see fig. 4) the beam signal does not contain information about the frequency difference $f_s°-f_c$. Therefore an exact suppression of the transient does not reduce the useful signal, but even increases the signal to noise ratio by suppression of the shot noise of the beam signal during this time interval.

5. Discussion of phasemodulation-sidebands

5.1 Minimum sideband-spacing

The output-signal s(t) at frequency f_s is produced in the frequency synthesis system by nonlinear processes from the input signal at frequency f_{IN}, $f_{IN} = 5$ MHz. Therefore s(t) cannot be a pure sine-wave signal at f_s, but contains some distortions, which appear periodically with period $T_p = 1/f_p$, where f_p is the common subharmonic frequency of f_s and f_{IN}. The value of f_p can be calculated from the ratio

$$\frac{f_s}{f_{IN}} = \frac{k}{l}, \text{ where k, l are integers without common divisor}$$

to be $f_p = \frac{f_s}{k} = \frac{f_{IN}}{l}$ (8)

The signal s(t) can be expressed as Fourier-series:

$$s(t) = \sum_{m=1}^{\infty} s_m \cos(m\omega_p t+\alpha_m) \qquad (9)$$

with $\omega_p = 2\pi f_p = 2\pi/T_p$

Here the component s_k at $f_s = Kf_p$ is the desired component whose magnitude usually is by far the largest of all.

The components at the frequency separation $\pm nf_p$ ($n = 1,2...$) around the carrier f_s are to be interpreted as phasemodulation-sidebands (amplitude-modulations are principally excluded in a phaselock loop synthesizer device). The phasemodulation $\varphi(t)$ corresponding to the sidebands can be written as

$$\phi(t) = \sum_{n=1}^{\infty} \phi_n \cos(n\omega_p t+\beta_n) \qquad (10)$$

and s(t) can be expressed as

$$s(t) = s_0 \cos(2\pi f_s t + \phi(t)) \qquad (11)$$

which is the preferable representation.

With (3) inserted into (1) f_s becomes

$$f_s = \left[1840 - \left(\frac{22}{15} + \frac{4m}{n_1 n_2}\right)\right] f_{IN}$$

and $$\frac{f_s}{f_{IN}} = \frac{27578\ n_1 n_2 + 60m}{15\ n_1 n_2} = \frac{k}{l}$$

For all values of f_s, f_p can be calculated by selecting the corresponding value of m.

It follows that

$$f_p = \frac{n}{3}\frac{f_1}{n_2} = \frac{n}{3} f_{H1} \quad ; \quad f_{H1} \approx W \qquad (12)$$

The integer n depends on the chosen value of f_s (see table 2)

f_s	$f_s^0 - \frac{W}{2}$	$f_s^0 + \frac{W}{2}$	$f_2^0 - \frac{3}{2}W$	$f_s^0 + \frac{3}{2}W$
CS1/2	$n=\underset{(3)}{1}$	$n=\underset{(6)}{1}$	$n=\underset{(12)}{1}$	$n=\underset{(15)}{5}$
CS3/4	$n=\underset{(6)}{2}$	$n=\underset{(3)}{1}$	$n=\underset{(3)}{1}$	$n=\underset{(60)}{10}$

Table 2: Values of n from which the sideband separation can be inferred in case of the synthesized frequencies f_s for the PTB clocks. The values in brackets are related to the relevant sidebands (see section 5.2)

5.2 The relevant sidebands

The sideband separation f_p calculated above gives the raster on the frequency scale where sidebands can be expected, but nothing is said about their magnitude. In fact, most of them can be neglected. The sidebands close to the carrier ($f_s \pm 100$ KHz) can be produced only by the 7,36 MHz-synthesis. Here two sources of disturbancies lead to significant phase modulations.

1) The mixing process of the two f_2-frequencies (fig. 2) gives origin to a direct phasemodulation of the 7,36 MHz with a modulation frequency of $2 f_2 \approx 70$ KHz. The corresponding sidebands can be held below -100 dBc.
2) The frequency f_2 coming out from the n_2-divider is phasemodulated itself (fig. 2). Due to the mixing of the two f_1-frequencies, the 29,6 MHz will be disturbed by phase modulations $\varphi(t)$ with $2f_1$ as modulation frequency and $\pm 2f_1$-sidebands arise being in the order of -55 dBc. The frequency f_s is obtained by dividing the 29,6 MHz by $n_2 = 850$; therefore, the resulting phase modulations of f_2 will be $\varphi(t)/850$ with the modulation frequency f_p equal to the common subharmonic of $2 f_1$ and f_2. These phase modulations fed into the 7,36 MHz phaselock loop with its specific filter effect result finally in some phase modulation sidebands of the order of -120 dBc situated very close to the carrier f_s and spaced by f_p(12), whereat for n the values in brackets have to be taken from table 2.

5.3 Effect of phase modulation-sidebands

In section 5.1 and 5.2 it has been demonstrated that phase modulations in the signal s at $f_s = f_s° + f_H/2$ are inherently different from those at $f_s = f_s° - f_H/2$. Thus, the effect on the determined line centre frequency has to be calculated separately. It is sufficient here to limit the discussion to one of the two cases.

Equation (10) describes the phase modulation $\varphi(t)$ when the frequency $f_s(t)$ has reached its stationary value $f_s° + f_H/2$. Therefore, it is valid during the effective measurement interval T_a. In the following it is assumed that t=0 always coincides with the begin of each of the measurement intervals. After each switching at $t = -T_b$ the excursion of phase with time $\varphi(t)$ will be started again. In general, the initial conditions expressed by the angles β_n in (10) will be different from cycle to cycle. But as the period $T_m = 1/f_m$ of the square wave modulation of f_s and the synthesized frequency f_s are both derived from $f_{IN} = 5$ MHz, after some modulation periods T_m the same phase modulation scheme repeats. Therefore, it cannot be assumed a zero net effect on the determined centre frequency by averaging over a long time interval. An extreme situation appears if the modulation period T_m is chosen as an integer multiple of $T_p = 1/f_p$, the period of the phase modulation. Then the phase modulation scheme repeats yet after one modulation period T_m. This situation will be considered for a discussion of the frequency errors.

Some suppositions are made

1) The flight time through the final sections of the resonator is neglected.
2) The phase modulation $\varphi(t)$ is small.
3) Only for simplicity, a monokinetic beam is assumed.

The first supposition allows the conclusion, that the atoms leaving the resonator at t will have detected a phase deviation of

$$\Delta\phi(t) = \phi(t) - \phi(t-T) \qquad (13)$$

referred to φ_0, where $\varphi_0 = 2\pi(f_s - f_c)T$ is the phase difference observed by the atoms in the case without phase modulations.

The second supposition means that $\Delta\varphi(t)$ will be transformed linearly into beam signal variations which will be processed in the servo loop in a frequency error signal of

$$\Delta f(t) = \frac{\phi(t) - \phi(t-T)}{2\pi T} \qquad (14)$$

Representative for all components of the Fourier-series of $\varphi(t)$ (see (10)) the effect of the nth component $\varphi_n(t) = \varphi_n \cos(n\omega_p t + \beta_n)$ will be calculated now:

$$\Delta f_n(t) = \left(\phi_n(t) - \phi_n(t-T)\right)/2\pi T$$

or

$$\Delta f_n(t) = \frac{-\phi_n \sin(n\omega_p T/2)}{\pi T} \sin\left(n\omega_p t - n\omega_p T/2 + \beta_n\right) \quad (15)$$

$$\text{Then } \langle \Delta f_n \rangle_{T_a} = \frac{1}{T_a} \int_0^{T_a} \Delta f_n(t)\,dt \quad (16)$$

is the frequency error produced by all atoms leaving the resonator during the measurement interval. It is obtained

$$\langle \Delta f_n \rangle_{T_a} = \frac{-\phi_n}{\pi T_a} \frac{\sin(n\omega_p T/2)}{n\omega_p T/2} \sin(n\omega_p T_a/2) \cdot \sin(n\omega_p T_a/2 - n\omega_p T/2 + \beta_n) \quad (17)$$

This equation may be valid for n=1...5 (see supposition 1). If the averaging time T_a is chosen an integer multiple of T_p

$$T_a = KT_p = K/f_p;\ K = 1, 2 \ldots \quad (18)$$

the effect of the phase modulation $\varphi(t)$ is averaged out to zero. This can be concluded from (17), because of each n $\langle \Delta f_n \rangle_{Ta}$ results to be zero. This favourable result remains valid even if the suppositions 1 and 3 are dropped.

For controlling eq. (17) the frequency synthesis circuits have been modified in a way that sidebands with $\pm nf_{H1}$ (n = 1, 2 ...; these sideband separations are equal to those of the relevant sidebands, see section 5.2) artificially are produced during one half of the modulation period T_m.

This has been done by feeding the frequency f_1 additionally to the f_2-mixer (see fig. 2). T_m has been chosen to $T_m = 16 \cdot T_p = 16/f_{H1}$ with the consequence - as mentioned before - that the same phase modulation scheme repeats each T_m. The measurement interval $T_a = (T_m/2) - T_b$ could be varied by changing $T_b = r/f_{H1}$ with r variable between 1 and 3,5. The sidebands observed around f_s have been chosen to be

-50 dBc for $\pm f_{H1}$
-62 dBc for $\pm 2f_{H1}$
-63 dBc for $\pm 3f_{H1}$
-70 dBc for $\pm nf_{H1}$; n≥3

If one restricts to the $\pm f_{H1}$ sidebands and takes into account (by multiplication eq.(17) with the factor 1/2) that the introduced phase modulations have been present only during one half of the modulation period, the relative frequency error $\Delta f/f_0$ turns out to be

$$\Delta f/f_0 = \frac{-\phi_1}{\pi T_A f_0} \sin(\pi f_{H1} T_a) \sin(\pi f_{H1} T_a - \pi/2 + \beta_1) \quad (19)$$

where $\omega_p = 2\pi f_{H1}$ and $f_{H1} = W = 1/2T$ have been used, or

$$\Delta f/f_0 = -5{,}25 \cdot 10^{-13} \left(\frac{8}{8-r}\right) \sin r\pi \cos(r\pi - \beta_1) \quad (20)$$

when $T_a = (8-r)/f_{H1}$ and $f_{H1} = 61$ Hz have been set into (20).

In fig. 5, $\Delta f/f_0$ has been plotted for several β. The results found with CS2 during some weeks of experiments are indicated by dots. They fit well to the calculated curve with $\beta_1 = 60°$. The value of $\Delta f/f_0$ for $T_b = 2/f_{H1}$, where zero net effect is expected, has been carefully examined comparing the measurement results with the CS2 output frequency when no additional sidebands are introduced. The "zero" effect has been confirmed with a statistical error of $1{,}5 \cdot 10^{-14}$ (1σ).

The worst case is present if $\beta_1 = \pm\pi/2$ and $r = (2K-1)/2$ with K=1, 2, ... A realistic choice for r could be r=2,5. Then $\Delta f/f_0$ is found to be $\Delta f/f_0 = 7{,}6 \cdot 10^{-13}$ with the strength of the $\pm f_{H1}$ sidebands as given above. Even in this worst case, however, in the regular mode of operation of the electronics system, the frequency error is smaller than 10^{-15} since the sidebands are -120 dB below the carrier.

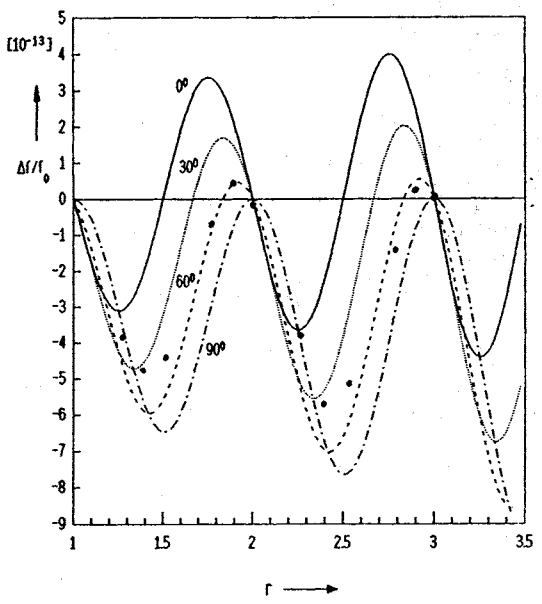


Fig: 5: Frequency error $\Delta f(r)/f_0$ caused by $\pm f_{H1}$- sidebands with -50dBc magnitude
lines: theoretical results, calculated with (20),
dots: experimental results, measured with the CS2,
$r = T_b f_{H1}$

The discussion conducted up to now has started from the extreme situation that the phase modulation scheme repeats with each cycle of the square wave modulation. A different choice of T_m by selecting the division factors n_3 and n_4 properly (see fig. 2) may generate a repetition period of the phase modulation scheme of some T_m over which the effect of the phase modulation can be smeared out better as this is possible in an interval with duration T_m. Therefore a characteristic for $\Delta f/f_0$ is to be expected with less pronounced extreme values as those from fig. 5. But, nevertheless, the choice of $T_a = KT_p$ (K integer) will be prefered here, too, being this the only sure guarantee for a zero net effect of the phase modulations.

In our case n_3 and n_4 have been chosen to

$$n_3 = 4n_2 \qquad n_4 = 4$$

Then $T_m=16/f_{H1}$, $T_b=2/f_{H1}$ and $T_a=6/f_{H1}$ are found, being $T_a=2T_p$ with $T_p=3/f_{H1}$ (see section 5.1) therefore an integer multiple of T_p. The scheme of the phase modulations will repeat after $3T_m$.

6. Conclusion

The electronic system to be used with the four PTB cesium clocks should not contribute to the overall uncertainty budget by more than 10^{-15} in relative units. This demand is fulfilled by the system presented in this paper. Here, the decision to use a square wave frequency modulation for the determination of the line centre has been proven as a favourable one. A clear separation of the modulation period in a sequence of time intervals can be done. After switching the frequency, the transient process of the beam signal can be suppressed during the first interval. In this way an error is suppressed to much below 10^{-15} which without transient suppression would remain in the order of some 10^{-13}. This error is known as second harmonic distortion error in cesium frequency standards where a sinusoidal phase modulation is used and the transient suppression can not be applied. In the second interval - the so-called measurement interval - the signal is processed for the determination of the line centre. It is disturbed by the phase modulations originating from the mechanism of the frequency synthesis which in worst case situations could give rise to an error of about $1 \cdot 10^{-15}$. But when the length of the measurement interval is properly chosen, an averaging of the effects of the phase modulations takes place which reduces the contribution to the uncertainty budget to zero.

References

/1/ "Comptes rendus de la treizième Conférence Générale des Poids et Mesures" (Paris October 1967). Paris: Gauthier-Villars 1968, p. 103.

/2/ N.F. Ramsey, Rev.Mod.Phys., Vol.62, 1990, p. 541-552.

/3/ J. Vanier, C. Audoin, The Quantum Physics of Atomic Frequency Standards, Vol. 2, Adam Hilger, Bristol and Philadelphia, 1989, p. 732.

/4/ W. Leonhard, Einführung in die Regelungstechnik, Vieweg, Braunschweig, p. 157-160.

Acknowledgement

The author expresses his thank to Ch. Richter and L. Rohbeck for their contribution in realizing the frequency synthesis and A. Bauch for helpful discussion during the preparation of this paper.

Session 4B Number 34

THE OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD AT PEKING UNIVERSITY

WANG Yi-Qiu*
YANG Dong-Hai**

* Department of Radio-Electronics
Peking University, Beijing, 100871, P.R. China
** Beijing Institute of Modern Physics, Peking University

ABSTRACT

An optically pumped cesium beam frequency standard with bias-saturated absorption lock of laser diodes has been built up at Peking University. The frequency stability of $1.2*10^{-11}/\tau^{1/2}$ was obtained. The new idea of sharp angle incidence of probing laser beam has been tested experimentally and fairly good results were achieved. Further improvements are discussed.

I. INTRODUCTION

The research on the optically pumped cesium beam frequency standard at Peking University has been undertaken since 1983. Works done at earlier stage included : the theoretical studies on searching for an optimum scheme for optical pumping and detection [1,2], the tests of the performance and development of frequency stabilization and noise suppression techniques for semiconductor laser diodes, building up an experimental set-up and observation of the clock signals, and finally, the realization of a primitive experimental frequency standard [3,4]. Main part of those work was reported in [4].

During recent two years we have significantly improved the experimental set-up and much better results acieved : the signal to noise ratio for the Ramsey pattern of clock transition reached from < 300 to 6000 in 1 Hz bandwidth and the frequency stability of the standard was $1.2*10^{-11}/\tau^{1/2}$ (including the instability of the measuring standard HP 5061A) [5]. These results have been obtained by employing some important measures including bias-saturated absorption lock for noise reduction and frequency stabilization of laser diodes and use of cycling transition for optical detection of clock signals. In section II these measures will be described in detail. Last year we proposed a new method to reduce the influence of laser frequency noise on the clock performance. It has been tested experimentally and the preliminary results, which will be presented in section III were attractive. Further improvements are possible and being under way.

II. PRELIMINARY RESULTS OF THE FREQUENCY STANDARD WITH BIAS LOCK AND CYCLING TRANSITION DETECTION

The experimental set-up is shown in Fig. 1 and is described in detail in [5]. The U-shape cavity drift length is 168.7 mm. In experiments the beam current was

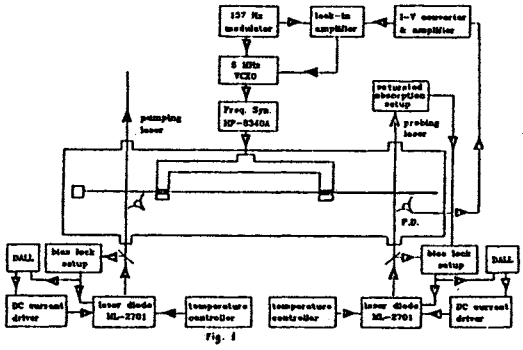


Fig. 1

Diagram of the experimental set-up with bias-saturated absorption lock.

about 10^{10} atoms/sec in 12 mm^2 cross section at detection region while the oven temperature was kept about 100° C.

Two TJS single mode GaAlAs laser diodes (ML-2701) with linewidth about 25 MHz and output power about 8 mW were used for optical pumping and probing respectively. The pumping laser frequency was locked to F = 4, F' = 4 transition, while the frequency of probing laser was locked to F = 4, F' = 5 cycling transition [6]. With this two laser scheme the strongest lock signal can be obtained.

The bias lock technique was employed for locking the laser frequency to suitable atomic transition. As shown in Fig. 2, the side slope of cesium gas cell absorption line contour was used as frequency discriminator to reduce the low frequency FM noise of diode laser and the time constant of few milliseconds was chosen for the servo-loop in this scheme. The gas cell was temperature stabilized within 0.05°C and the centre frequency of laser diode was fixed by setting a proper bias with a very stable voltage source. For probing laser this bias lock only is not enough because the F = 4, F' = 5 transition is located near the top of the gas aborption line and the requirement to the frequency stability of probing laser is more strict. So a second loop locking the laser frequency at the saturated absorption peak of F = 4, F' = 5 transition [7] was added. The bias lock loop operates as a fast coarse lock loop,

while the saturated absorption loop serves as a slow fine lock. This technique we call the bias-saturated absorption lock.

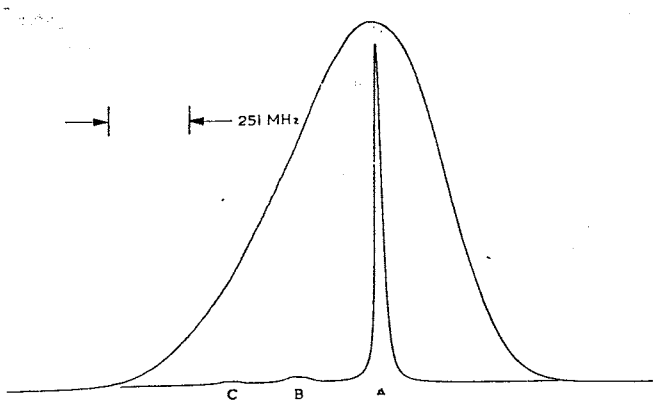


Fig. 2

The gas cell absorption line contour and three fluorescence lines for F = 4, F' = 5, 4, 3 (corresponding to A, B, C, respectively) D_2 hyperfine transitions of cesium beam. The side slope of absorption contour is used for frequency discriminating.

For long term operation a digital auxiliary lock loop (DALL) was added to the above described frequency locking scheme in order to overcome the frequency drift problem caused by change of environment condition. An electronic circuit automatically adjusts stepwise the injection current to a new value when laser frequency drifts beyond some limit and the loop can hardly function. This scheme equivalently enlarges the dynamic range of frequency lock loop. With DALL added the laser frequency can be satisfactorily locked for more than 100 hours.

With the set-up described above a derivative Ramsey resonance signal of S/N about 6000 (1 Hz bandwidth) was obtained. The Allan variance of the frequency stability was $1.2*10^{-11}/\tau^{1/2}$ (Fig. 3) compared with a commercial cesium standard (HP-5061A). The contribution of its instability ($5*10^{-12}$) was included.

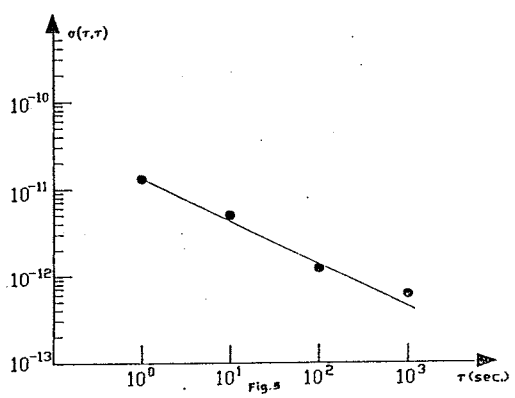


Fig. 3

The measured frequency stability of the optically pumped cesium beam frequency standard in Allan variance compared with HP-5061A.

III. THE STANDARD WITH SHARP ANGLE INCIDENCE PROBING LASER BEAM

In an optically pumped atomic beam frequency standard the effect of laser frequency noise always limits the signal to noise ratio of clock signal. The problem is less severe for the pumping laser because usually the transition is highly saturated and the frequency response of pumping transitions becomes relatively flat. At detection region, however, the probing cycling transition used is difficult to saturate. Therefore, sharp response of clock resonance signal to the frequency variation of probing laser will deteriorate the signal to noise ratio of Ramsey pattern. Besides the sophisticated frequency stabilization and line narrowing techniques generally adopted, an alternative way to solve this problem is possible. It is to intentionally broaden the probing transition line. This idea was realized by making a sharp angle of probing incidence beam with atomic beam [8]. In our case the probing transition line is Doppler broadened from 10 MHz (natural linewidth) to 125 MHz (with an angle 66.5° and oven temperature 100°C). This angle is chosen for conveniance of locking the shifted cycling transition frequency to the one of saturated absorption line peaks.

The first such an experiment was done at National Institute of Metrology of China (NIM) using a somewhat modified long cesium tube of cavity length 3.68 m [9]. The diagram of experimental arrangement is shown in Fig. 4. The cesium oven was operated at about 150° C

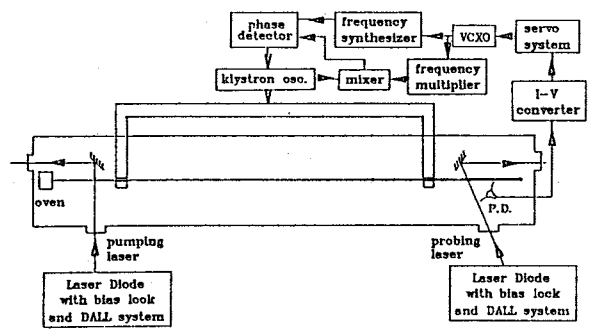


Fig 4

The experimental set-up with sharp angle incidence of probing laser beam.

and the angle between probing laser beam and atomic beam was 67.4 degrees. Both pumping and probing lasers was only bias locked but with DALL in operation. The former was locked to the F = 4, F' = 4 transition and the latter to the shifted F = 4, F' = 5 cycling transition. The derivative Ramsey pattern is shown in Fig.5, where much more peaks appear than that of the ordinary perpendicular case is due to the velocity selection coming from Doppler effect. The linewidth of central peak is 47 Hz, while the signal to noise ratio is about 1000 in 1 Hz bandwidth. S/N could be improved significantly if the cesium tube could be modified satisfactorily and the experimental condition would be

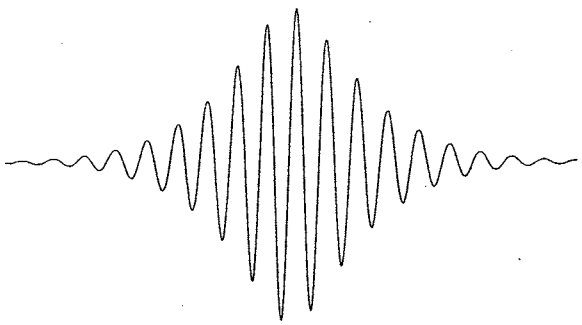


Fig. 5

The derivative Ramsey pattern with sharp angle incidence of probing laser beam. The width of central peak is 47 Hz.

arranged properly so that the scattered light background and stray pick-up level would be much lower. Nevertheless, this primitive test has already proved the idea that the effect of probing laser FM noise on the clock signal can be reduced by widening the atomic transition spectral line. This is essential because the sophisticated laser line narrowing techniques of either optical of electrical feedback require strict environment conditions which can not be satisfied everywhere.

Recently, we have built a new cesium tube with optical windows specially for the sharp angle probing laser beam incidence. The inner parts remain as the same as described in section II. Preliminary experiment gave encouraging results. The signal to noise ratio of Ramsey pattern increased to 8000 (Fig. 6, in 1 Hz bandwidth). Many peak configuration of the resonance pattern is in agreement with theoretical expectation provided that velocity selection gives relative width of velocity window $\Delta v/v$ being about 20 %, corresponding to the ratio of laser linewidth about 25 MHz to the Doppler broadened atomic transition linewidth 125 MHz. With this results we may expect better frequency stability of the standard. The work is being continued.

Furthermore, we have indicated that the narrow velocity distribution of atoms contributing to the clock resonance signal will improve also the accuracy of the standard because better evaluation of the cavity phase dependent shift and second Doppler shift can be made in this case [10]. The loss of signal intensity by velocity selection of atoms in beam will be compensated by reducing the noise coming from the frequency variation of probing laser.

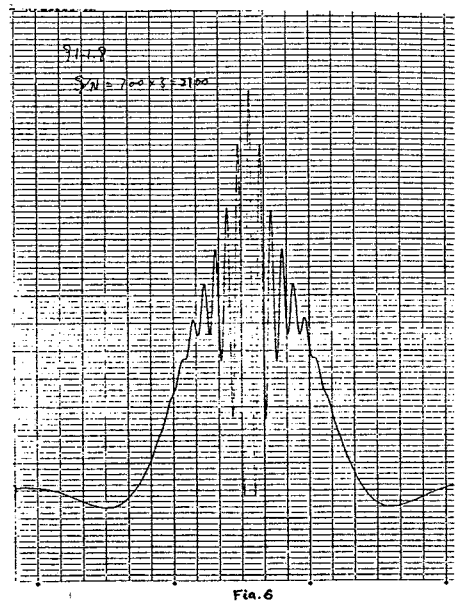


Fig.6

The Ramsey resonance pattern with sharp angle incidence of probing laser beam, for a small Cs tube.

IV. FURTHER IMPROVEMENTS

The presented above techniques provided a very simple, compact, reliable and easy to carry out method for moderate suppression of FM noise (- 20 dB below 200 Hz) and frequency stabilization of laser diodes. The arrangement of sharp angle incidence of probing laser beam further reduces the effect of laser frequency noise on the clock signal. The preliminary experiments proved the feasibility of this technique. However, all these experiments were conducted with a cesium tube, the inner parts of which were originally designed for magnetic state selection, and the environment conditions were primitive. Therefore many measures to improve the experiments have to be undertaken.

1. A laboratory with better mechanical and temperature stability and with better grounding will ensure further increase the signal to noise ratio of clock resonance.

2. The cavity C-field and magnetic shieldings structure must be specially designed and reconstructed to provide better magnetic homogeneity and polarization purity of laser light.

3. The electronic circuits and parameters of operation both for laser frequency lock and for frequency control of standard could be optimized. For example, the gain and time constant in bias lock loop might be adjusted in order to get even better reduction of FM noise of diode laser.

4. Long term and systematic studies of performance of the frequency standard must be carried out in near future in order to finally evaluate the stability and accuracy of the standard and to further improve the experimental arrangement.

ACKNOWLEDGEMENT

The work was supported in part by Natural Science Foundation, Doctorate Foundation of State Education Commission and former State Bureau of Metrology of P.R. China.

REFERENCES

1. WU Xinxin, XU Yueting, TIAN Kunyu, YAO Shutong, XIE Linzhen, WANG Yiqiu, Acta Metrologia Sinica 6 15 (1985)
2. WU Xinxin, SHANG Songquan, XIE Linzhen, WANG Yiqiu, Acta Metrologia Sinica 7 269 (1986)
3. WANG Depei, XIE Linzhen, WUANG Yiqiu, GUO Shu jun, Chinese Phys. Lett. 5 97 (1988)
4. WANG Yiqiu, XIE Linzhen, WANG Depei, YAO Shutong, Frequency Standards and Metrology, Proc. of 4th Symposium, Ed. A. de Marchi, Springer-Verlag 1988 p. 126
5. YANG D-H, WANG Y-Q, to be published in IEEE Trans. IM
6. WANG D-H, WANG Y-Q, Optics Comm. 73 285 (1989)
7. YANG D-H, WANG Y-Q, Optics Comm. 74 54 (1989)
8. YANG D-H, WANG Y-Q, Digest CPEM'90 1990 p. 244
   YANG D-H, WANG Y-Q, METCALF H. to be published in IEEE Trans. IM
9. YANG D-H, LI E-X, CHEN L-M, HUANG B-Y, WANG Y-Q, to be published in Optics Comm.
10. YANG D-H, WANG Y-Q, to be published in Optics Comm.

SESSION 5B

TIME AND FREQUENCY III

Chairman : G. BUSCA

Session 5B Number 35

ATOMIC FREQUENCY STANDARDS FOR DIFFERENT SYSTEMS AND AIDS OF POSITIONING AND TIMING SERVICE : STATE-OF-THE ART OF DEVELOPMENTS AND PROSPECTS FOR INTERNATIONAL COOPERATION

Yuri G. Gouzhva, Arvid G. Gevorkyan, Alexey M. Moskovsky

Leningrad Scientific Research Radiotechnical Institute (LSRRI)
2, Rastrelli square, Leningrad, 193124, USSR

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 atomic frequency standards (AFS). The general coordination and creation of development programs for PTService systems and aids are being performed by "Interradionavigation" Committee of the Soviet Union. LSRRI is a leading organization in designing, production and implementation of a series of PTService 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 system, ground-based long-range pulse-phase "Chayka" radionavigation system and also State System of Common Time and Standard Frequencies of the USSR. 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 PTService 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 types being intended for the use in subsystems of different ground- and space-based complexes have been 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 and vehicle equipement as well as for onboard space vehicle equipment with accuracy performances at the level of 10^{-14} and increased operational life.

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^{-11} to 1.10^{-13}.

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 metrological performances, researches of monitoring and accounting methods of those factors in order to exclude its infuence 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 synchronisation system comprising the main synchronizer, onboard satellite frequency and time standard, facilities for satellite orbit monitoring and software for prediction of onboard clock rate. The latest achievment in the field of AFS development for the space equipment is an BFS with the daily instability of $\sim 10^{-13}$. For the main synchronizer, one usea ABFS with the instability of 1.10^{-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 ABFS 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 on Bread Board with beam pumping through one laser. The predicted value are

$F = 500$ and $\sigma(\tau) = 2.10^{-13} \cdot \tau^{-1/2}$.

The improvement of performances is realized on the basis of utilization of two laser's 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 foreing 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 ABFS - 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 5.10^{-14} ;

b) Joint realization of prodicted accuracy performances of onboard ABFS 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 onboard GLONASS/GPS space vehicles' equipment as well as for space vehicles of a new generation.

Session 5B Number 36

ESA Center of Excellence in Time and Frequency

S. J. Feltham
European Space Research and Technology Centre
2200 AG Noordwijk, The Netherlands

Abstract

Nearly all the Agency's programs including both ground and space segments require some form of stable timing device. The main difference between the various applications is the specified degree of stability.

The European Space Agency (ESA) is currently working on 3 large programs:

- Columbus Space Station
- The Space Plane Hermes
- Ariane 5

There are also a number of other projects within the various directorates, for example:

- Science; Cassini, Vulcan
- Earth Observation; Aristoteles
- Communications; Data Relay Satellite

In order to build up expertise within Europe and prepare European industry to meet the technical challenges which will arise from participation in these programs, ESA has established a centre of excellence (CoE) in time and frequency at the Observatory of Neuchatel, Switzerland.

This paper will discuss the aims and objectives of the CoE, outline Agency plans and briefly describe applications of time and frequency within planned Agency programs.

I Introduction

The Agency's purpose as described in its convention is to 'provide for and promote, for exclusively peaceful purposes, cooperation among European states in space research and technology with a view to their being used for scientific purposes and space applications systems. One area where the Agency foresees a need to increase expertise is in the field of time and frequency.

Nearly all the Agency programs including both ground and space segments require some form of stable timing device. The main difference between the various applications is the specified degree of stability.

ESA has been considering setting up a Time and Frequency laboratory at European Space Technology Centre (ESTEC). However it was eventually decided that the ESA activity would best serve the European space community if it were placed with industry under a CoE umbrella. This is intended to effect a long term improvement in the field of time and frequency.

The ultimate goal is to enable European industry to compete effectively in the world market as per the ESA convention.

II Definition of a CoE

The definition of a CoE as given in [1], is as follows:

A CoE is any entity which has proved to have an outstanding technical and commercial competence in a specific discipline.

The 5 selection criteria are as follows:

1. Technical; first and foremost a CoE must have a proven record of technical competence in the field concerned.
2. Programmatic; the expertise provided by the CoE must fulfill a long term need of the Agency.
3. Commercial; The CoE must have a proven record of successfully performing work at competitive prices. In addition the CoE must have demonstrated an ability to work successfully with other European firms and to undertake to continue to work with such firms as may be directed by ESA.
4. Industrial Policy; A CoE should normally only be recognised when no other equivalent entity exists in a small or low return country, in order to avoid conflict of interest and only when adequate competition exists in other member states in order to avoid the creation of a monopoly.
5. National Policy; A CoE should be supported by its national authorities in order to ensure the maximum technological improvement through both national and ESA investments.

The title CoE is conferred on qualifying organisations for an initial period of 5 years. The performances of the CoE will be regularly reviewed by the Agency's technical officers.

III Terms of Reference for an ESA CoE in Time and Frequency

The main objectives of an ESA CoE in Time and Frequency are:

- Establishment of a technical base for the realisation of space hardware with particular emphasis on the minimisation of mass, volume and dc power requirements.
- Fundamental research and development of frequency standards.
- Maintenance of a measurement system capable of performing state-of-the-Art frequency stability measurements, either on site or at customer premises.

- Support to ESA and industry in the preparation of program proposals and evaluations.
- General awareness and technical support in the domain of time and frequency including such areas as;
  1. GPS and GLONASS activities
  2. Time dissemination
  3. Gravitation and relative aspects

The major emphasis of the CoE will be directed towards the stimulation of fundamental research and development of atomic frequency standards. At the moment Rubidium, Cesium and Hydrogen masers are being studied with a view to miniaturisation of mass, volume and dc-power requirements for space applications.

IV The CoE at the Observatory of Neuchatel, Switzerland

The Observatory of Neuchatel (ON) is a non-profit, scientific and engineering organisation supported by the Canton of Neuchatel. ON personnel have a long and extensive experience in the field of time and frequency, particularly in basic research into atomic frequency standards. They are the only commercial European manufacturer of hydrogen masers. In the past ON has been involved with the definition of reference oscillator requirements for several ESA programs. They are currently cooperating with several industrial partners and are building up international contacts.

The Observatory fulfills all the stated selection criteria and are ideally suited to perform the duties of an ESA CoE in time and frequency.

V Outline of Running Activities

V.1 Rubidium Ultra Stable Oscillator (USO) Feasibility Study

The program of work involves the design of a pre-engineering physics package with a possible extension to a full engineering model development. The work follows on from a feasibility study performed under Agency contract, [2]. The performance of the miniaturised USO will be compared against commercially available devices and a hydrogen maser. The miniaturisation potential of the electronics package will be studied and a long duration life test on 3 physics packages initiated.

The main design goals are:

- Short term stability: $5 \times 10^{-12}\tau^{-1/2}$ for $1 \geq \tau \geq 100s$ (goal)
- Overall volume: $0.2l$ with miniature electronics
- DC power consumption: maximum $5W$ (goal $4W$)
- g-sensitivity: less than $4 \times 10^{-12}g^{-1}$
- Frequency drift: 1×10^{-11} per month

V.2 Hydrogen Maser Development

The ON are the only European commercial supplier of hydrogen masers and have to date delivered 13 units. The objective of the maser development is to improve performance and reliability whilst minimising maintenance. To achieve these goals, efforts have been directed towards improving the physical package technology and the electronic control circuitry.

The twin goals are to achieve a stability of better than 1×10^{-15} for a 1000 second averaging time and an unattended operating life of greater than 3 years. With the introduction of the following improvements:

- improved vacuum system
- low phase noise synthesiser
- microprocessor controlled monitoring system with remote interrogation
- automatic source oscillator tuning
- automatic cavity tuning

A long term stability of better than 3×10^{-16} at 10000 seconds averaging should be possible.

The study, production and test of new bulb coating materials and application processes is under evaluation. The quality factor of the atomic resonance is largely determined by the relaxation properties of hydrogen atoms upon collision with the storage bulb wall. A high Q with a long life and reliable wall coating is therefore essential for space applications. The objective is to improve both the initial value and the aging rate of the atomic line quality factor.

The design principles that govern the development of a miniaturised space qualified, hydrogen maser are currently being studied and will lead to a preliminary design.

V.3 Frequency Stability Mobile Measurement System (FSMMS)

An essential task of an ESA sponsored CoE will be the ability to perform high precision frequency stability measurements. In particular the development of a system to perform close to carrier phase noise and short term frequency stability measurements of ultra performing frequency sources is considered a priority.

The State-of-the-Art system will use a high resolution Numerically Controlled Oscillator (NCO) to maintain zero phase between the internal reference oscillator (3 specially selected B-VA crystal oscillators) and the device under test. The system will be adaptable for long term stability measurements using either a Cesium or Hydrogen maser as the frequency reference.

The equipment support structure will be vibration dampened to assure that the reference oscillator specifications will be met under normal environmental conditions.

The system will be transportable to allow measurements to be performed at remote locations, customer premises etc.

The predicted phase noise performance is:

0.5Hz	-140dBC
1Hz	-143dBC
10Hz	-153dBC
> 100Hz	-163dBC

Overall System Allan Variance = 2×10^{-14} $1 < \tau < 100$ sec.

VI Overview of ESA Programs with Reference to their Time and Frequency Requirements

VI.1 Vulcan

The Vulcan mission which is current being studied, will address fundamental problems in solar, heliosphere, planetary and relativistic science. The multi-disciplinary Vulcan mission will examine the solar wind plasma, the electromagnetic fields and particle populations in the inner most regions of the heliosphere where the solar wind is accelerated. Vulcan will offer the possibility to carry out simultaneous in site measurements and remote sensing of the solar corona. It will study the structures in the solar atmosphere and corona with un-precedented spatial resolution. Furthermore, Vulcan will perform several experiments to test theories of gravitation in addition to measuring the quadruple moment J2 and possibly also the total angular momentum of the sun [3]. A similar program called STAR PROBE is under study by NASA. A joint ESA/NASA program may finally be agreed.

The interesting aspects of the Vulcan mission with regards time and frequency involves the gravitational experiments. In summary, the scientific objectives of the gravitation experiments are :

- verification of the metric hypothesis of general relativity
- independent determination of J2 as function of time
- measurement of the effect of gravity on photo propagation
- search for gravitational waves.

Accurate time keeping and range rate data are crucial. A frequency standard with a relative stability of at least 10^{-14} over 24 hours is foreseen, probably a hydrogen maser. The Doppler tracking accuracy must be of the order of $0.1mms^{-1}$ to ensure that the scientific objectives of the gravity experiments can be realised.

VI.2 Cassini

The Cassini mission is a joint program by NASA and ESA to explore the Saturn system with a Saturn orbiter and Titan probe. As reported in the last forum, [4] in order to perform Doppler wind experiments in the atmosphere of Titan, a USO with a stability of 5×10^{-10} over a 3 hour period, will allow relative wind velocity measurements over the same time interval accurate to $2ms^{-1}$.

Most probably a quartz USO will be selected but it must have extremely good phase noise performance. Minimum volume, mass, dc-power requirements and insensitivity to "g" force are still of prime importance.

VII ARISTOTELES

Geodynamics is concerned with physical forces and processes that affect the Earth's core, mantle and crust. It considers also the interactions of the solid Earth with the atmosphere and the oceans. Space techniques provide a powerful tool to study effectively the links between these quantities. Possibly the most essential pre-requisite for understanding the structure, dynamics and evolution of planet Earth are the precise determination of the Earth's potential fields :

- the Gravity field,
- the Magnetic field.

A solid Earth mission was included in the long term space plan of the Agency, adopted at the council meeting of ministers in Rome 1985 [5]. ARISTOTELES is a joint ESA-NASA program which involves a low Earth orbit satellite to determine the fine structure of the Earth gravity and magnetic fields on a global scale from a near sun-synchronous polar orbit [6]. A good description of the gravity field mission for planet Earth is given by Lambeck [7].

The time management will use a P-code GPS receiver developed by Jet Propulsion Laboratory (JPL). It is a miniaturised version of an existing geodetic receiver. Exploitation of the onboard receiver will afford a means of refining the long wavelength terms of the Earth's gravity field from measurements of the relative velocities between ARISTOTELES and the GPS satellites. GPS receivers will also support ARISTOTELES in the following areas :

- orbit determination,
- data time tagging,
- atmospheric research,
- high-low orbit satellite - satellite tracking,
- scientific data processing of both magnetometry and gradiometry.

When GPS is fully operational a minimum of 8 satellites will be simultaneously visible at any time.

VII.1 Columbus and Hermes

Columbus and Hermes are two of the larger ESA programs. Both are planned to be launched towards the end of this century.

Columbus is a manned laboratory designed for scientific experiments in fluid physics, life sciences, material sciences and microgravity. The system consists of 3 elements :

- attached laboratory,
- free flying laboratory called EURECA
- polar platform.

These elements will form part of the International Space Station, FREEDOM.

Hermes, the messenger of the Gods, is a delta wing re-usable European space plane designed to service Columbus.

Both programs will use the GPS system for primary navigation with an inertial system as back up. Typical foreseen requirements are :

	Position (m)	Velocity (ms^{-1})
Absolute	30	0.12
Relative	1	0.05 - 0.01

The relative measurements are important for rendez-vous maneuvers between the various elements. GPS in the differential mode and radio altimeter measurements, will be used as the main navigation sensor during the automatic glide landing phase of Hermes.

A GPS receiver pre-development program has been initiated with ALCATEL-SEL [8]. Two representative elegant breadboards are currently being evaluated in the ESA radio navigation laboratory [9]. Absolute and relative position and velocity performances are being determined using a 6 channel GPS satellite simulator.

The Columbus program offers an ideal opportunity for Europe to demonstrate the capability to develop and deploy atomic frequency standards in space.

VII.2 EURECA

In order to prepare for Columbus utilisation, ESA intends to carry out a precursor flight program in the 1994-1997 time frame. This flight program will consist of 2 spacelab flights and 2 EURECA flights. The EURECA flights may be conducted on a cooperative basis with NASA and thus gives scope for joint US-European proposals.

EURECA is an unmanned European retrievable carrier which will be launched by the shuttle orbiter. EURECA represents a bridge between space-lab and the space station. Its primary objectives will be microgravity research and flight demonstrations of new technology.

A joint ESA/NASA experiment to fly a hydrogen maser on EURECA is currently under discussion. The objective is to develop and test in space a hydrogen maser clock system designed for long term space applications. The original proposal [10] involved the evaluation of frequency stability and operation for a minimum period of 30 days in orbit. The maser frequency being compared to that of a reference maser used in the NASA Deep Space Network (DSN).

A second proposal under discussion [11] involves flying both a European and American maser. This has the advantage of cross-correlation of the phase data originating from two different high performance clocks and eliminates noise contributions of each clock. This will increase the precision of the measurement of gravitational effects.

The twin objectives of the experiment are:

- The evaluation of maser technology via frequency stability measurements.
- Evaluation of very high precision time and frequency transfer by microwave links and its comparison with laser time transfer. Propagation effects being eliminated by the use of 2 frequencies.

Applications include very long baseline interferometry in space, high precision relativity experiments, space-earth time and frequency transfer and in the longer term the detection of gravity waves.

VII.3 Data Relay Satellite System (DRSS)

The next Agency communications spacecraft will be the DRSS. An outline of the system and USO requirements were given in the 1990 forum [4]. Since that time the development of the engineering model has started. The development is only in its early phase. Preliminary design considerations and trade-off of different configurations such as electrical, mechanical and reliability aspects are addressed in the configuration status report [12].

VIII Conclusions

In many current and future ESA programs high stability Time-Frequency references are required. The oscillator parameter specifications vary with the application, with various emphasis on short, medium or long term stability, phase noise, warm up time, size etc... Different oscillator types are better matched to different applications e.g. quartz and Rubidium oscillators, Cesium beam clocks and hydrogen masers.

ESA is currently involved in study, design and development of these frequency references.

Instead of establishing a Time and Frequency laboratory at ESTEC the work has been placed in industry under the umbrella of an ESA CoE.

It is hoped that the CoE will benefit the whole European Time and Frequency community with the stimulation of new ideas and proposals for programs, experiments and technological advances.

The Columbus program offers an ideal opportunity for European industry to demonstrate its ability to design and manufacture atomic frequency standards for long duration space applications. Active participation of the Time and Frequency community is sought, which hopefully will result in the initiation of new experimental studies and programs.

(1) ESA/IPC paper (89)39 Paris 15 March 1989

(2) P. Thomann, G. Busca, Oscilloquartz Final Report ESTEC Contract 7720/88/NL/JG

(3) The Vulcan Mission ESA Publication SCI (88) Sept.88

(4) S.J. Feltham, Frequency Sources for Space Applications, 4th E.F.T.F. Neuchatel March 1990 p.429

(5) Space Science Horizon 2000 (ESA-SP-1070)

(6) ARISTOTELES Program Proposal, ESA Publication ESA/EOSTAG (91)1

(7) K. Lambeck, ARISTOTELES: an ESA mission to study the Earth's Gravity field, ESA Journal 90/1

(8) R. Lucas, ESA Contract number 8339/89/

(9) R. Lucas et al "Overview of ESA-ESTEC GPS Activities", ESTEC, Keplerlaan, 2200 AG Noordwijk, Holland

(10) R. F. C. Vessot, E. M. Mattison, Development and in-space evaluation of a High Stability Hydrogen maser clock, Proposal P2223-12-89.

(11) G. Busca, S. Starker, Observatory of Neuchatel, Notification of interest, Research opportunities in Space, 1-2-91

(12) V. Pointet, Ref. Oscillator for the DRS satellite, Oscillator configuration status Report N°SRO-RPT-011-CIR. CIR, Neuenburg strasse 7, CH-2076 Gals/BE, Switzerland.

Session 5B Number 37

SA AND CONTROL OF LOCAL OSCILLATOR IN GPS RECEIVERS

Kalevi Kalliomäki (1,2), Tapio Mansten (1)

(1) Technical Research Centre of Finland, Otakaari 7 B, SF-02150 Espoo
(2) University of Oulu, Dept. of Electrical Engineering, SF-90570 Oulu

Abstract

We have compared three GPS-receivers, Kinemetrics Truetime, Datum GPS model 9390 and Odetics Kode SatSync Model 325. The test period has been from 40 to 60 days in 1990 and 144 datapoints per day was collected. As an independent local reference we have two rubidium oscillators and three Loran-C receivers. The systematic time difference between those receivers varied up to 500 ns. The short term noise was in best case less than 50 ns rms. SA could cause microsecond level time jumps and exp-10 frequency errors.
As a conclusion time and frequency control software of GPS-receivers seems to work well at least 95% of time but sometimes, especially during SA the behaviour may be bad.

Introduction

The number of GPS-receivers has increased rapidly during the last few years and there are several manufacturers producing different designs. The receiver usually controls a high stability oscillator, which works as a flywheel smoothing timing jitter and offering continuous time scale. This oscillator is either an oven controlled crystal oscillator or a rubidium oscillator.
Only national time laboratories can take advantage of common view techniques to eliminate most of timing errors eg. due to SA (Selective Availability). Conventional users have to rely on time and frequency outputs of a local receiver. Those outputs are subject to many errors depending on current satellite constellation, possible SA, intelligence level of the control software and the stability of local oscillator.
We have compared three GPS-receivers, Kinemetrics Truetime, Datum GPS model 9390 and Odetics Kode SatSync Model 325. The test period was from 40 to 60 days in 1990 and 144 datapoints/day was collected.

Methods

As an independent local reference we have two rubidium oscillators and three Loran-C receivers (Sylt, Bryansk). To improve the performance of those instruments and our own GPS receiver, the temperature of the clock cabinet is carefully stabilized to +/-0.1 degrees C level. In addition, the 27 V DC supply is double stabilized. A computer controlled data logging system (Kalliomäki 1990) reads the phase differencies between various clocks and receivers every ten minutes.

The resolution of the counter is 1 ns and typical daily residual phase variation 10 nsrms between the rubidiums. GPS receivers under test were not placed into the same cabinet as our own clocks but the temperature variation in the air conditioned time laboratory is only +/- 0.5 degrees C maximum. The GPS antennas were on a flat roof about 15 meters from the receivers. To improve the accuracy of time, the coordinates were usually fixed during the tests and only block I satellites were selected if SA condition was to be expected.

Datum GPS model 9390, test 1

Datum receiver was tested at the beginning of the year 1990, MJD 47903...47943. During that time the satellites even in block II behaved quite nicely, only SV 14 was sometimes noisy. The coverage of satellites was good, thus no "gaps" were observed in phase records.
It was possible to adjust the receiver delay to compensate for delays in antenna cables etc. Unfortunately the delay setting was not direcly calibrated in nanoseconds. Therefore we followed the rules of thumb given in the manual. The systematic time difference between our own GPS receiver and Datum was suprisingly high, about 500 ns and there was a statistically significant trend of 1.2 ns/d (Fig 1).
The standard deviation on the time difference was 60 ns (2 h means) / 22 ns (24 h means). More careful analysis pointed out that Datum was more noisy than Kinemetrics.
During the test period the average residual standard deviation of raw data was 84 ns for Datum and 46 ns for Kinemetrics. There was a clear daily variation about 1:2 in the noise level. In addition we observed peculiar unstabilities in both receivers. On MJD 47905.0 Datum got grazy; its second tick varied up to 25 us for half a day (Fig 2), Kinemetrics was stable. Later on MJD 47941.4 it was Kinemetrics turn, its time jumped about 100 us for half an hour (Fig 3). In both cases receivers recovered itself.

Odetics Kode SatSync Model 325, test 2.

Odetics receiver was tested at the middle of the year 1990, MJD 48056...48125. During that time there was at least four SA periods, during which the satellites in block II got grazy for several days. The coverage of satellites was limited because we must use only block I satellites for most of the time. Thus there are daily "gaps" in phase records.
The systematic time difference between the receivers was 150 ns, Kinemetrics was late as in the previous test. The trend in difference was 0.1 ns/day only (Fig 4). In spite of worse satellite situation, the standard deviation of the difference was lower than earlier, 22 ns (2 h means) and 15 ns (24 h means). Now the short term noise levels of raw data of the individual receivers were the same, 32 ns.

During the test period, we had to restart Odetics twice because it suddenly started to behave as an ordinary crystal clock, drifting at a rate of about 2 us/day. First time we waited for one week for automatic recovery. In addition there were two 1000 us (=C/A code length) time jumps lasting 2...3 hours. Kinemetrics was stable in all those cases. We have seen earlier a 1000 us jump in Kinemetrics, too.
By allowing free selection of satellites in SA case simply doubled or tripled the timing jitter (Fig 5), Odetics seems to tolerate quite well the SA situation. In survey mode the effect was more dramatic; timing noise increased to 200 ns rms level.

Comments about Kinemetrics Truetime

We have had this receiver since 1989. It has an internal rubidium oscillator (Efratom). Most of the time the receiver behaves nicely but SA condition is difficult. As a rule the rubidium control software gets grazy and a restart, even a cold start is needed. In quiet situation the software works well as a rule but sometimes it allows the time to drift out of specifications.

Conclusions

The systematic time difference between the three receivers varied up to 500 ns. The short term noise was in best case less than 35 ns. SA could cause increased timing jitter or ruin the operation of local oscillator control software. As a conclusion time and frequency control software of GPS-receivers seem to work well at least 95% of the time but sometimes, especially during SA the behaviour may be bad.
We had to reset (cold/warm start) the receivers several times, because the control software got grazy. By selecting block 1 satellites only, the problems seem to disappear. In addition we observed that during good conditions, the time of GPS clocks may sometimes deviate exactly 1 ms (C/A code length) from nominal time for several hours. Therefore independent local clocks are still needed to observe time/frequency jumps. The control software, "disciplined control" clearly needs improvements to tolerate or remove bad satellites, too.

Reference

Kalevi Kalliomäki, Tapio Mansten and Jari Mannermaa; A computerized time and frequency standards monitoring system, EFTF 90, Neuchatel.

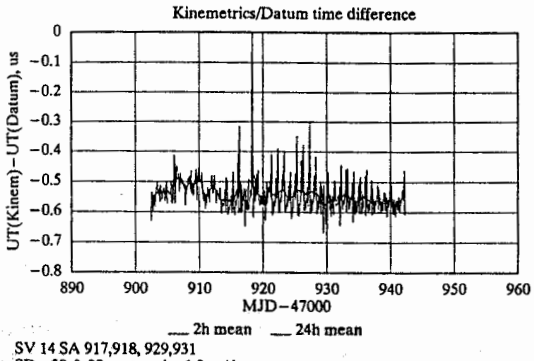


Fig. 1 Time difference (Datum)

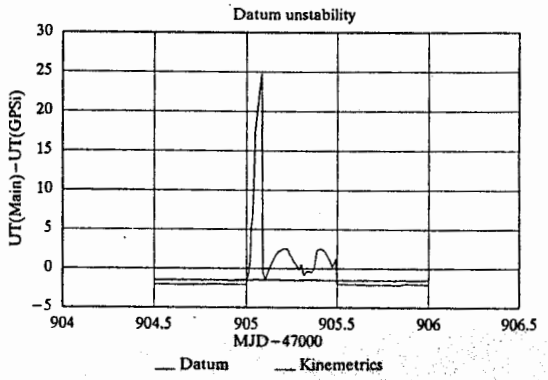


Fig. 2 Datum unstability

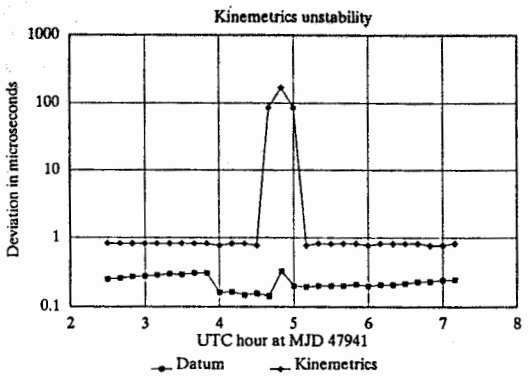


Fig. 3 Kinemetrics unstability

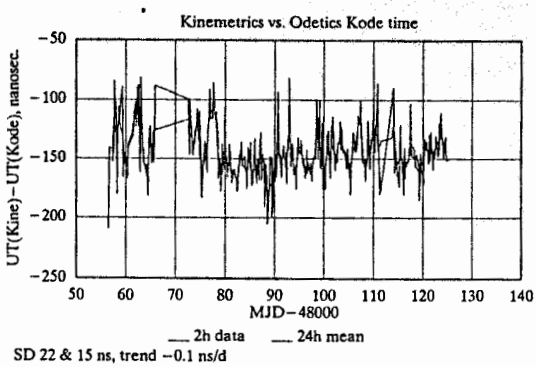


Fig. 4 Time difference (Ode.)

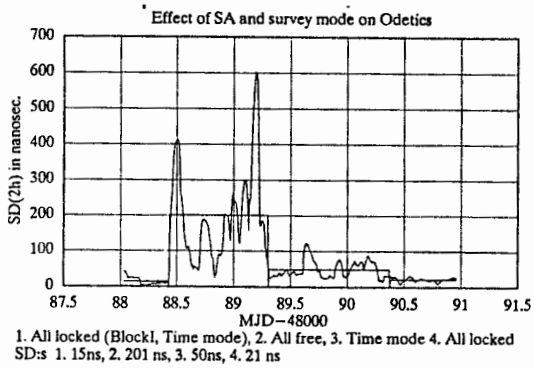


Fig 5 Effect of SA on Odetics

Session 5B Number 38

GPS TIME CLOSURE AROUND THE WORLD USING PRECISE EPHEMERIDES, IONOSPHERIC MEASUREMENTS AND ACCURATE ANTENNA COORDINATES.

W. Lewandowski, G. Petit, C. Thomas
Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex, France

M.A. Weiss
Time & Frequency Division
National Institute of Standards and Technology
325 Broadway, Boulder C0 80303, USA

ABSTRACT

Over intercontinental distances the accuracy of Global Positioning System (GPS) time transfers ranges from 10 to 20 ns. The principal error sources are the broadcast ionospheric model, the broadcast ephemerides and the local antenna coordinates.

Previous work has already shown the impact of the use of measured ionospheric delays as provided by dual frequency codeless ionospheric measurement systems. The role of precise GPS satellite ephemerides, now available from the US Defense Mapping Agency (DMA), has already been investigated. The improvement brought by the corrections to local GPS receiver antenna, resulting from the BIPM differential positioning method, has also been demonstrated for short-distance time comparisons.

At present, ionospheric measurement systems of the type designed by the National Institute of Standards and Technology (NIST) operate on a regular basis at the NIST in Boulder (Colorado, USA) and at the Paris Observatory (OP) in Paris (France), and systems of the type designed by the Communications Research Laboratory (CRL) operate at the CRL in Tokyo (Japan). Broadcast ephemerides are currently recorded in Mojave (California, USA) and at the BIPM (France). The GPS antenna coordinates of OP, NIST and CRL are now unified in IERS Terrestrial Frame. In this paper, we realize for the first time the closure around the world obtained by the combination of time transfers OP-NIST, NIST-CRL and CRL-OP after reduction of the three major error sources. It gives the evidence of improvement in accuracy for GPS time transfer with using precise ephemerides, measured ionospheric delays and accurate antenna coordinates.

INTRODUCTION

The excellence of worldwide time unification depends on the quality of the clocks kept by national timing centers and on the means of time comparisons. Rapid development of the Global Positioning System since 1983 has led to major improvements in the precision and accuracy of time metrology. Using commercially available C/A Code* GPS time receivers, time comparisons can easily be performed with an accuracy of 10 to 20 nanoseconds over intercontinental distances. However, it should be possible to improve this performance by removal of residual systematic errors [1]. In GPS time transfers, the three principal sources of error are the local antenna coordinates, the broadcast ionospheric model and the broadcast ephemerides. Previous work shows the effect of correcting these errors individually [2] [3] [4]. We present here a five-month experiment in which three long-distance time links are performed with simultaneous reduction of these error sources. The laboratories involved are the Paris Observatory (Paris, France), the National Institute of Standards and Technology (Boulder, Colorado, USA) and the Communications Research Laboratory (Tokyo, Japan). Closure around the world is obtained by combination of these time links: since the closure leads in principle to a zero sum, we have a clear test of accuracy for overall GPS time transfer.

* *Acronyms are listed at the end of the text.*

1. THE EXPERIMENT

The three long-distance time transfers UTC(OP) - UTC(NIST), UTC(NIST) - UTC(CRL) and UTC(CRL) - UTC(OP) were computed using the common-view method [5], for a 162-day period, from 1990 June 16 (MJD 48058) to 1990 November 24 (MJD 48219).

The GPS data taken at the three sites correspond to the International Schedule N°15, issued by the Bureau International des Poids et Mesures, and implemented on 1990 June 12 (MJD 48054). This schedule includes Block I and Block II satellites. For one part of the period under study (1990 June 16 to 1990 August 10), the intentional degradation of GPS signals, known as Selective Availability (SA), was turned on for Block II satellites. It can be shown however that it affected only the satellite clocks, producing a phase jitter which is completely removed by strict common views [6].

In our experiment only common views with the same starting time and the same track length are kept. Time comparison values UTC(Lab1) - UTC(Lab2) are obtained for each observed satellite at the time, T_{mid}, of the midpoint of the track.

At the time of the experiment some satellites were in their rephasing manoeuvre. In consequence, their elevations were sometimes too low to be observed from one of the sites although, in principle, they formed part of the international Schedule.

The GPS receivers used at OP and NIST come from the same maker and use the same software to treat the short-term data. This enhances the symmetry of the experiment for the time link OP-NIST. This is not the case for the NIST-CRL and CRL-OP links, the GPS receiver in regular operation at CRL coming from another maker. Detailed characteristics of the three time links are given in Table 1.

TABLE 1: *Time links OP-NIST, NIST-CRL and CRL-OP.*

	OP-NIST	NIST-CRL	CRL-OP
Baseline	7388km	8522km	8816km
CV/day	9	7	8
p	61%	62%	48%

* *CV/day gives the number of scheduled common-view tracks by day for each pair of laboratories*

* *p gives the fraction of scheduled tracks for the period of study (162 days) which correspond to perfect common views and which could simultaneously be corrected for antenna coordinates, measured ionospheric delay and precise ephemerides.*

Before analysing the results of this experiment, we review the methods for reducing the three main sources of errors: antenna coordinates, broadcast ionospheric model and broadcast ephemerides [7].

1.1 Accurate antenna coordinates

Accurate antenna coordinates can be determined with uncertainties of a few centimeters using geodetic methods to obtain [8] the relative position of the antenna with respect to the nearest IERS site. The BIPM has also developed a method of differential positioning between GPS antennas using the data of the time comparisons themselves [2]. The consistency of the coordinates obtained by this method is within 50 cm for distances up to 1000 km. By combining these techniques, all national laboratories equipped with GPS receivers have, over the last few years, been linked to IERS sites [9]. On 1990 June 12 at 0h00 UTC (MJD 48054), as suggested by the BIPM, these corrected coordinates were introduced into the GPS time receivers, ensuring worldwide homogeneity of the coordinates in the IERS Terrestrial Reference Frame (ITRF).

Thus, at the beginning of our experiment, the OP antenna coordinates were known with an uncertainty of 50 cm. They were obtained, by the BIPM differential positioning method, with respect to the Grasse ITRF SLR site from data covering the period 1987 December 15 to 1988 June 21 [2]. For the NIST, the GPS antenna has coordinates known with an uncertainty of 30 cm. These were obtained by GPS geodetic differential positioning with respect to the Platteville VLBI site on July 1989 [8]. The CRL antenna coordinates are known to within 10 cm, relative to the Kashima VLBI site, by local survey on March 1989 [10].

1.2 Measured Ionospheric Delay

In the usual GPS data files, the correction used for ionospheric refraction comes from a model [11], the parameters of which are included in the GPS message. At radio frequencies, however, the ionosphere is a dispersive medium so its effect on time comparison between local and GPS satellite clocks can be estimated by dual-frequency methods. Dual-frequency receivers, which do not depend on knowledge of the P-Code, have recently been developed [12] [13]. They give measurements of ionospheric delay along the line of sight of satellites with uncertainties of 1 to 2 ns [14]. The gain in precision for long-distance time comparisons has already been demonstrated when the two branches of the link are corrected with measured ionospheric values [15]. The gain in accuracy was also confirmed by the study of the closure around the world via NIST, OP and CRL through the use of such measurements [3].

The OP and NIST are equipped with similar dual-frequency GPS receivers of the NIST type [13] (NIST Ionospheric Measurement System). In their present configuration, these devices give values of ionospheric delay for all satellites in view every 15 seconds. These data are stored after a linear fit over 15 minutes at a time corresponding to round quarters of hours: 0h00 UTC, 0h15 UTC, 0h30 UTC, etc.

The CRL is equipped with another type of dual-frequency GPS receiver designed by the CRL [12] (Realtime TECmeter). This device has a single channel and is programmed to observe by priority the satellites suggested by the BIPM International Schedule. It operates with 4-minute sequences: about 1 minute to point its directional antenna and about 3 minutes to perform the observation. Values of the ionospheric delay are taken every 6 seconds for a given scheduled satellite. The mean and the standard deviation of these data are provided for the mid-point of the observation sequence.

From the ionospheric data provided by one or another device are made estimates of the value of the measured ionospheric delay for the mid-point (T_{mid}) of a given 13-minute track of satellite s. To do this, we use several measurements for the same satellite s, surrounding T_{mid}. A polynomial fit is then performed (linear to cubic depending on the number of values used). The estimated value at T_{mid} is deduced by interpolation and is used to correct UTC(Lab) - GPS time. This polynomial fit is never extrapolated and measurements from other satellites are never used.

1.3 Precise ephemerides

The GPS precise ephemerides were computed at the U.S. Naval Surface Warfare Center (NSWC) from the beginning of 1986 to 1989 July 29. Since then they have been produced by the Defense Mapping Agency (DMA). These ephemerides are received on a regular basis at the BIPM. Their estimated accuracy is of order 3 meters. At present, the delay of access to precise ephemerides is about two months, so the period we study here is limited to mid-November 1990.

In practice, computations with precise ephemerides require knowledge of the broadcast ephemerides used, during the observation, by the receiver software in order to apply differential corrections [4]. To perform the closure around the world it is necessary to have access to recorded broadcast ephemerides on at least two sites correctly situated in the world. The BIPM started the regular collection of GPS broadcast ephemerides in May 1990. For this experiment, these data were used together with broadcast ephemerides recorded in Mojave (California, USA) by the U.S. National Geodetic Survey (NGS).

The precise ephemerides, PE_i, are provided in cartesian coordinates (expressed in WGS 84 reference frame) at time T_i corresponding to round quarters of hours: 0h00 UTC, 0h15 UTC, 0h30 UTC etc. It is then necessary to compute, from the broadcast Keplerian elements of satellite s, its positions BE_1, BE_2 and BE_3 for three times T_1, T_2 and T_3, such that:

$$T_1 < T_{start} < T_2 < T_{stop} < T_3$$

where T_{start} and T_{stop} are the starting time and the stopping time of the usual 13-minute tracking. The ephemeride corrections $PE_i - BE_i$, for i = 1,2,3, are transformed in a frame linked to the satellite (On-track, Radial, Cross-track) and a quadratic polynomial in time is computed to represent each component. A quadratic representation is also computed in the same frame for the vector satellite-station. The inner product of these quadratic representations provides the corrections to GPS measurements each 15 seconds. A linear fit over 13 minutes on these short-term corrections gives the correction at the middle-time T_{mid} of the track.

1.4 Combining the data

Data files of measured values UTC(Lab) - GPS time and of ionospheric measurements for the three laboratories involved, as well as data files of precise ephemerides produced by the DMA and broadcast ephemerides from Mojave and Paris, were gathered at the BIPM for processing. Only the tracks corresponding to perfect common views, which could be simultaneously corrected for ionospheric model and precise ephemerides (see last line of Table 1), were retained. The results given in Sect. 2 involve only the values UTC(Lab1) - UTC(Lab2) corresponding to these particular trackings, whether the tracks include the corrections or not.

2. Results

The gain in precision from the simultaneous use of corrected antenna coordinates, ionospheric measurements and precise ephemerides has already been demonstrated for an observation period of 67 days for the time link UTC(OP) - UTC(NIST) [7]. We present here similar results for a longer period and for three time links. We then analyse the results concerning the closure around the world.

2.1. Precision of time comparisons

The corrections to the antenna coordinates being already introduced, four different cases may be distinguished for each pair of laboratories:

* non-corrected values
* values corrected for ephemerides only
* values corrected for ionosphere only
* values corrected for both ephemerides and ionosphere.

For each case, a Vondrak smoothing [16] is performed on the values UTC(Lab1) - UTC(Lab2). The standard deviations of the residuals to the smoothings for the complete period under study are given in Table 2. The smoothing used acts as a low-pass filter with a cut-off period of about 3 days.

TABLE 2: *Standard deviation, in nanoseconds, of the residuals to the smoothed values UTC(OP) - UTC(NIST), UTC(NIST) - UTC(CRL) and UTC(CRL) - UTC(OP) for the whole period under study, with application or not of the corrections.*

	OP-NIST	NIST-CRL	CRL-OP
non corrected	9,7	15,4	10,4
ephem. only	8,9	8,9	8,1
iono. only	5,9	12,3	6,7
ephem. + iono.	4,3	4,5	4,6

Application of individual corrections decreases the total standard deviation, for each link, with a clear improvement when both corrections are applied simultaneously. This improvement is linked to the length of the baselines involved. For such long baselines, common-view observations are mostly at low elevations and so are more sensitive to ionospheric effects and to satellite positions.

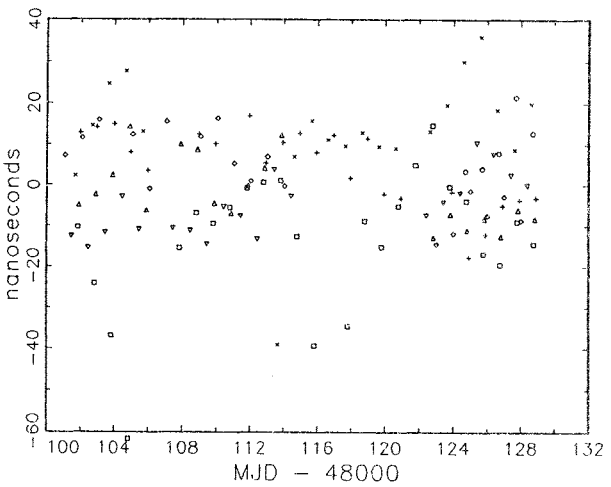


FIGURE 1: *GPS time transfer UTC(NIST) - UTC(CRL): residuals to the smoothed non-corrected values (a different symbol is used for each track).*

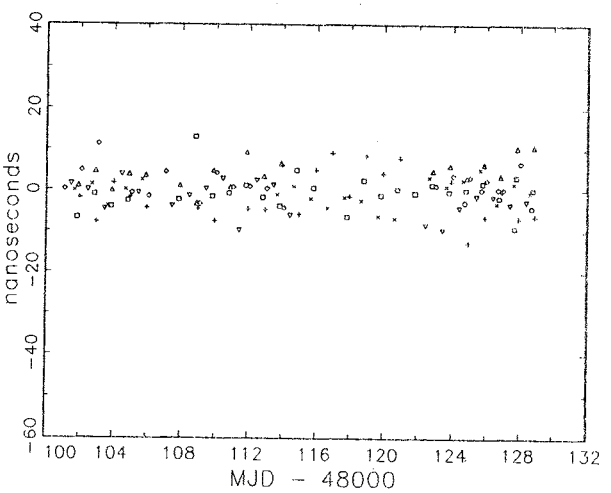


FIGURE 2: *GPS time transfer UTC(NIST) - UTC(CRL): residuals to the smoothed values previously corrected for precise ephemerides and measured ionospheric delays.*

The residuals to the smoothed values UTC(NIST) - UTC(CRL) for each data point of a given period of the experiment are presented in Figs. 1 and 2. Comparison of these figures shows the improvement of precision in the determination of UTC(NIST) - UTC(CRL) when the error sources are corrected simultaneously. The daily standard deviations of the residuals drop to values around 4 ns (see Fig. 3).

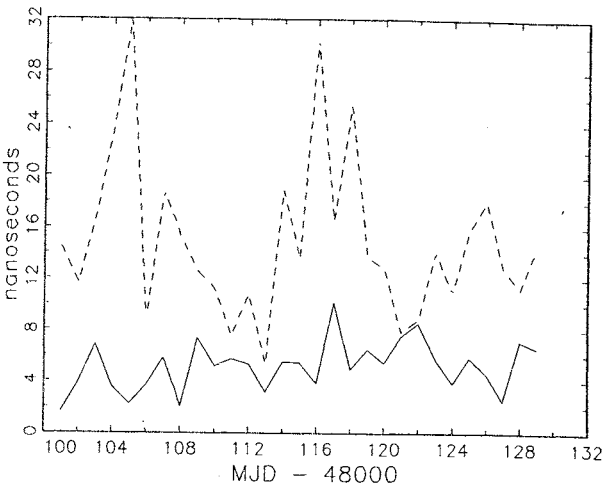


FIGURE 3: *GPS time transfer UTC(NIST) - UTC(CRL): daily standard deviations of the residuals obtained with:*
— — — — — non-corrected data,
———— data corrected for precise ephemerides and measured ionospheric delays.

Another example of the improvement in precision is given in Fig. 4 and concerns the time link OP-NIST in July-August 1990. The residuals to smoothed values of UTC(OP) - UTC(NIST) are shown for only one daily common view corresponding to satellite 12. With non-corrected data, satellite 12 presented very large residuals for some days in August. This effect completely disappears only when corrections for precise ephemerides are applied and was thus linked to very poor broadcast ephemerides. In general, the use of precise ephemerides helps to smooth the daily residuals for each common-view track while the use of ionospheric measurements tends to decrease the biases between satellites [7]. These two effects combine to reduce the uncertainty of the time comparison.

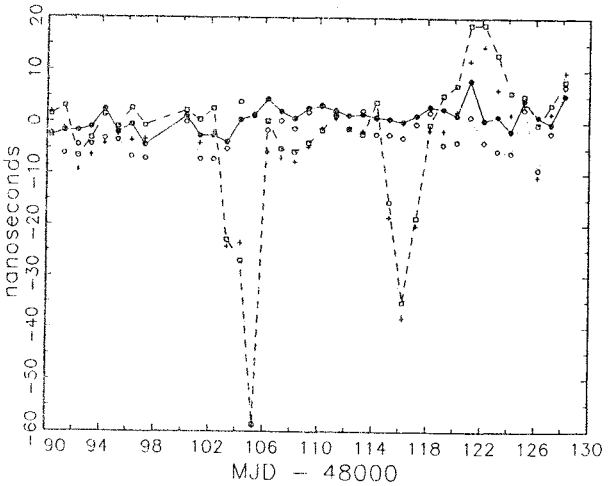


FIGURE 4: *GPS time transfer UTC(OP) - UTC(NIST): residuals to the smoothed values for one daily common-view track (satellite 12).*
—□—□—□—□—□— non-corrected data,
·····o·o·o·o·o·o·o·o· data corrected for precise ephemerides only,
+···+···+···+···+··· data corrected for ionosphere only,
—•—•—•—•—• data corrected for precise ephemerides and ionosphere.

Daily values of UTC(CRL) - UTC(OP), at 0h00 UTC, were estimated from the smoothed data points. The Allan deviation of these daily non-corrected and corrected values is given in Figs. 5 and 6 with a basic sample duration equal to one day. It is impossible to estimate the Allan deviation on a shorter evenly-spaced time interval since the scheduled common views CRL-OP are spread over 11 hours each day.

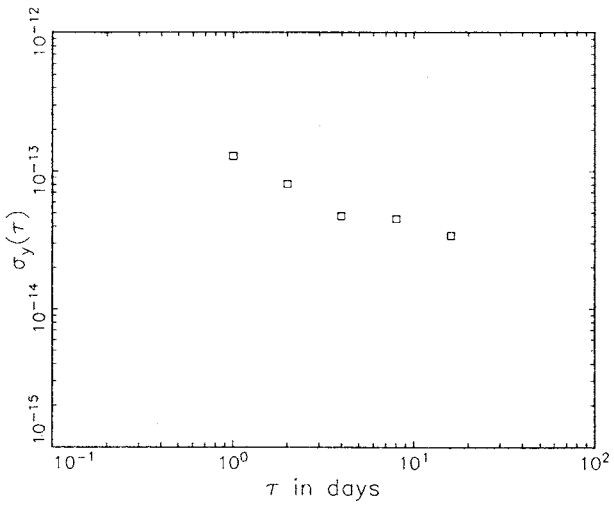


FIGURE 5: *GPS time transfer UTC(CRL) - UTC(OP): Allan deviation of the non-corrected values.*

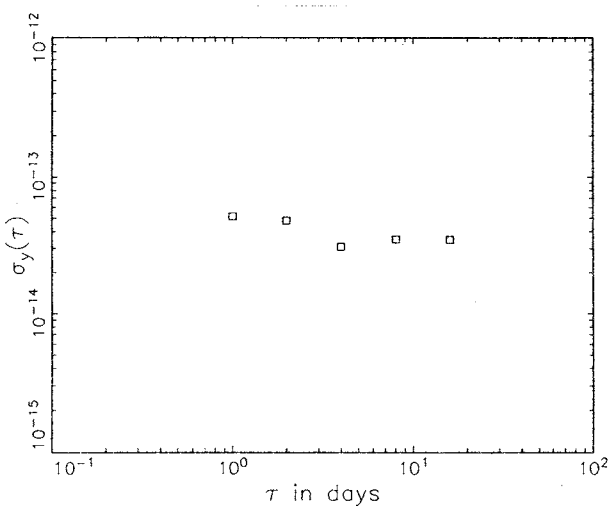


FIGURE 6: *GPS time transfer UTC(CRL) - UTC(OP): Allan deviation of the corrected values for precise ephemerides and ionosphere.*

Raw data are affected by white phase noise whose origin is the time difference measurements. It is smoothed out by averaging over 5 to 6 days. When corrections for precise ephemerides and measured ionospheric delays are applied, the measurement noise is already smoothed out when averaging over one day. The real performance of the local clocks, white frequency modulation for that averaging time, is then accessible to examination.

2.2. ACCURACY TEST: CLOSURE AROUND THE WORLD

A test of accuracy for GPS time transfer can be performed by computing the closure around the world via OP, NIST and CRL. Daily values of UTC(OP) - UTC(NIST), UTC(NIST) - UTC(CRL) and UTC(CRL) - UTC(OP) were estimated from the smoothed data points (the Vondrak smoothing was performed with the same smoothing degree for each time link). The resulting daily values of the deviation from closure, for the whole period under study, are shown in Fig. 7 without correction and in Fig. 8 with corrections applied. Fig. 8 provides evidence of a gain in accuracy when all three long-distance time links are computed using the corrections for precise satellite ephemerides and measured ionospheric delays.

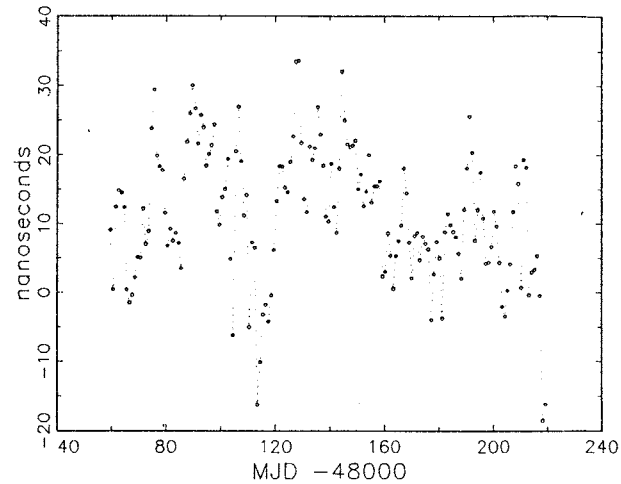


FIGURE 7: *Deviation from closure around the world via OP, NIST and CRL with non-corrected GPS data.*

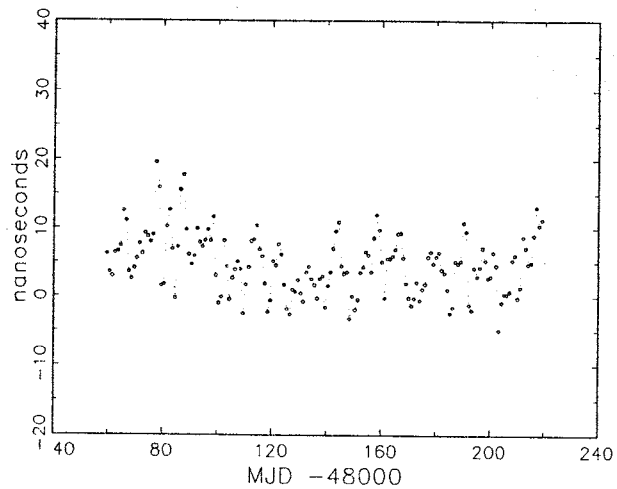


FIGURE 8: *Deviation from closure around the world via OP, NIST and CRL with GPS data corrected for precise ephemerides and measured ionospheric delay.*

To analyze the closure results in detail, Allan deviations were computed with and without corrections and plotted in Figs. 9 and 10. The figures exhibit common behaviour for the Allan deviation as a function of the inverse of the averaging time. The noise is lower when corrections are applied. It should be noted that the bending for $\tau < 2$ days is due to the smoothing that was applied on individual time links and has no physical interpretation.

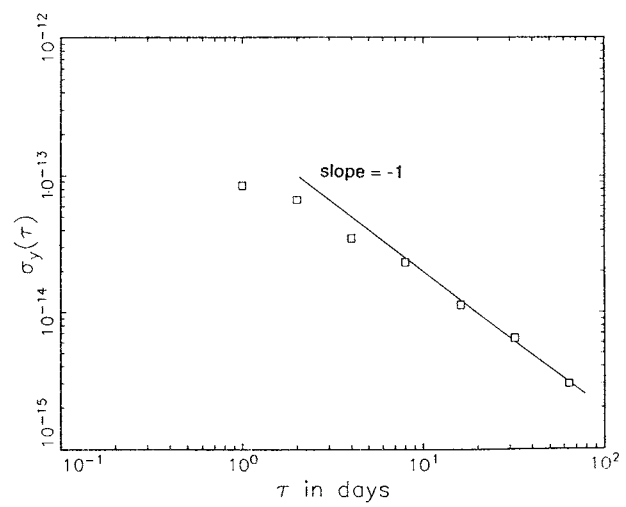


FIGURE 9: *Square root of the Allan variance of the deviations from closure obtained with non-corrected data.*

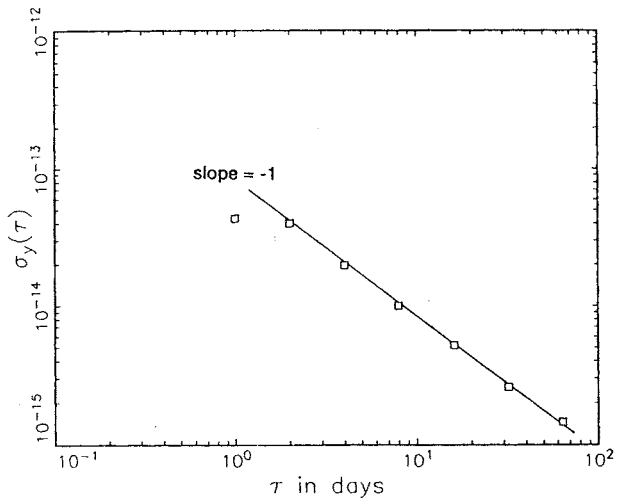


FIGURE 10: *Square root of the Allan variance of the deviations from closure obtained with data corrected for precise ephemerides and measured ionospheric delay.*

Use of the modified Allan variance (Figs. 11 and 12) allows us to characterize the noise affecting the values of the deviation from closure. Non-corrected data exhibit flicker phase noise. When corrections are applied, the flicker phase noise is not present and the obtained values, for the closure, then exhibit white phase noise up to an averaging time of about 20 days. This justifies computation of mean values of the deviation from closure for periods of duration up to 20 days and corresponding standard deviations.

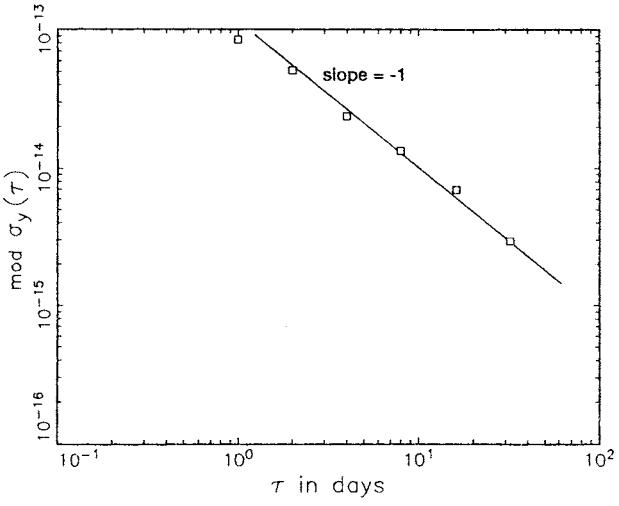


FIGURE 11: *Square root of the modified Allan variance of the deviations from closure obtained with non-corrected data.*

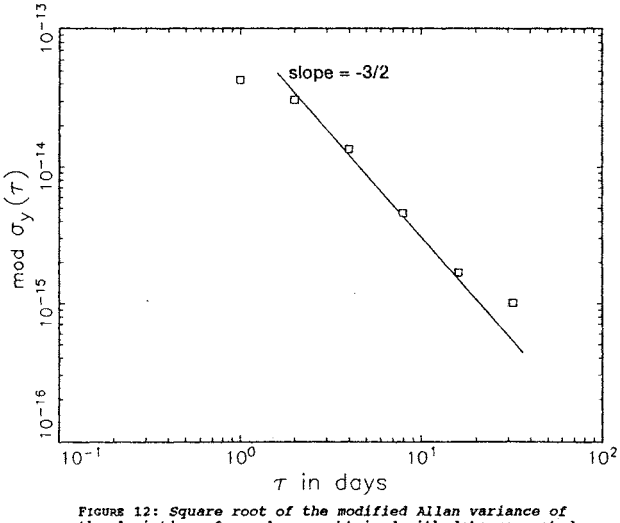


FIGURE 12: *Square root of the modified Allan variance of the deviations from closure obtained with data corrected for precise ephemerides and measured ionospheric delay.*

For 20-day averaging period the results are as follows (Table 3):

Table 3: *Mean values of the deviation from closure around the world estimated by the quantity [UTC(OP) - UTC(NIST)] + [UTC(NIST) - UTC(CRL)] + [UTC(CRL) - UTC(OP)] and standard deviations.*

Averaging period (MJD)	Mean deviation from the closure (ns)	Standard deviation (ns)
48059-48078	7,9	1,0
48079-48098	8,1	1,0
48099-48118	3,6	0,8
48119-48138	2,1	0,6
48139-48158	4,1	1,0
48159-48178	4,0	0,8
48179-48198	3,9	0,8
48199-48218	4,1	1,0

Table 3 shows a residual systematic effect of a few nanoseconds. Such a bias may be attributed to residual errors in antenna coordinates. The geographical positions of the involved sites necessitate that the useful common views are observed from one site for privileged directions. This keeps the coordinate errors from averaging and could produce a systematic bias.

Another point of interest is that the mean value of the closure depends on the period over which it is estimated. This may arise from the use of precise ephemerides which are not computed in a reference frame consistent with the IERS Terrestrial Reference Frame. This would introduce rotations of the reference frame of the precise ephemerides which would induce slowly varying shifts in the closure values. It might be possible to correct the precise ephemerides for this effect; further investigation of this topic is in progress at the BIPM.

Finally, one other phenomenon that could contribute to the bias in the closure is a systematic error due to the ionospheric measurements [14]. This probably depends on ionospheric conditions, in particular on whether observations are taken by day or by night. This would affect values of the deviation from closure obtained through time links having baselines long enough that one of the lines of sight corresponds to day-time while the other one corresponds to night-time. Such a bias will slowly evolve as the dates of observation change due to the sidereal orbits of satellites.

CONCLUSIONS

A five-month study of the GPS time transfers OP-NIST, NIST-CRL and CRL-OP shows that the consistency of long-distance GPS time comparisons is greatly improved through use of accurate antenna coordinates, precise satellite ephemerides and measured ionospheric delays. At the precision achieved with reducing these error sources, local time scales can be fully compared for averaging times of the order of one day. This performance matches with that observed for short-distance time comparisons [1].

The computation of the closure around the world via OP, NIST and CRL shows that corrections for precise satellite ephemerides and measured ionospheric delays greatly improve the accuracy of GPS time transfers. A slowly varying bias, of the order of a few nanoseconds, remains and is the subject of further investigations.

The GPS has proved to be an efficient tool for long-distance time comparison at the nanosecond-level both in precision and accuracy. At present, GPS is better than any existing system or method; yet this performance is still open to improvement with better conditions of operation: improved precise ephemerides, improved precise antenna coordinates, improved measurements of ionosphere, calibration of GPS receivers, unification of receiver software, control of multipath propagation, use of Earth tides model *etc.* The implementation of SA would be a severe drawback in terms of direct access to time, but its effects could be partially or completely removed for accurate time comparisons.

ACKNOWLEDGEMENTS

The authors are grateful to Pierre Uhrich of the Laboratoire Primaire du Temps et des Fréquences (Paris, France) for his kindness in collecting ionospheric measurement from the NIST Ionospheric Measurement System at the LPTF and sending them to the BIPM.

They also thank William Strange, of the U. S. National Geodetic Survey (Rockville, Maryland, USA), for providing access to NGS archives, and Miranda Chin for her kindness in assembling weekly files of broadcast GPS ephemerides recorded in Mojave and in sending them to the BIPM.

ACRONYMS

BIPM	Bureau International des Poids et Mesures
C/A Code	Coarse/Acquisition Code
CRL	Communications Research Laboratory
CV	Common View
DMA	Defense Mapping Agency
IERS	International Earth Rotation Service
ITRF	IERS Terrestrial Reference Frame
LPTF	Laboratoire Primaire du Temps et des Fréquences
MJD	Modified Julian Date
NIMS	NIST Ionospheric Measurement System
NIST	National Institute of Standards and Technology
NGS	National Geodetic Survey
NSWC	Naval Surface Warfare Center
OP	Observatoire de Paris
P-Code	Precision Code
SLR	Satellite Laser Ranging
VLBI	Very Long Baseline Interferometry
WGS	World Geodetic System

REFERENCES

[1] W. Lewandowski and C. Thomas, 'GPS time transfer', *IEEE Special Issue on Time*, 1991, in press.

[2] B. Guinot and W. Lewandowski, 'Improvement of the GPS time comparisons by simultaneous relative positioning of the receiver antennas', *Bulletin Géodésique*, 63, pp. 371-386, 1989.

[3] M. A. Weiss, T. Weissert, C. Thomas, M. Imae and K. Davies, 'The use of ionospheric data in GPS time transfer', *Proc. 4th European Frequency and Time Forum*, pp. 327-333, 1990.

[4] W. Lewandowski and M. A. Weiss, 'The use of precise ephemerides for GPS time transfer', *Proc. 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 95-106, 1989.

[5] D. W. Allan and M. A. Weiss, 'Accurate time and frequency transfer during common-view of a GPS satellite', *Proc. 34th Annual Symposium on Frequency Control*, pp. 334-346, 1980.

[6] D. W. Allan, M. Granveaud, W. J. Klepczynski and W. Lewandowski, 'GPS time transfer with implementation of Selective Availability', *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting*, 1990, in press.

[7] W. Lewandowski, G. Petit, C. Thomas and M. A. Weiss, 'The use of precise ephemerides, ionospheric data and corrected antenna coordinates in a long-distance GPS time transfer', *Proc. 22nd Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting*, 1990, in press.

[8] W. Lewandowski, R. J. Douglas, W. J. Klepczynski, W. Strange, J. Suter and M. A. Weiss, 'Positioning of GPS antennas in time-keeping laboratories of North America', *Proc. 43rd Annual Symposium on Frequency Control*, pp. 218-224, 1989.

[9] IERS Annual Report for 1988, Paris, 1989.

[10] CRL private communication to the BIPM.

[11] J. A. Klobuchar, 'Ionospheric correction for the single frequency users of the Global Positioning System', *IEEE Trans. NTS*, 1982.

[12] M. Imae, W. Lewandowski, C. Thomas and C. Miki, 'A dual-frequency GPS receiver measuring ionospheric effects without code demodulation and its application to time comparisons', *Proc. 20th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 77-86, 1988.

[13] D. Davis, M. A. Weiss and M. Vidmar, 'A codeless ionospheric calibrator for time transfer applications', *Proc. of the Institute of Navigation Satellite Division*, 2nd International Technical Meeting (ION GPS-89), 1989.

[14] M. A. Weiss and C. Thomas, 'A study of data from two GPS ionospheric calibrators', *Annual Report of the BIPM Time Section*, 2, pp.D1-D22, 1989.

[15] M. Imae, M. Miranian, W. Lewandowski and C. Thomas, 'A dual-frequency codeless GPS receiver measuring ionospheric effects and its application to time comparison between Europe and USA', *Proc. 3rd European Frequency and Time Forum*, pp. 89-93,1989.

[16] J. Vondrak, 'A contribution to the problem of smoothing observational data', *Bull. Astron. Inst. Czechoslovakia*, 20, pp. 349-355, 1969.

POSTER SESSION

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 39

DETERMINATION OF A COMPLETE FULLY CONSISTENT DATA SET OF THE FUNDAMENTAL PROPERTIES OF HIGH PURITY SYNTHETIC QUARTZ

Dr B J James

GEC-Marconi Limited, Hirst Research Centre
East Lane, Wembley, Middlesex, HA9 7PP, UK

In order to design electronic devices utilising quartz, a reliable set of the material constants and their temperature coefficients is required. The previously available data sets suffer from inconsistencies and inaccuracies in experimental and/or theoretical approach. The objective of this joint MoD/GEC-Marconi research programme currently underway at the Hirst Research Centre is to provide a data set of the fundamental properties of modern high purity synthetic quartz.

The purpose of this project is to measure the physical properties of quartz to high resolution and high absolute accuracy, referenced to National Standards or fundamental constants. Effort has been made to achieve universal consistency and to conform to the conventions of the current IEEE Standard on Piezoelectricity.

Important factors influencing the accuracy of the data obtained include:

Indium cold weld bonding of transducers allowing a very repeatable, well characterised bond to be produced and modelled using a six-layer transmission line model.

Use of a comprehensive "angular spectrum of plane waves" diffraction analysis using a parabolic approximation to the slowness surface, for all modes.

Interferometric sample length measurement giving an accuracy of ±5ppm.

A measurement of the density of quartz to an accuracy of ±50ppm.

Reference of all measured quantities to fundamental constants or National Standards of frequency, length, mass, temperature and resistance.

A measurement of the dielectric moduli, to ±0.2%, and their temperature coefficients over the range -60°C to +120°C was performed and the result incorporated in the determination of the elastic moduli.

Those parameters already measured include the second order elastic moduli and their temperature coefficients, the high frequency dielectric permittivities and their temperature coefficients, and the density. Measurements are currently being performed to ascertain the piezoelectric moduli and their temperature coefficients, the expansion coefficients, and the third order elastic moduli and their temperature coefficients.

QUARTZ FUNDAMENTAL PROPERTY MEASUREMENT PROGRAMME

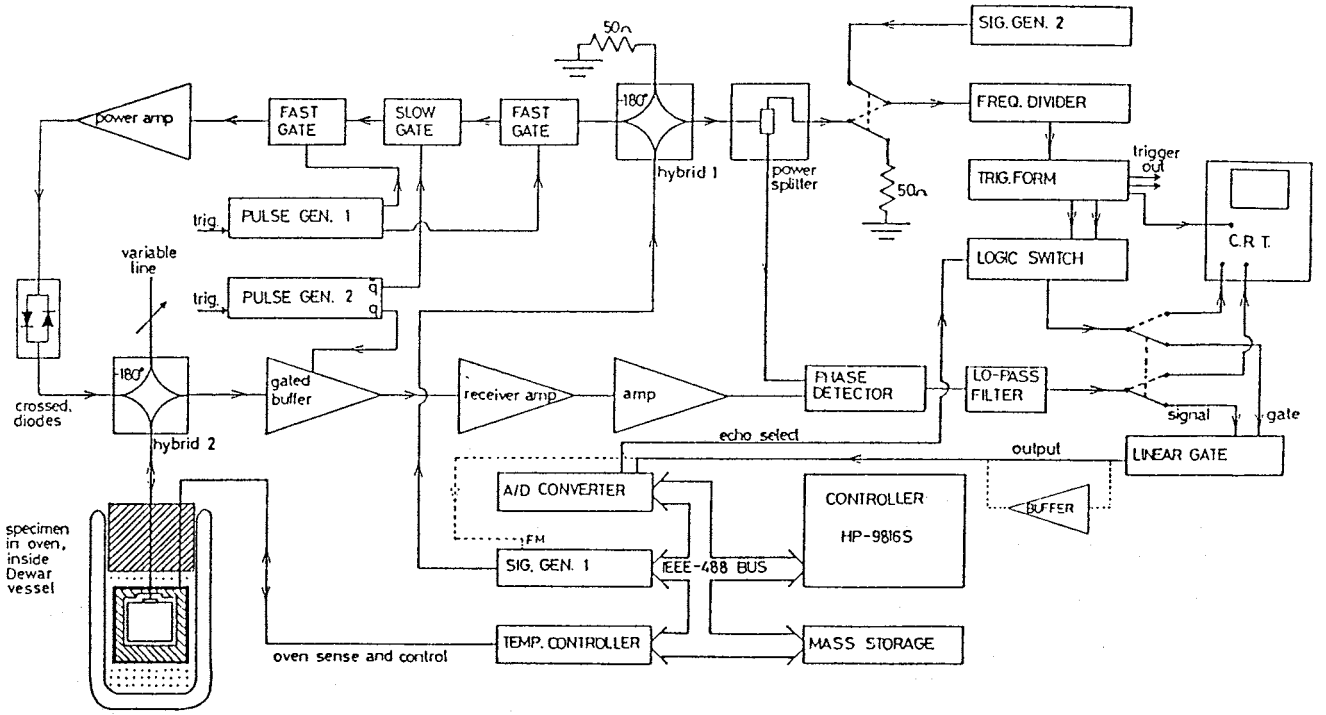
Material Properties Needed for Device Design/Simulation

Properties already measured

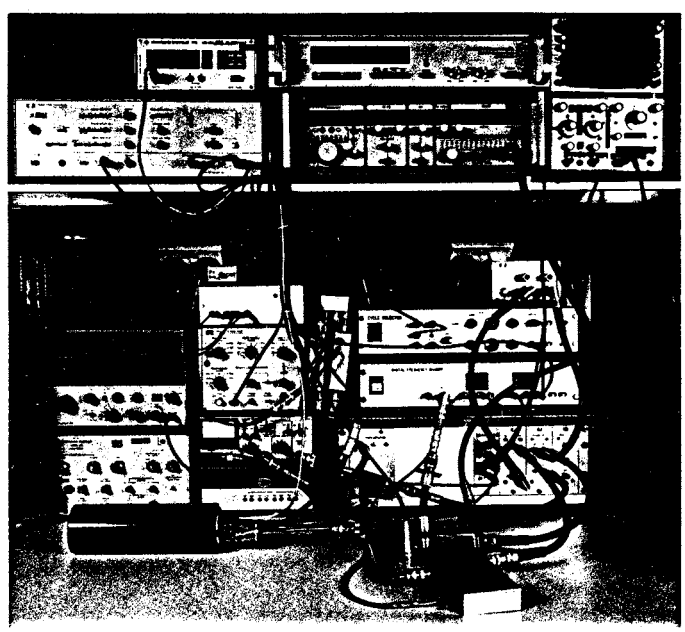
	Property	Method
1	Second order elastic moduli	Acoustic delay line echo phase comparison
2	Second order elastic moduli temperature coefficients	Phase locked echo phase comparison
3	High frequency dielectric permittivity	Precision capacitance determination
4	High frequency dielectric permittivity temperature coefficients	Capacitance change with respect to temperature
5	Density	Precision immersion technique

Properties currently under investigation

	Property	Method
1	Piezoelectric moduli	Precision resonator parameter measurement
2	Piezoelectric moduli temperature coefficients	Resonator parameter change with respect to temperature
3	Thermal expansion coefficients	Polarising interferometric dilatometer
4	Third order elastic moduli	Hydrostatic and uniaxial pressure derivatives of ultrasonic velocity using pulse-echo-overlap technique
5	Third order elastic moduli temperature coefficients	Velocity pressure derivatives with respect to temperature

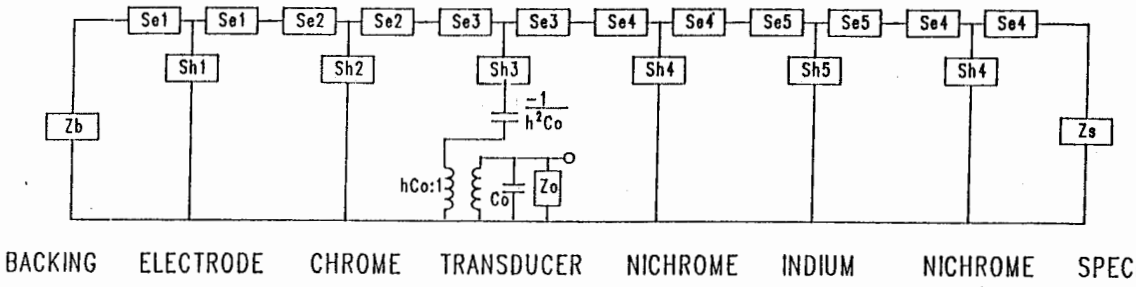


Schematic diagram of pulse-echo phase detection system



Automated pulse-echo phase detection system to operate over the frequency range 10-220 MHz. Temperature is controlled to 0.05°C.

$$Se_i = jZ_i \tan\theta_i \qquad Sh_i = \frac{-jZ_i}{\sin 2\theta_i} \qquad \theta_i = \frac{\pi t_i}{\lambda_i}$$



An important aspect of the measurement of the elastic moduli is the modelling of the indium cold-weld bonded transducers using transmission line techniques.

Data Set for Quartz

Parameter (f)	Units	f (25° C)	Δf (uncertainty)	$T_f^{(1)}$	$T_f^{(2)}$	$T_f^{(3)}$
Elastic Moduli	Nm^{-2}					
C_{11}		8.6790×10^{10}	1.3×10^{7}	-4.3591×10^{-5}	-1.1098×10^{-7}	-9.9879×10^{-11}
C_{12}		6.7901×10^{9}	3.7×10^{7}	-2.6401×10^{-3}	-3.2587×10^{-6}	-2.4878×10^{-9}
C_{13}		1.2009×10^{10}	1.8×10^{8}	-5.9227×10^{-4}	-1.2185×10^{-6}	6.8606×10^{-10}
C_{14}		1.8116×10^{10}	6.3×10^{7}	1.0259×10^{-4}	-3.0802×10^{-8}	-4.5928×10^{-10}
C_{33}		1.0579×10^{11}	2.1×10^{7}	-1.9032×10^{-4}	-1.6531×10^{-7}	-1.9089×10^{-11}
C_{44}		5.8212×10^{10}	1.8×10^{7}	-1.7159×10^{-4}	-2.5449×10^{-7}	-2.4495×10^{-10}
C_{66}		4.0000×10^{10}	1.7×10^{7}	1.7679×10^{-4}	1.5618×10^{-7}	1.0280×10^{-10}
Dielectric Moduli	Fm^{-1}					
$\varepsilon_{11}^{T}\varepsilon_o$		3.992×10^{-11}	1×10^{-13}	7.06×10^{-6}	1.632×10^{-8}	-2.206×10^{-10}
$\varepsilon_{33}^{T}\varepsilon_o$		4.104×10^{-11}	1×10^{-13}	1.88×10^{-5}	3.29×10^{-8}	-1.92×10^{-10}
$\varepsilon_{11}^{S}\varepsilon_o$		3.916×10^{-11}	——	1.47×10^{-5}	1.45×10^{-8}	-2.10×10^{-11}
Piezoelectric Moduli	Cm^{-2}					
e_{11}		0.1711	9.4×10^{-4}	-1.6×10^{-4}	0	0
e_{14}		0.0406	——	-1.44×10^{-3}	0	0
Expansion Coeff's						
α_{11}		l_x (25° C)		1.377×10^{-5}	1.303×10^{-8}	-6.329×10^{-12}
α_{33}		l_z (25° C)		7.483×10^{-6}	9.405×10^{-9}	-5.440×10^{-12}

$$\varepsilon_o = 8.85416\times10^{-12}\ Fm^{-1}$$

Density ρ (25° C) = $2648.38\pm0.03\ kgm^{-3}$

$$T_f^{(n)} = \frac{1}{n!}\frac{df^n}{f(25)\ dT^n} \qquad f(T) = f(25)\left[1+\sum_n T_f^{(n)}(T-25)^n\right]$$

Session POSTER Number 40

An alternative and economic conception for digital compensated crystal oscillators

Wolfgang Balzer

KVG GmbH, R&D crystal components, D-6924 Neckarbischofsheim

Abstract

Alternative in this conception are following items:
- The memory of the microprozessor from the equipment (radio) is used instead an internal microprocessor.
- Reduced number of measurement points by using interpolation formulas to calculate the oscillators temtemperature characteristics.
- Data transfer of the coefficients for the control voltage Uz and temperature sensor voltage Ut between user and producer via data disc.

Summary

Digital compensations for crystal oscillators have been known for a long time, but for most users, the advantages of this compensation method have not been substantial enough to outweigh the disadvantages of higher costs and larger sizes in comparison with temperature compensated crystal oscillators (TCXO) with thermistor-resistor networks. Especially in the past it was not possible to fulfill the requirements for mobile radios.
This paper describes an economic conception to produce in connection with the user, a crystal oscillator with following features:
- Stability +- 1,5 ppm (within the temperature range -25°C to + 75°C).
- Small enclosures (DIL-14).
- Low power consumption <1,5 mA at +5V supply voltage.
- Low costs.

Introduction

AT-cut crystals are used in most TCXO applications. The frequency-temperature characteristics of these crystals are shown in figure 1.

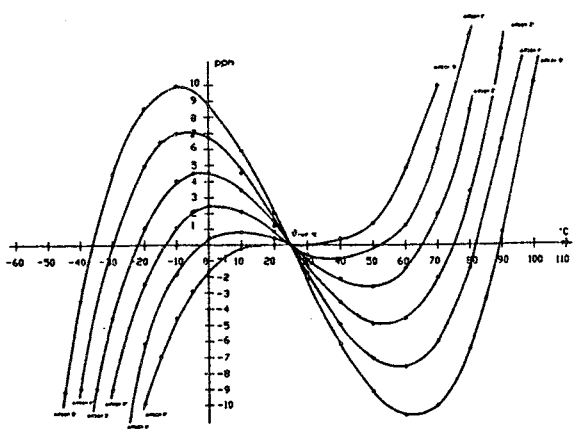


FIG.1 Frequency-temperature characteristics of AT-cut crystals.

These curves follow a cubic equation of the formula

$f = A1 * (t-to) + A2 * (t-to)^2 + A3 * (t-to)^3$

An improved stability over a wide temperature range will be obtained by using a varactor diode to change the load capacitance of the crystal and thus the crystal's thermal drift. In analogue TCXO's a thermistor-resistor network generates a temperature-varying voltage that changes the varactor capacitance in a predetermined manner to compensate the crystal frequency-temperature characteristics.

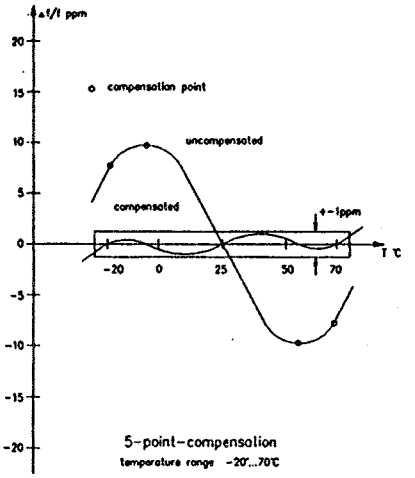


FIG.2 F-T characteristics of an oscillator with analogue compensation.

The analogue compensation is limited by the slope of the tuning voltage, the varactor and the temperature characteristics of the crystals and thermistors.

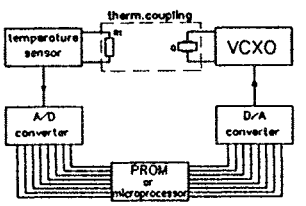


FIG.3 Block diagram of a DTCXO

By the digital compensation the analogue signal of a temperature sensor is digitized by an A/D converter and used as an address for a memory (PROM). For each temperature address the compensation voltage for the varactor is stored. The temperature compensation data in the PROM are read on the basis of the temperature data and converted to the corresponding analogue compensation voltage through a D/A converter in order to pull the oscillator back to the nominal frequency.

The functional course of the compensation voltage can also be approximated by a polynominal. The PROM is replaced by a microprocessor. The microprocessor has calculating capabilities and is able to reduce the storage capacity for the temperature data.

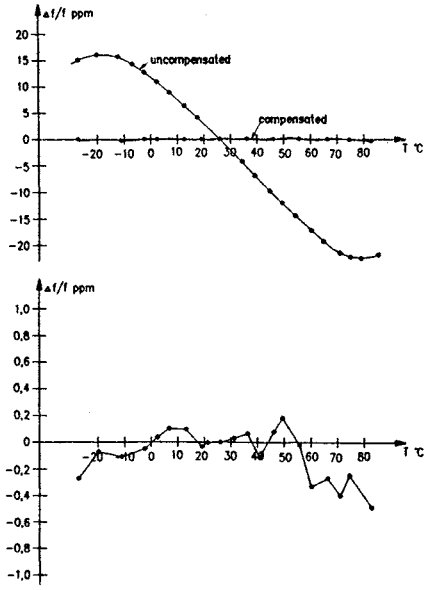


FIG.4 F-T characteristics of an oscillator with digital compensation.

Alternative design conception of DTCXO

All components which are necessary for a complete DTCXO increase the costs extremely. As in all modern radio equipments microprocessors are applied and the available capacity is rarely used to its limits, the new KVG conception uses the free capacities of the microprocessor inside the radio equipment.

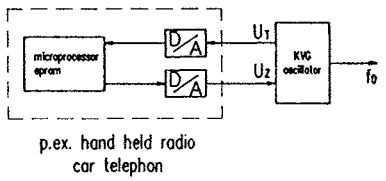


FIG.5 Components of the complete DTCXO.

The A/D- and D/A converter are also inside the radio equipment.
The oscillator consists only of a voltage controlled crystal oscillator (VCXO) and a temperature sensor, therefore the enclosure could be reduced to a DIL-14 package.

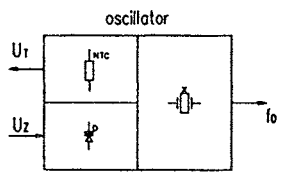


FIG.6 Schematic diagram of KVG oscillator D-0Z.

Figure 7 shows the temperature characteristics of an oscillator with digital compensation. A computer was used as processor. The oscillator was measured at 50 temperature points and compensated on these points.

The sawtooth can be improved by using interpolation formulas or more temperature steps. The problem for series-production are costly temperature measurements, especially the long time of measurements and the amount of temperature data for the PROM in the microprocessor. With an interpolation only 4 points are nescessary to calculate the temperature characteristics of an AT-cut crystal and therefore for an oscillator. The temperature sensor inside the oscillator supplies a voltage Ut which is equivalent

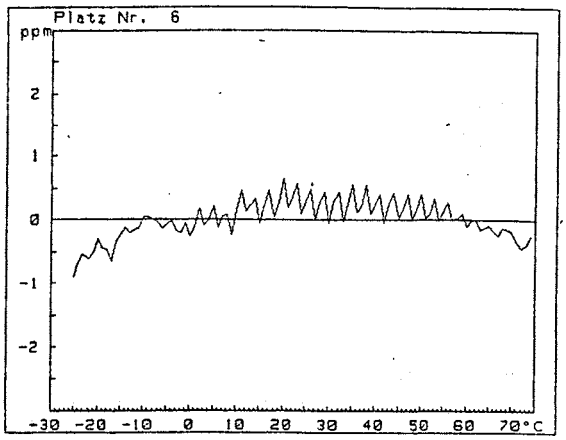


FIG.7 Frequency-temperature characteristics of a DTCXO with a computer as microprocessor.

to the temperature. The mathematic relation between the temperature sensor voltage Ut and the compensation voltage Uz can be expressed by the following cubic equation:

$$Uz = Ao + Ar * Ut + Br * Ut^2 + Cr * Ut^3$$

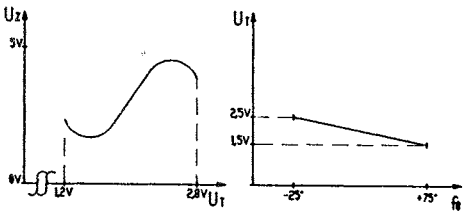


FIG.8 Relation between Uz and Ut.

In order to reach a better approximation of the function Uz = f (Ut) KVG has chosen 7 temperature measurement points to calculate the coefficients Ao, Ar, Br and Cr. The advantage of this methode is the short measurement time and only 4 coefficient have to be stored in the PROM of the microprocessor.

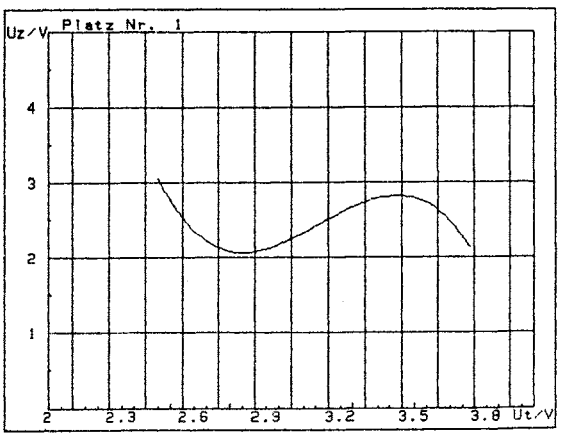


FIG.9 Uz = f (Ut) after a pre-measurement.

Figure 9 shows the calculated function Uz = f (Ut) after the temperature measurement of the oscillator.
The coefficients Ao, Ar, Br and Cr can be supplied by KVG to the user of the oscillator in 3 different versions, as marking on the case of the oscillator, as a barcode or on a disc. On a disc all parameters for 50 oscillators are stored.

Platz Nr. 4
DTCXO - NR.
14

Ao =194.972 Ar =-189.688

Br = 61.539 Cr = -6.5786

19497218968861538865786

FIG.10 Oscillator series number and coefficients Ao, Ar, Br and Cr.

The DIL-14 enclosure has only 4 pins. However 5 pins are nescessary for the whole oscillator(Ut, Uz, supply voltage, ground and RF-output). Therefore we use pin 1 for both suppply voltage and RF-output. Figure 11 shows the pin configuration of the oscillator D-02 and the external components.

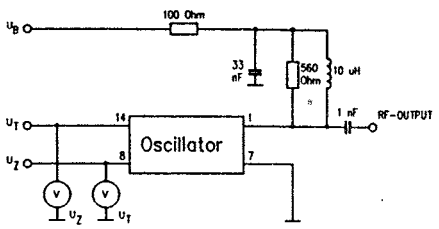
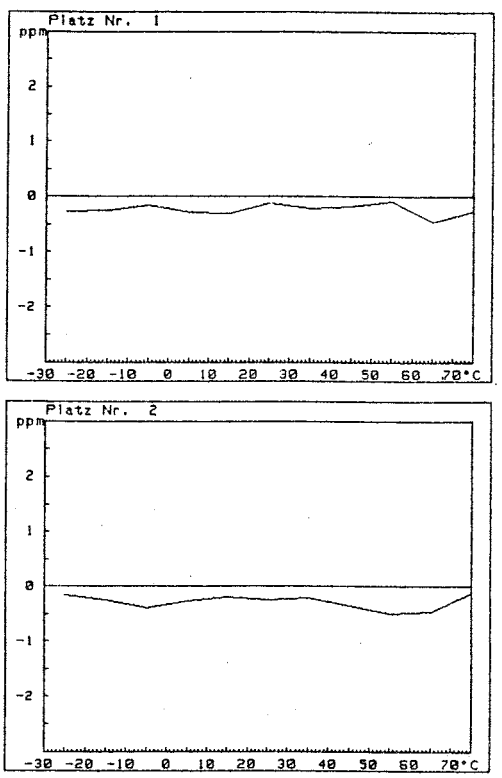


FIG.11 Connection diagram the oscillator D-02.

With this conception of a digital compensation it is possible to reach a stability lower than +- 1.5 ppm within the temperature range -25°C to +75°C without problems.



Ergebnisse der Endmessung (bezogen auf 25°C) 08:09:28

Platz Nr. 1
+ .118 ppm
-.334 ppm

Platz Nr. 2
+ .25 ppm
-.256 ppm

FIG.12 Final measurements results on 2 oscillators D-02 (KVG-test chamber).

Sources of errors.

1. Thermal gradients between crystal and temperature sensor.

   To minimize the thermal gradients between the temperature sensor and the crystal for the best thermal tracking it is nescessary to couple both thermally as tightly as possible.
   With an 8-bit D/A converter (256 steps) and a control range from 0V to 5V the solution of the D/A converter is 5V/256 steps = 19,5 mV/step.
   With a maximum tuning range of +- 46 ppm and a voltage range of Uz = 0,75 V ... 5,0 V (conditioned by the varactor) the minimal adjustment precision per step is given as follows:

   21,64 ppm/V * 19,5 mV/step = 0,42 ppm/step.

   The deviation of one step can cause an error of 0,42 ppm.

2. Linearity errors of the A/D- and D/A converters.

   Depending on the A/D - D/A converter errors of linearity can occur in the range of 5 mV (0,25 step) to 30 mV (1,5 step). A linearity deviation from 30 mV can cause an error of 0,63 ppm (1,5 steps * 0,42 ppm/step = 0,63 ppm).
   Therefore it is necessary to use A/D-D/A conerters with mimimal linearity errors (max. 10 mV).

3. Supply voltage changes.

   The supply voltage of the A/D-converter must be stabilized to min. 5V +- 1%. A supply voltage change about 50 mV can lead to frequency error of 1,08 ppm (2,56 steps * 0,42 ppm/step = 1,08 ppm)
   The disadvantage of this conception is that the voltage regulator drift over the temperarture range cannot be compensated.

Conclusion.

All errors discribed above can't be compensated as in a complete DTCXO. The available stability of this conception is not so good as in a DTCXO.
The comparison of figure 12 and figure 13 shows the difference between the temperature characteristics with compensated errors and the results after the oscillator is installed in a mobile radio.

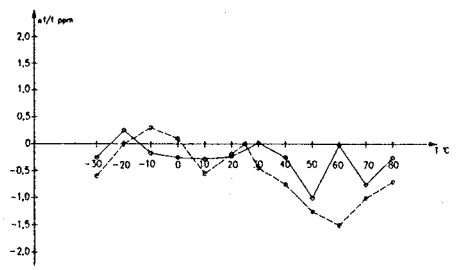
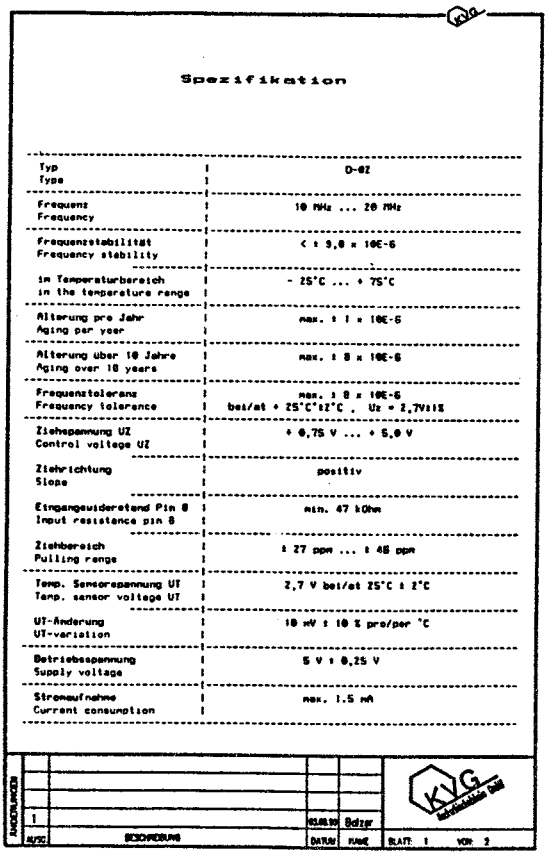


FIG.13 Frequences-temperature characteristics of the complet oscillator in a mobile radio.

Specification of the oscillator D-02



Spezifikation

Typ Type	D-02
Frequenz Frequency	10 MHz ... 20 MHz
Frequenzstabilität Frequency stability	< ± 5,0 x 10E-6
im Temperaturbereich in the temperature range	- 25°C ... + 75°C
Alterung pro Jahr Aging per year	max. ± 1 x 10E-6
Alterung über 10 Jahre Aging over 10 years	max. ± 8 x 10E-6
Frequenztoleranz Frequency tolerance	max. ± 8 x 10E-6 bei/at + 25°C±2°C , U_B = 5,7V±1%
Ziehspannung UZ Control voltage UZ	+ 0,75 V ... + 5,0 V
Ziehrichtung Slope	positiv
Eingangswiderstand Pin 8 Input resistance pin 8	min. 47 kOhm
Ziehbereich Pulling range	± 27 ppm ... ± 45 ppm
Temp. Sensorspannung UT Temp. sensor voltage UT	2,7 V bei/at 25°C ± 2°C
UT-Änderung UT-variation	10 mV ± 10 % pro/per °C
Betriebsspannung Supply voltage	5 V ± 0,25 V
Stromaufnahme Current consumption	max. 1.5 mA

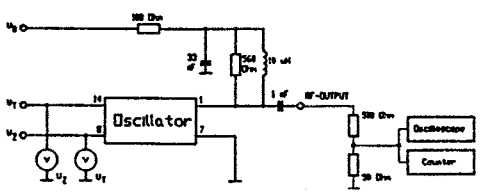
ÄNDERUNGEN	1		Bdzr	
	BLATT	BEZEICHNUNG	DATUM	NAME

KVG Quartz Crystal GmbH

BLATT 1 VON 2

Ausgangssignal Output signal	sinus, 500 mVpp
Last load	560 Ohm
Betriebstemperaturbereich Operable temperature range	- 35°C ... + 85°C
Lagertemperaturbereich Storage temperature range	- 55°C ... + 105°C

Testschaltung Test circuit

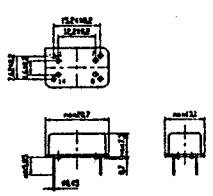


Pinbelegung/pin configuration

1. HF-Ausgang,Betriebsspannung UB
7. Masse,Gehäuse
8. Ziehspannung UZ
14. Temperatursensorspannung UT

1. RF-output, supply voltage UB
7. Ground, case
8. Control voltage UZ
14. Temperature sensor voltage UT

HF96A



BLATT 2 VON 2

References

1. Fry, Steven: digital temperature compensation for oscillators, RF-design 1990.
2. Albrecht, Harald: Digital temperature compensated oscillators, KVG-report 1989.
3. T. Miyayma, Y. Ikeda and S. Okano: a new digitally temperature compensated crystal oszillator for a mobile telephone system, 1989.
4. Frerkin, Marvin: crystal oscillator design and temperature compensation, Van Nostrand Reinhold Company, 1978.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 41

A LASER COOLED CESIUM ATOMIC FOUNTAIN: TOWARDS A HIGH PERFORMANCE CLOCK

André Clairon

Laboratoire Primaire du Temps et Fréquences
61 avenue de l'Observatoire, 75014 Paris, France

Christophe Salomon

Laboratoire de Spectroscopie Hertzienne
Département de Physique de l'ENS
24 rue Lhomond, F-75231 Paris Cedex 05

Saida Guellati

Laboratoire Aimé Cotton
Batiment 505, 91405 Orsay, France

William D. Phillips

National Institute of Standards and Technology, PHY B160, Gaithersburg, MD 20899, USA.

Abstract:

We report the realization of Zacharias's proposal for observing a Ramsey resonance in an atomic fountain. Launched up from a moving optical molasses where they have been cooled to a few microkelvins, the atoms pass once through a microwave cavity, continue to the summit of their trajectory, then fall again through the same cavity, completing the separated fields interaction. 2.0 Hz wide Ramsey fringes have been observed, corresponding to a 0.25 s ballistic time spent by the atoms above the cavity. The present S/N of 30 for a 1 s integration time would correspond to a 3 10^{-12} $\tau^{-1/2}$ short term Allan variance. With presently achieved atom density and temperature, one envisions cold cesium fountain clocks with unprecedented performances both in stability (10^{-16} / day) and accuracy (10^{-16}).

PACS NOS: 32.80.Pj, 42.50.Vk

Atomic motion is one of the fundamental limits to high resolution spectroscopy and a major source of inaccuracy in atomic frequency standards. The reduction of kinetic motion of atoms using momentum and energy exchange with laser fields has produced extremely low kinetic temperatures in recent years both for trapped charged particles and for neutral atoms [1,2]. These results offer new possibilities in the field of high precision physical measurements as well as for neutral atom frequency standards. Of particular interest is the primary time and frequency standard, the Cesium standard, using the ground state hyperfine transition F = 3, $m_F = 0$ (level $|1\rangle$ F = 4, $m_F = 0$ (level $|2\rangle$) at 9.192 GHz. Recent laser cooling experiments have produced cesium atoms at ~ 2 μKelvins, corresponding to an r.m.s. velocity of ~ 1 cm/s [3] which allows one to envision useable interaction times in the 1 second range. By contrast, conventional Cesium clocks use 400 K thermal beams and are typically limited to 10 ms interaction times. For freely moving atoms, the original Zacharias idea of using an atomic fountain seems, in the presence of gravity, the only realistic technique to achieve with the help of the Ramsey separate oscillatory field method, such long interaction times [4]. His proposal was to send an atomic beam upwards through a microwave cavity where a first interaction creates a coherent superposition of the $|1\rangle$ and $|2\rangle$ states. During the ballistic flight of the atoms above the cavity, this coherence freely precesses and is finally probed when the atoms, after having turned around, traverse again the same cavity downwards. This produces the well-known Ramsey fringe pattern in which the fringe width is inversely proportional to the ballistic time T spent by the atoms above the cavity. A 30 cm high fountain is required for achieving a 1 Hz fringe width.

A first atomic fountain was recently demonstrated in [5]: Na atoms were launched upwards by the radiation pressure of a pushing laser beam and a 2 Hz fringe width has been measured by using a pulsed microwave excitation, the atoms always staying inside the microwave region. In [6], laser cooled Cs atoms have been simply released from a laser trap and irradiated by two microwave pulses during their free fall, giving a 8 Hz wide fringe on the clock transition.

In this paper, we report the realization of the Zacharias fountain: cesium atoms confined in optical molasses are launched upwards at ≈ 5 μKelvins through a microwave cavity. 2.0 Hz wide Ramsey fringes have been obtained, which correspond to a 0.25 s total time spent by the atoms above the cavity . The atoms reach 6.4 cm above the cavity. The signal to noise ratio is 30 for an integration time of 1 second and should lead to a fractional stability of $3\times10^{-12}\ \tau^{-1/2}$ for an oscillator locked to such a signal. This short term stability would be comparable to or better than any present day Cs clock, and represents a factor 50 improvement over the previous Na fountain [5]. We also discuss potential improvements and give estimates for the stability and accuracy of a cold Cs fountain standard. In particular, we show that a 10^{-16} stability and accuracy for a one day integration time is a very realistic target for such a Cs fountain clock. This would then compare very favorably with the trapped ion microwave clocks which have virtually infinite interrogation times but have on the other hand a relatively poor S/N ratio imposed by the fact that only a very small number of ions may be used in high accuracy frequency standards [2].

Our experimental set-up has been partially described in [3,7]. Cesium atoms from a laser decelerated atomic beam are caught in a region of optical molasses [8] at the intersection of three orthogonal standing waves (Fig. 1). The molasses beams have an intensity of 5 mW / cm^2 and are derived from a frequency stabilized dye laser tuned -10 MHz below the $6\ S_{1/2}$ (g), F = 4 → $6\ P_{3/2}$ (e), F' = 5 transition. The horizontal beams are apertured to about 10 mm in diameter, whereas the vertical beams are limited to 8.5 mm by the holes in the microwave cavity. The vertical and one of the horizontal beam pairs consist of linearly and orthogonally polarized counterpropagating waves (Lin ⊥ Lin configuration). The other horizontal standing wave has a linear vertical polarization and will allow Zeeman or hyperfine optical pumping (see below). A repumping diode laser beam tuned to the g, F=3 → e, F' = 4 transition is combined with the two horizontal molasses beams. After loading the molasses for 0.5 s, about 7 10^6 atoms are confined and cooled in the ≈0.5 cm^3 molasses volume. A TE_{102} microwave cavity made of a piece of standard X band waveguide and shorted at both ends is located 2.7 cm above the molasses. Two holes have been drilled in the narrow side of the waveguide half a guide wavelength away from short-circuits ($\lambda_g / 2$). The microwave magnetic field is vertical. Its amplitude is zero on the lower and upper side of the waveguide and follows half a sinewave in between [9]. The cavity is overcoupled to produce a quality factor of 1000. The 9.192 GHz microwave frequency is generated by a Dielectric Resonator

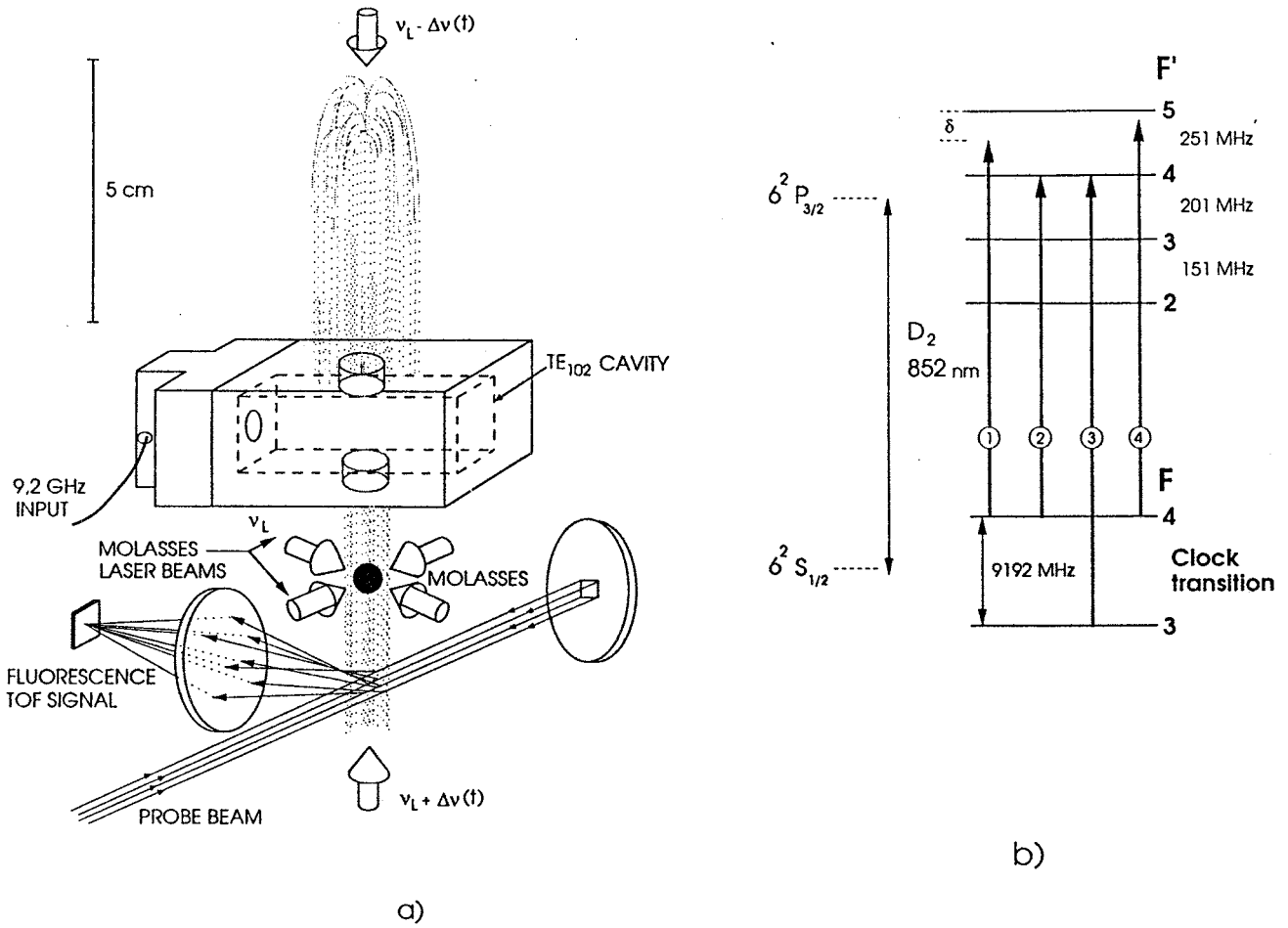


Fig. 1(a), Apparatus: Atoms from a laser-cooled atomic beam (not shown) are further cooled and confined in optical molasses at the intersection of three pairs of mutually orthogonal, counterpropagating laser beams. Then, by properly changing the frequencies of the vertical beams, one creates a 'moving'molasses that drags the atoms upwards and subsequently cools them to about 5.5 μK in a frame moving up at a chosen velocity. On their way up, the launched atoms interact a first time with the field of a microwave cavity which creates a coherent superposition of two atomic states. During the time T of the ballistic flight of the atoms above the cavity, this coherence freely precesses and is finally probed when the atoms, after having turned around, interact a second time with the cavity on their way down. This produces the Ramsey fringe pattern of Fig 4 for the transition probability between the two levels. The width of the fringes is simply 1/2T or 2.0 Hz for a 0.25 s ballistic time. The microwave resonance is detected optically by measuring the probe induced fluorescence when the atoms cross the probe beam on their way down.

(b) Relevant cesium transitions. Molasses loading and launching: 1 + 3. Hyperfine pumping: 2. Detection: 4.

Oscillator (DRO) that is phase-locked to the sum of the 918th harmonic of a 10 MHz Quartz oscillator and a tunable 12 MHz from a frequency synthesizer. The 9.2 GHz frequency is measured by a counter referenced to a rubidium clock. A weak ($B = 3\ 10^{-6}$ T) vertical magnetic field is applied to resolve the hyperfine transitions sensitive to B at first order and record only the clock transition sensitive to B at second order. The maximum field inhomogeneity over the whole fountain is smaller than $\pm\ 10^{-6}$ T.

Efficient detection of the atoms and precise temperature measurements are accomplished by an optical time of flight method (TOF). A probe laser beam having a 10×10 mm square crossection is centered 25 mm below the molasses center. It is tuned to the g, F = 4 $\rightarrow$ e, F' = 5 transition and drives the fluorescence of the atoms when they fall through the probe after their parabolic flight through the microwave cavity. The probe beam is derived from a linewidth narrowed diode laser, has an intensity of 0.15 mW / cm^2, is σ^+ polarized, retroreflected and detuned half a linewidth below resonance to minimize longitudinal heating . Each atom scatters an average of 10^4 photons during its ~ 10 ms long fall through the probe. With a collection efficiency of 10 %, the noise of the detection system per shot and in a 2 kHz bandwidth (needed for having a distortion free TOF signal) allows us to detect the signal of only 200 atoms with a S/N of 1. Therefore atom shot noise limited detection is expected for a number of atoms per shot greater than $4\ 10^4$.

Clearly the launching method is a crucial element for a fountain experiment: the transverse temperature of the launched atoms determines the number of atoms detected after the fountain through a given area, and hence the S/N ratio. The technique of 'moving molasses' [10,11,12,22] is a simple and powerful way to implement the launching while maintaining the very low molasses temperature. The idea is to create an optical molasses that is moving upwards at the required velocity: this is accomplished by shifting the frequency of the vertical beams while keeping unchanged the frequency ν_L of the horizontal beams. If the two vertical beams have the same intensity and if the upward traveling wave frequency is increased by $+\Delta\nu_L$ (frequency $\nu_L + \Delta\nu_L$) and the downward traveling wave frequency is decreased by $\Delta\nu_L$ (frequency $\nu_L - \Delta\nu_L$), this creates a moving standing wave at a velocity $v_z = \lambda \times \Delta\nu_L$. Because of the Doppler shift, atoms moving up with velocity v_z see the vertical beams at frequency ν_L the same frequency as the horizontal beams: this realizes a static molasses in the frame moving at v_z and therefore the temperature should be the same as for static molasses in the Lab frame. Now, if $\Delta\nu_L$ is slowly swept in time between 0 and a maximum value $\Delta\nu_L^{final}$, the atoms, initially at thermal equilibrium in the laboratory frame at a given temperature, will follow adiabatically the change of frequency and stay nearly at the same temperature in the moving frame: the temperature should not change during the launching process. In Fig.2, we show that this launching method is indeed able to produce a fountain of atoms at 5.5 ± 1 µK, with practically no loss of atoms during this process. The frequency difference $\Delta\nu_L$ is swept between 0 and 1.6 MHz in 1 ms and the molasses laser is at a detuning of - 2 Γ and high intensity (5mW / cm^2). With these parameters, the temperature during the acceleration phase is about 60 µK. Then a 1.6 ms cooling time is allowed at fixed $\Delta\nu_L$ (1.6 MHz) during which the molasses intensity is divided by four and the detuning increased to 70 MHz. During this phase the temperature drops to 5.5 µK. Special care has been taken in order to insure a perfectly symetric frequency sweep of the vertical beams, to avoid atom heating and fluctuations in the TOF mean arrival time. The measured shot to shot fluctuations in the vertical mean velocity are about 1 mm/s or $7\ 10^{-4}$ of the 1.5 m / s launching velocity.

Once the atoms have been launched and cooled to the desired temperature, a 2 ms long laser pulse, which is tuned to the g, F = 4 $\rightarrow$ e, F' = 4 transition and σ polarized with respect to the guiding magnetic field, quickly pumps all the atoms in the g, F = 3 level, with negligeable heating . See Fig. 3 which presents the detailed time sequence of the molasses loading + fountain cycle. Since the probe is tuned to the g, F = 4 $\rightarrow$ e, F' = 5 , the probe fluorescence is a measure of the microwave induced transition probability on the clock transition and the signal (area of the TOF peak) appears on a dark background. In order not to be limited by the shot to shot fluctuations in the total number of atoms in the molasses, which are rather large in our system (± 30 %), we measure, as in [6], not only the F = 4 population but also the total F = 3 + F = 4 population in each fountain cycle. We call $A_{F=4}$ the area of the first half of the TOF signal. At the central value of the TOF peak, a repumping laser beam, which is superimposed on the probe and detuned 1.5 GHz from resonance, is quickly (< 100 µs) made resonant with the g, F = 3, $\rightarrow$ e, F' = 4 transition so that the area $A_{F=3\ +F=4}$ of the second half of the TOF signal is a measure of the total number of atoms in a given fountain cycle. The clock transition probability is then simply the ratio $A_{F=4} / A_{F=3\ +F=4}$. See Fig. 3.

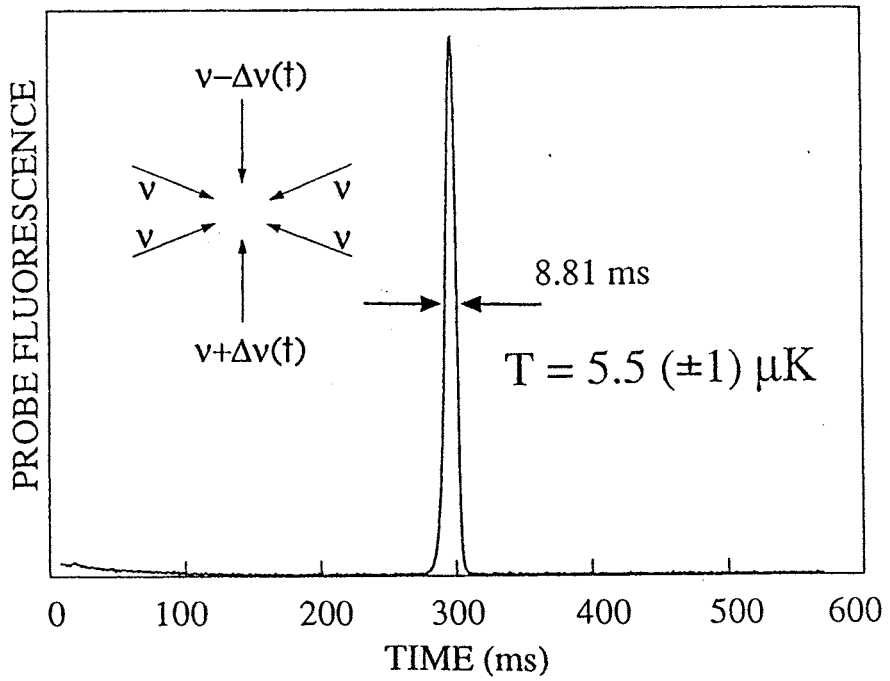


Fig. 2. Time of flight signal of Cs atoms launched at 1.4 m / s by the method of 'moving molasses' and subsequently cooled to 5.5 ± 1 μK. If the horizontal beams have the frequency ν_L, and if the vertical beams (±) have frequencies $\nu_L \pm \Delta\nu_L(t)$, one creates an optical molasses moving up at velocity $v_z = \Delta\nu_L(t) \times \lambda$ which drags the atoms upwards at the velocity v_z. If the acceleration dv_z / dt is not too large, the temperature of the atoms, which is initially at equilibrium in the molasses, will stay nearly at equilibrium during the acceleration phase. Here $\Delta\nu_L(t)$ is swept between 0 and 1.6 MHz in 1 ms, launching the atoms at $v_z = 1.4$ m / s. Then a 1.6 ms cooling phase at constant $\Delta\nu_L^{final}$ during which the molasses beam intensity I is divided by four and the detuning increased to 70 MHz cools the atoms to 5.5 μK. For I = 5 mW / cm^2 and a detuning of - 2 Γ, we find that for (dv_z / dt) < 2 10^3 m/s^2, nearly all the atoms of the molasses are launched and cooled. In order to reduce the geometrical contribution of the size of the molasses to the width of the TOF peak, a 3 mm high slice of the molasses is selected as in [3] by optically pumping all atoms of the molasses in g, F = 3 except those in a 3 mm high shadow. This is done in less than 1 ms by the F = 4 → F'= 4 beams.The probe beam vertical dimension is also reduced to 3 mm.

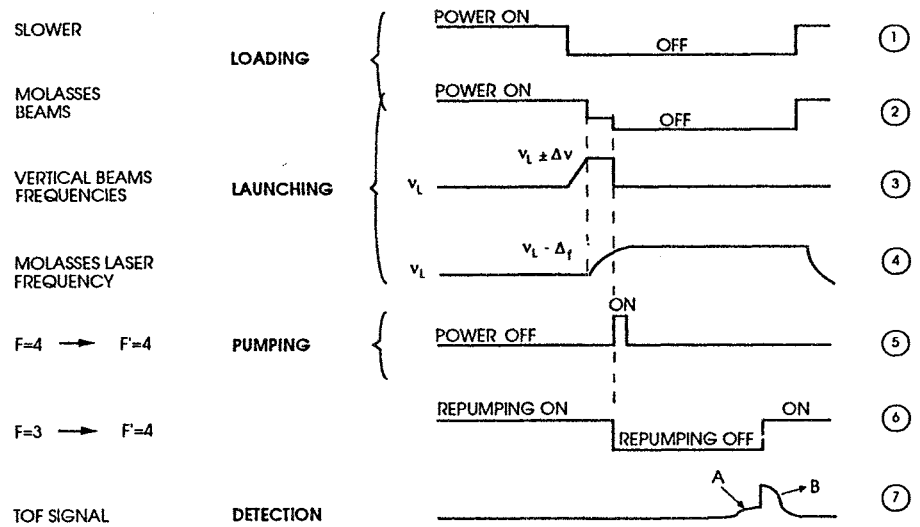


Fig.3. Time sequence and normalization procedure.The atomic fountain is operated in a pulsed mode. The total cycle time (1.1 s) includes molasses loading (0.5 s), launching (1 ms), cooling of the launched atoms (1.6 ms), hyperfine pumping (1-2 ms), fountain precession time and detection (0.6 s). In order to evaluate the transition probability in each fountain cycle , we first measure the F = 4 population (proportional to the area A) and then the total population in F = 3 and F = 4 (area B). The clock signal is then A / B. This is done by tuning a repumping laser into resonance with the g, F = 3 → e, F' = 4 transition at a fixed time near the peak of the TOF signal.

With no Zeeman optical pumping, one then expects that about 1 / 7 of the total number of atoms in the molasses contributes to the $F = 3, m_F = 0 \rightarrow F = 4, m_F = 0$ clock signal, in good agreement with the measured value (11 ± 2 %).

In order to illustrate the flexibility of the fountain method, Ramsey fringes on the clock transition are presented in Fig. 4 , with different resolutions and molasses temperatures. In Fig 4 a, the atoms are relatively hot (26 μK), and their trajectories reach only 3.4 cm above the cavity. The fringe width is 3.0 Hz and the Rabi pedestal is about 50 Hz wide. While the microwave power is optimum for the central fringe (transition probability ~ 1 at resonance), the contrast of the side fringes is not as good. This is due on the one hand to the distribution of interaction times, a combination of velocity ($\delta v / v = \pm 3$ %) and position distribution of the launched atoms, and on the other hand to the μwave field inhomogeneity across the 8.5 mm diameter holes (15 %).

In Fig.4 b, atoms are launched at the maximum height allowed by our vacuum chamber apparatus at a temperature of 5.5 μK. The atoms reach 6.4 cm above the cavity, the Ramsey interrogation time is 0.25 s and the resulting fringes have a full width of 2.0 Hz. Note the excellent contrast of the six central fringes. The dramatic improvement in the side fringe contrast is not due to the reduced velocity distribution ($\delta v / v = \pm 1.5$ %), but rather to the reduction of the effect of the position distribution because of the higher launch. Fig. 4 c presents the central fringe of Fig 4 a, recorded with a 0.2 Hz frequency step. Each data point is the average of 5 (molasses loading + fountain) cycles (5.5 s). The continuous line is a least square fit with the usual $1/2 [1 + \cos(\omega_{\mu w} - \omega_{at}) T]$ transition probability, which is valid for monokinetic atoms . The standard deviation is 2.8 % and is a measure of the actual S/N including both the fluctuations in the normalized fluorescence signal and the effect of the frequency noise of the 9.2 GHz source. The frequency (horizontal) standard deviation is 50 mHz or $5.5\ 10^{-12}$ and comparable to the measured Allan variance of our quartz crystal reference oscillator ($5\ 10^{-12}$ between 1 and 100 s). As a result, we believe that the present limitation to the S/N on this recording is the quartz frequency instability. We have indeed observed that the S/N improves when the Ramsey fringes are broader. In addition, in separate S/N measurements, for instance in conditions where the fringe contrast was deliberately deteriorated in order to reduce the effect of the quartz frequency noise, we measured a S/N of 30 for one fountain cycle (1.1 s). This should have given a S/N ratio of 70 for the conditions of Fig 4 c (where the integration time is 5.5 s), more than 2 times better than the measured one. Clearly a very good fly-wheel oscillator such as a H maser is needed for these high resolution fountain experiments.

The stability of a clock is conveniently characterized by the square root of the Allan variance [13,14]:

$$\sigma(\tau) = (\delta\nu / \pi \nu) \cdot N / S \cdot (T_{cycle} / \tau)^{+1/2} \qquad (1)$$

where $\delta\nu$ is the fringe width at half maximum, T_{cycle} the time for a complete molasses loading and fountain cycle ($T_{cycle} = T_{load} + T_f$) and τ the integration time in s. With a S / N ratio of 30 for one cycle ($T_{cycle} = \tau = 1.1$ s), measured under conditions where it is not limited by the quartz frequency instability, one gets a short term stability of $\sigma(\tau) = 3 \cdot 10^{-12}\ \tau^{-1/2}$ which is comparable or better than any existing primary or commercial Cs standard. Note that this stability is not limited by the atom shot noise (about $2\ 10^5$ atoms per toss contribute to the Ramsey fringe signal), but rather by technical fluctuations in the normalization procedure. If atom shot noise limited detection could be achieved in our present set-up, the short term stability of an oscillator locked to this Cs fountain would be $3 \cdot 10^{-13}\ \tau^{-1/2}$ or 10^{-15} per day. This would outperform by more than a factor 10 all existing Cs clocks. Since the number of atoms involved in the clock signal is susceptible of considerable improvements, for instance by producing a more intense atomic beam using atom funnels [11,15] and/or by using a magneto-optical trap to collect the slow atoms [5,6,16], we believe that a 10^{-16} stability per day is a realistic target.

To illustrate this point, it is interesting to discuss further expression (1) in the limit where the fountain time $T_f >>$ molasses loading time. If we are able to detect all the atoms after the fountain (for instance by using an open Fabry-Perot cavity and using a large area probe beam) and if the detection is limited by the atom shot noise, it is easy to show that $\sigma(\tau)$ decreases as the square root of the fountain time T_f. However a natural limit is set by the transverse residual Doppler effect that will tend to wash out the fringes after velocity integration. One would then need to resort to sub-Doppler techniques, and this discussion would apply for instance for two-photon Ramsey fringes in the optical domain [17].

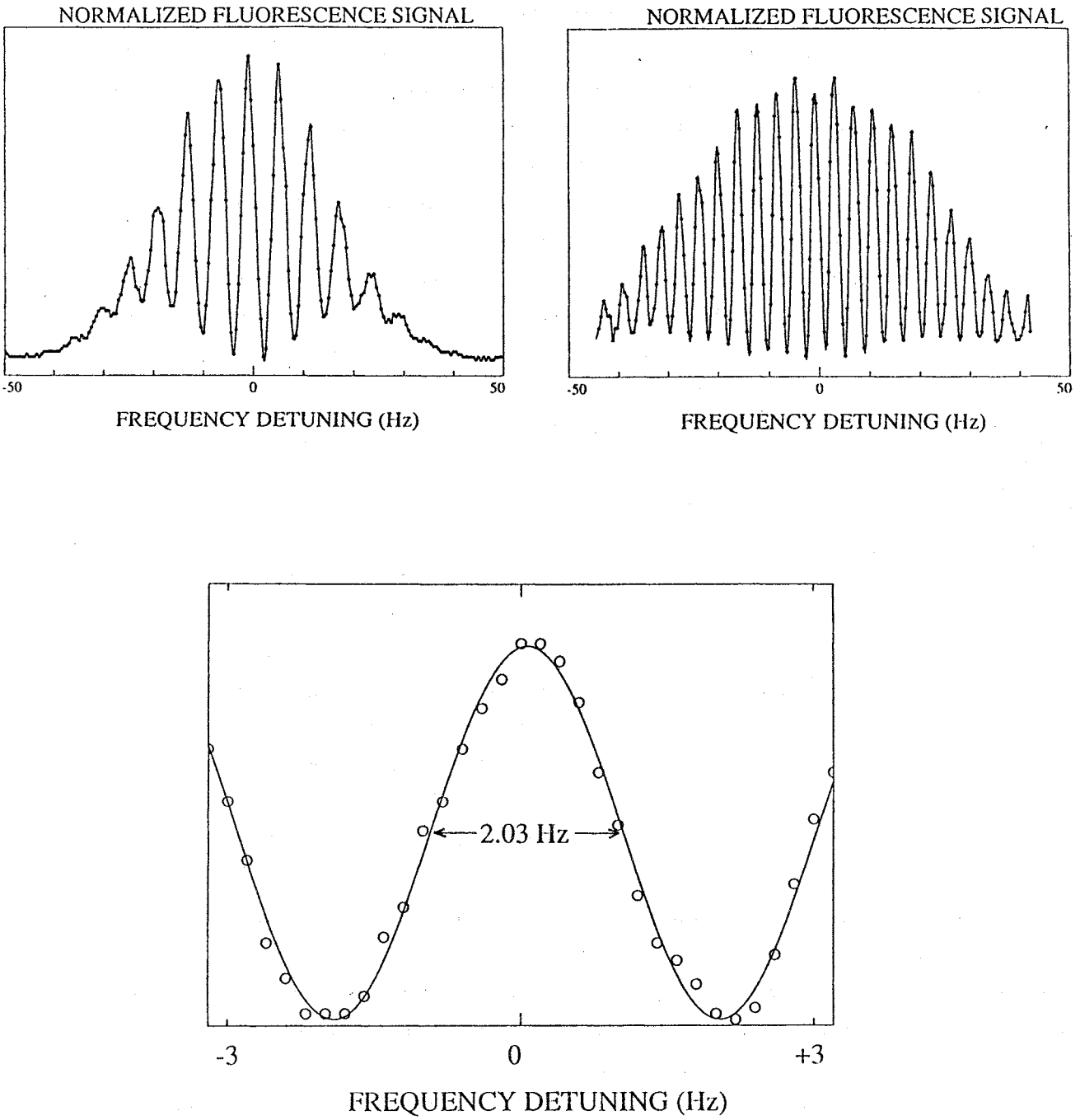


Fig.4. Cold fountain Ramsey fringes for two different conditions:

a) v_z = 1.19 m/s, T = 26 μK. The atoms reach 3.4 cm above the cavity. The fringe width is 3.0 Hz. Each dot is the average of 5 fountain cycles. 0.5 Hz per frequency step.

b) v_z = 1.41 m/s, T = 5.5 ± 1 μK. Fringe width 2.0 Hz.

c) Central fringe of b) with a 0.2 Hz frequency step. Each dot is the average of 5 cycles. The continous curve is a least square fit with the usual 1/2 [1 + cos($\omega_{\mu w}$ -ω_{at}) T] transition probability . The standard deviation is 2.8 %. The frequency (horizontal) standard deviation is about 50 mHz and is comparable to the Allan variance of our Quartz crystal reference oscillator (50 mHz for times between 1 and 100 s). We have chosen this variance for the diameter of the data points. The quartz frequency noise is the dominant limitation to the S/N ratio on this recording.

More serious for a clock in the microwave domain, is the effect of the residual Doppler efffect on the accuracy of such a clock: the unavoidable traveling wave component of the microwave field in the cavity induces a Doppler shift of the fringes (distributed cavity phase shift). Increasing the Q factor of the cavity would be a solution but one then approaches the maser threshold and cavity pulling effects become very severe [9]. One is thus led to reduce the interaction volume in the microwave cavity to regions where the phase is as constant as possible. This is done for instance, by having small holes ($r < 5$ mm) on the axis of a highly symetrical cavity such as a cylindrical TE_{011} cavity where the phase varies radially only quadratically. In this case, it is the transverse temperature which sets a limit to the number of detected atoms. Neglecting the initial molasses size and asssuming $v_{rms}\, T_f >>$ cavity hole radius r, we find for $r = \lambda_g / 8$ and for atom shot noise limited detection:

$$\sigma(\tau) = [\, (32/\pi^2)\ (k_B T / M\, c^2)(1/N_0)(T_f/\tau)\,]^{1/2} \qquad (2)$$

where $N_0$is the initial number of atoms in the $F = 3$, $m_F = 0$ state, M the Cs mass, and k_B the Boltzmann constant. This shows that the stability degrades only as the square root of the temperature and improves, as expected, as the square root of the initial number of atoms per molasses toss and of the number of tosses. With presently achievable atom number (10^7) and temperature (5 μK), we find $\sigma(\tau) = 2\ 10^{-14}\ \tau^{-1/2}$ which equals 10^{-16} per day of integration. This short term stability would be equivalent to that of the best H masers.

Not only is the stability potentially very high, but also the potential accuracy of such a Cs fountain is very attractive. Most of the limiting factors of a conventional clock are considerably reduced [18]: the second order Doppler shift and the gravitational red shift are small ($\approx 10^{-17}$) and calculable. The residual first order Doppler effect or distributed cavity phase shift is proportional to the fringe width. Having a phase shift as small as that in [19] would give an effect of about 10^{-16}. . In addition the use of a single cavity suppresses the effect of unequal electrical lengths in the two arms of a conventional clock. For $Q_{cav} = 10^4$, $N_0 = 10^7$ and a 1 Hz fringe width, the power dissipated by the cavity is only 10^4 larger than the power absorbed by the atoms in the cavity. As a consequence, the cavity pulling effect is no longer proportional to $(Q_{cav} / Q_{at})^2$ as usual in Cs clocks, but tends to be proportional to Q_{cav} / Q_{at}. The relative cavity frequency detuning must then be controlled to better than 10^{-7} to reach a 10^{-17} accuracy level [9]. Interestingly, the maser oscillation could easily be achieved with a superconducting cavity. Finally a magnetic field map with a 10^{-10} T resolution over the ≈ 1 cm^3 volume of the launched atoms can simply be recorded by measuring the frequency of a field dependent transition as a function of the height of the fountain. With a constant B field of 10^{-7} T, the resulting error is less than 10^{-16}. As the clock runs in a pulsed mode, no laser light-shift is present. On the other hand the black-body shift [20] must be considered ($-\ 1.69\ 10^{-14}$ at 300 K) and the temperature of the fountain region must be known to within ± 0.5 ° to allow a 10^{-16} uncertainty level. Obviously collisions could also become a problem at this level of precision. Changing the initial atomic density over a broad range is easily achieved with such a fountain and this should give access to the measurement of collisional shifts of the clock transition. Whatever the ultimate source of limiting systematic errors, the excellent linewidth and frequency stability of a cold atom fountain clock should allow unprecedently accurate evaluation of these errors.

Finally, most of the atomic beam manipulation techniques presented here can be extended to optical frequency standards in the visible or uv domain [21], with perhaps even more frequency stability and accuracy capabilities.

We wish to thank Alain Aspect, Claude Cohen-Tannoudji, J. Dalibard for helpful contributions and discussions. The C.N.E.T. is aknowledged for the loan of a Rb clock. This work was supported in part by Collège de France, DRET, and BNM.

REFERENCES

* Laboratoire de spectroscopie Hertzienne is an Unité de recherche de l'Ecole Normale Supérieure et de l'Université Paris 6, associée au CNRS - U.R.A. 18)

1. See the special issues of J. Opt. Soc. Am. B2 (1985) and B6 (1989). C. Cohen-Tannoudji and W. D. Phillips, Physics Today, Oct 1990, p. 33.

2. D.J. Wineland, W.M. Itano, J.C. Bergquist, J.J. Bollinger, F. Diedrich and S. Gilbert, 4th symp. on Freq. Standards and Metrology, ed. A. de Marchi, p. 71, Springer verlag (1989). W. M. Itano, J. C. Bergquist, and D. Wineland, Science **237**, 612 (1987).

3. C.Salomon, J.Dalibard, W.D.Phillips, A.Clairon, and S.Guellati, Europhys.Lett.,12, 683, (1990).

4. N. F. Ramsey, Molecular beams Oxford University Press, Oxford (1985).

5. M. Kasevich, E. Riis, S. Chu and R. de Voe. Phys.Rev.Lett.**63**, 612 (1989).

6. C. Monroe, H. Robinson, and C. Wieman, Opt. Lett.**16**, 50,(1991)

7. C. Salomon and J. Dalibard, C. R. Acad. Sci. Paris, **306**, 1319 (1988). J. Dalibard, C. Salomon, A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, in Atomic Physics 11, S. Haroche, J.-C. Gay, and G. Grynberg, eds. (World Scientific, Singapore, 1989) p. 199.

8. S. Chu, L. Hollberg, J. Bjorkholm, A. Cable, and A. Ashkin, Phys. Rev. Lett. **55**, 48 (1985).

9. J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Adam Hilger (1989).

10. P. Lett, W. Phillips, S. Rolston, C. Tanner, R. Watts, and C. Westbrook, J. Opt. Soc. Am. B6, 2084 (1989).

11. E. Riis, D. Weiss, K. Moler and S. Chu, Phys. Rev. Lett, **64**, 1658, (1990).

12. C.Salomon, J.Dalibard, W.D.Phillips, A.Clairon, and S.Guellati, 12th Int. Conf. Atomic.Physics, Ann Arbor, (1991)..

13. D.J. Allan, Proc. IEEE, **54**, 221 (1966).

14. W Phillips et al. 4th Eur. Time and Freq. Forum, Neuchatel (1990).

15. J. Nellessen, W. Werner and W. Ertmer, Opt. Commun.**78**, 300 (1990).

16. D. Grison, B. Lounis, C. Salomon, J.Y. Courtois and G. Grynberg, to appear in Euro. Phys. Lett. (1991).

17. G. Beausoleil and T. Hansch, Phys. Rev. A **33**, 1661, (1986).

18. A. de Marchi, Metrologia, **18**, 103 (1982).

19. A. Bauch, and T. Heindorff, 4th symp. on Freq. Standards and Metrology, ed. A. de Marchi, Springer verlag (1989).

20. W.M. Itano, Phys. Rev. A, **25**, 1233, (1982)

21. J.L. Hall and M. Zhu, J.Opt.Soc.Am. B6, 2194 (1989).

22. J.L. Hall and M. Zhu presented at Q.E.L.S. Anaheim (1990), unpublished.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 42

Environmental Sensitivity and Long Term Stability Limitations Induced by C-field Variations in Commercial Cesium Beam Frequency Standards*

A. De Marchi E. Rubiola
Politecnico di Torino — Torino, Italy

Abstract

Measurements of sensitivity of C-field to different environmental factors are reported, as taken on a small sample of commercial standards, and results are related where possible to the characteristics of the electronics or of the physical package of the standard under test. It is shown that effects on the C-field may well be responsible for residual long term frequency fluctuations of the standard, once the sensitivity to microwave power (and hence its changes) is removed.

1 Introduction

Both long term frequency stability and environmental sensitivity of existing commercially available Cesium beam frequency standards were in recent years recognized to be in most cases related to microwave power sensitivity caused by the pulling effects of neighbouring transitions [1] to [7]. In fact it was shown that performances can be improved in both perspectives by operating a standard in such a way that its microwave power sensitivity be minimized. This can be done, for example [5], by setting the C-field at a value for which these pulling effects cross zero, which they usually do several times [1,7] in their oscillating behaviour versus C-field. By doing so, long term stabilities of the order of 10^{-14}, and temperature sensitivities of $2{\cdot}10^{-14}$/K, were routinely obtained with supertube standards [2]. The question arose of what would be the next most important limitation in these respects, once the standard is made insensitive to microwave power.

In this paper we report on measurements of stability and environmental sensitivity of C-field taken on a small number of commercial standards. Due to the limited sample set of devices available, the results are obviously not conclusive, however they can be considered representative at least of the model of instrument that we have examined. The effects on the C-field of temperature, humidity, and magnetic field variations were measured by locking the standard under test to the $m = 2$ neighbouring line. The resulting effects on the output frequency in normal operation can be induced from these results.

It is shown that the above mentioned limits found for long term stability and temperature sensitivity are in agreement with variations that can be introduced by this mechanism. This is also true for previously unexplained [8,9,10] sensitivity to environmental humidity found for such standards in recent years.

2 Measurement technique

In order to acquire information on environmental coefficients of C-field in a meaningful way, two types of measurement procedures were followed: in one, variations of one parameter at a time were imposed on the standard under test, while keeping the other parameters that were thought to be relevant under active or passive control; in the other one, the standard was observed for a long time together with the relevant environmental parameters, correlations and variances were then calculated for all the experimental data. With the first technique short term data can be taken, so that transitory behaviour can be observed and time constants evaluated; with the second, a clear picture of what processes are most relevant can be obtained with a statistical analysis.

The standard's C-field value was monitored during the tests by closing the frequency loop onto the $m = 2$ field dependent transition central Ramsey fringe signal by substituting an external synthesizer for the standard's fixed 12.6 MHz internal synthesizer. In this way the nominally even 5.0 MHz output carries the desired information on C-field variations, although attenuated in relative value by a factor $2f_z/\nu$, which amounts to $1.16{\cdot}10^{-5}$ at the field for which the Zeeman frequency is 53 kHz (typical for commercial clocks).

The effect induced by C-field changes on the frequency standard in normal operation can be inferred

*Work performed at NIST in Boulder, CO, with the support of NIST and US Naval Observatory.

by the formula:

$$\frac{\delta\nu}{\nu} = 4\left(\frac{2f_z}{\nu}\right)^2 \frac{\delta f_z}{fz}$$

This means that the amplification obtained in the observation of C-field changes by locking to the $m = 2$ line instead of the $m = 0$ line is $\nu/8f_z$, or $2.15{\cdot}10^4$ at $f_z = 53$ kHz.

3 Results on magnetic field variations

Because a connection between residual long term instabilities, annual periodic terms in the frequency variations, and environmental sensitivity of C-field was suspected, coefficients were measured for those environmental parameters which exhibit an annual term, namely temperature, humidity, and the earth's magnetic field. The annual term of the first two is obvious, and depends on location; for the third it is caused by the eccentricity of the earth's orbit and it's similar almost everywhere on the globe, at the level of about 30 nT peak to peak in locations sufficiently removed from the poles.

Shielding factors against external field and its variations were measured for the component parallel to the C-field inside the tube, which is vertical in the tested standards. An external mmf was imposed by driving current in a winding placed around the tube. With the standard inside a magnetic shield, and initially degaussed, increasing field intensities were imposed, up to 3 g. After each step up small cycles were performed down from the reached value and back up. In this way an estimate was obtained of the shielding factor for variations of the external field, which depends on the reversible permeability of the shield material. Then the reached maximum field value was reversed twice in order to obtain an estimate of the shielding factor for fixed fields, which depends on the permeability of the shielding material. Results obtained for the tube under test are summarized in Fig.1, where the dependence of the two types of shielding factors on the mentioned permeabilities is obvious.

From the reversible shielding factor obtained for the tube under test, an annual term of the order of 10^{-17} peak to peak is expected for this standard, due to the earth's field yearly cycle. Although standards contributing to TAI may not all be this well shielded, it appears unlikely that the annual term in TAI (about $2{\cdot}10^{-14}$ peak to peak) be produced by this mechanism. Exceptional μT bursts, however, may show at the 10^{-15} level even in very well shielded tubes.

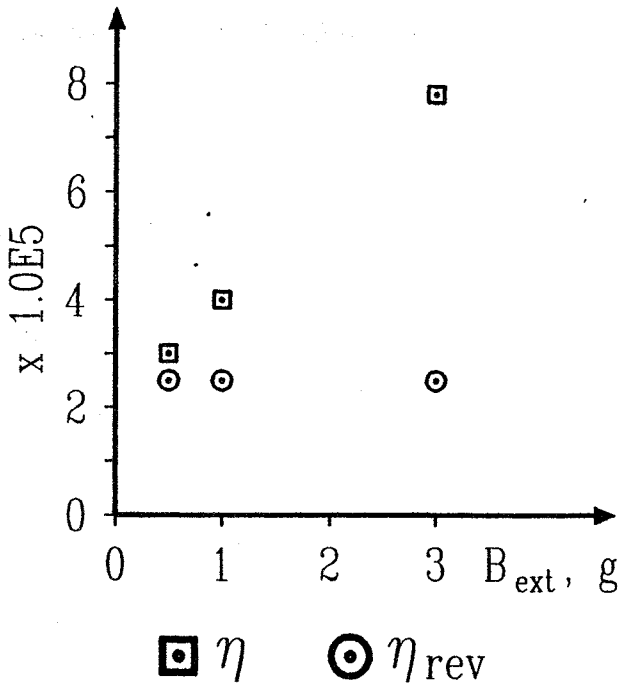


Figure 1: Measured magnetic shielding factors for continuous (η) and varying fields (η_{rev}) of a commercial Cesium beam tube (Hewlett Packard option 004).

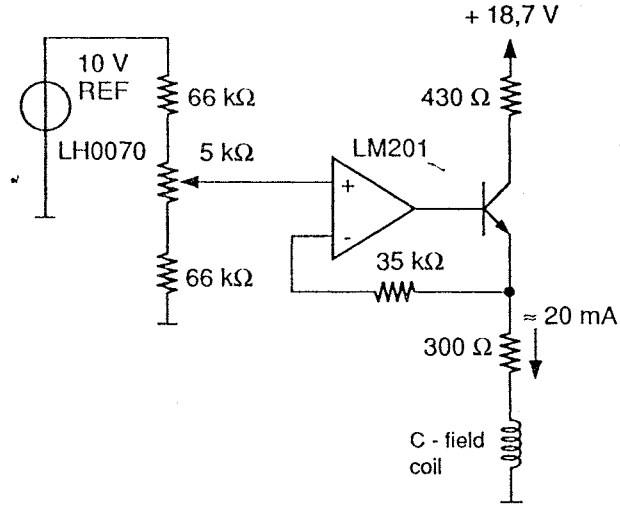


Figure 2: C-field supply (Hewlett Packard).

4 Expected stability of the C-field supply

The C-field circuit, shown in Fig 2, is a voltage source driving the coil, and therefore is sensitive to load variations.

Three main mechanisms can be identified for temperature sensitivity of C-field:

- power supply voltage changes,
- coil resistance variations,
- dilation of the inner magnetic shield.

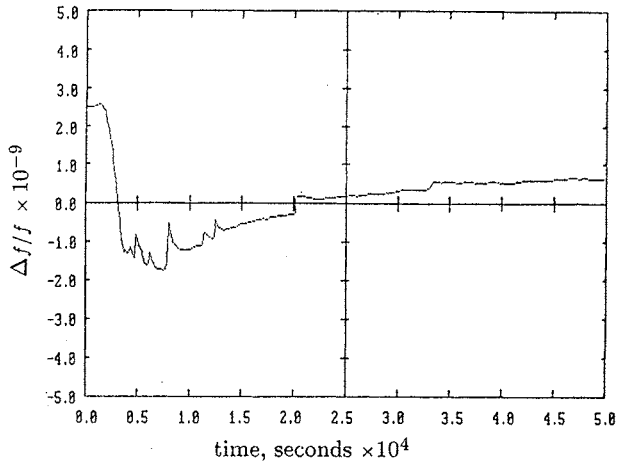


Figure 3: Temperature sensitivity of the standard HP0523 locked to the $m = 2$ transition. Conditions: constant humidity (35%), 8 °C temperature step (55 to 63 °C); measured: HP1449 – HP0523.

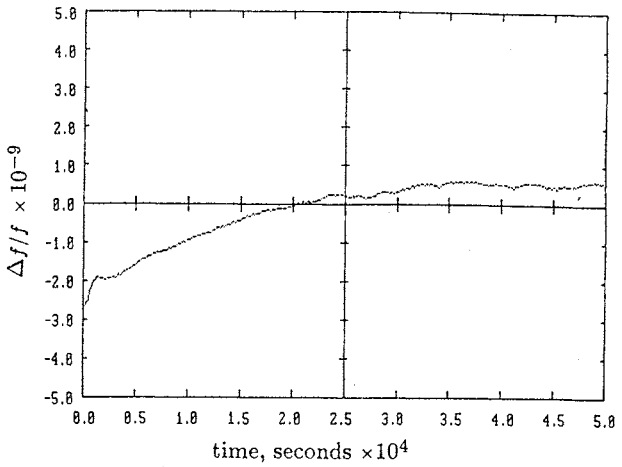


Figure 4: Humidity sensitivity of the standard HP0523 locked to the $m = 2$ transition. Conditions: constant temperature (63 °C), 30% humidity step; measured: HP1449 – HP0523.

The expected power supply voltage changes due to temperature variations are of the order of 10^{-6}/°C; this value comes from the devices data sheet. A more important effect is due to the load or coil resistance thermal coefficient, of the order of 10^{-3}/°C. The distance between the two opposite inner shield walls comes directly in the A/m computation. The effect of this dilation can thus be expected to be of the order of 10^{-5}/°C. It must be pointed out that the time constant of these last two phenomena is expected to be much higher than that of supply variations because coil and shields are inside the tube.

As for the flicker noise, it can be assumed that it all come from the power supply. Assuming a typical flicker noise of 20 μVpp from 0.1 to 10 Hz, the coefficient $h_{-1} \simeq 2{\cdot}10^{-12}$ V^2/Hz can be calculated for the voltage reference. At the Op-Amp input the resulting flicker coefficient is $h_{-1} \simeq 1{\cdot}10^{-12}$ V^2/Hz, greater than the flicker of the operational amplifier.

As a consequence of the voltage reference instability, the expected flicker level of the C-field is $\sigma = 2.5{\cdot}10^{-7}$ ($m = 2$). This agrees with the flicker noise of $3{\cdot}10^{-12}$ measured for $\tau \leq 100$ s on the $m = 2$ line. The flicker level which can be inferred for the clock transition is 10^{-16}.

5 Results on temperature variations

In order to evaluate the frequency variations, the standard under test (HP532) was placed in a temperature controlled chamber, while the other one was taken as reference. Five sides of the chamber were insulated, the sixth being a thin plastic cover hermetically sealed.

A temperature step $\Delta t = 8$ °C was given to the chamber, and the temperature was monitored with a sensor (AD-590) placed on the operational amplifier; unfortunately no temperature sensor was placed in this experiment on the tube input flange in order to get an estimate of the C-field coil temperature.

The result of this experiment (Fig. 3) shows two superimposed effects with different time constants. The fast response mechanism ($\tau_1 \simeq 2{\cdot}10^3$ s) is probably a variation in the electronics, while the slow one ($\tau_2 \simeq 2{\cdot}10^4$ s) is most likely a variation inside the tube, probably in the coil resistance (the sign would agree and the temperature variation would be about 50 mK if it were a coil resistence change).

The relative frequency change is $5{\cdot}10^{-10}$/°C for the fast phenomenon, and $2{\cdot}10^{-10}$/°C total, when the time constant effects are exhausted. An overall temperature coefficient of $1{\cdot}10^{-14}$/°C can be inferred for the clock transition, in agreement with previous results on standards that were made insensitive to the microwave power [2].

After analyzing long term data taken with the standards in the NIST time scale environmental chambers for about 40 days, no significant correlations have been found between frequency and temperature. The temperature effect is probably masked by humidity driven variations.

6 Results on humidity variations

The effect of humidity has been tested with the same experimental setup as in the previous section, keeping the temperature at 63 °C and changing the chamber humidity. The humidity was varied from 5% to saturation at the plastic wall temperature; this corresponds to a humidity step of 30% inside the chamber.

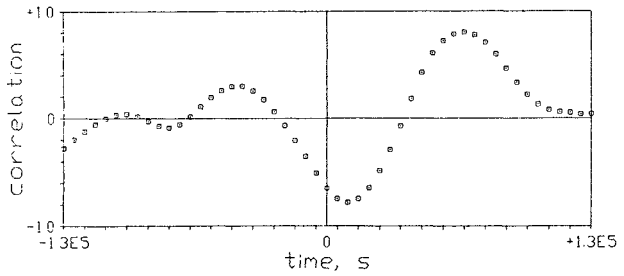


Figure 5: Correlation coefficient of the frequency standards HP0523 and HP1449, both locked to the $m = 2$ transition, as function of the delay introduced on the data of the first with respect to the second.

The results of the test, performed on one standard (HP0523) are shown in Fig 4. The output relative frequency changes by $-3{\cdot}10^{-9}$, with a time constant of $2{\cdot}10^4$ s. The coincidence of this time constant with the slow mechanism of temperature effects leads one to think that the two may be related to the same phenomenon.

The humidity coefficient of this standard turns out to be $-1{\cdot}10^{-10}/\%$ at the neighbouring line, or $-5{\cdot}10^{-15}/\%$ on the clock transition.

Another set of measurements was taken in long term with both standards in the time scale environmental chambers. In these, the relative humidity, although controlled, changes by a couple of percent peak to peak with a noticeable dayly term.

Correlations coefficients $\rho_{xy}(\tau)$ between humidity variations and the frequency of the two standards were computed (Fig. 6), as well as the correlation between the standards (Fig. 5).

The high correlation peak in the latter (about 0.8) implies that the frequency variations of the two standards are due to the same physical phenomenon. On the other hand, both turn out to be correlated to the chamber humidity. Delays of the standards from humidity variations agree in these correlation plots, as well as signs of the humidity coefficients.

Another estimate of the humidity coefficient can be obtained from the Allan variance measured on the $m = 2$ transition, under the assumption that humidity is responsible for most of the frequency fluctuations, by comparison with the variance of the humidity. The result for HP532 is a few units in $10^{-10}/\%$, in fair agreement with the coefficient obtained with the other method.

This coefficient is similar to those obtained in other laboratories, which are reported for comparison in Tab. 1. This suggests that the humidity sensitivity of Cesium standards may be traceable to C-field related problems.

Measured $\sigma_y(\tau)$ for the two standards are shown in Fig. 7, where scales are given for both $m = 2$ (measured) and $m = 0$ (inferred) transitions.

7 Conclusions

Preliminary results on measurements of stability and environmental sensitivity of C-field were reported, as taken on a small set of commercial Cesium beam standards by monitoring Zeeman frequency variations while changing environmental parameters or keeping them as constant as possible. The focus of this inquiry was on identifying the sources of long term frequency instability when microwave power sensitivity is unimportant, and the mechanisms which can determine annual terms in Cesium clocks. In order to check the hypothesis that both would be related with C-field sensitivity to environmental parameters, the latter was measured specifically for those parameters which show an annual term. The effect of humidity was found to be rather strong, probably acting through heat conductivity driven thermal balance variations in the standard, while variations of temperature and Earth's magnetic field were found to be probably less relevant for most commercial devices in the environ-

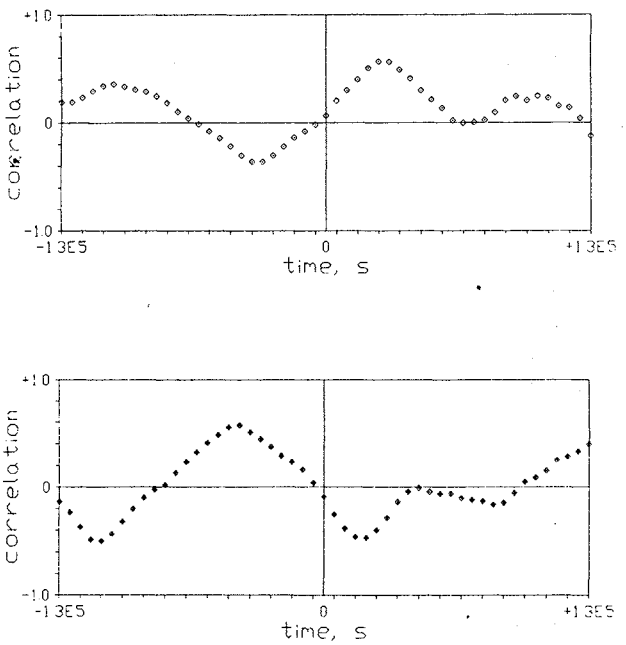


Figure 6: Correlation coefficient of chamber humidity and frequency standards locked to the $m = 2$ transition, as function of the delay time. Top: HP523. The peak indicates that the time delay is about $2{\cdot}10^4$ s and that the coefficient is negative; A delay periodicity of the humidity can be desumed from this plot. Bottom: HP1449. A delay of $3{\cdot}10^4$ s and a positive correlation can be found.

Reference	Standard	f_z (kHz)	Humidity effects on $m = 0$	Humidity effects on $m = 2$	Measurement conditions
IEN [8]	CS5	42.8	$+24{\cdot}10^{-15}$ / (g/m^3) ¶	$+1.6{\cdot}10^{-10}$ / % §	Temp. 24 °C $m = 0$
	CS6	42.8	$+22{\cdot}10^{-15}$ / (g/m^3) ¶	$+1.5{\cdot}10^{-10}$ /% §	
	CS9	42.8	$-48{\cdot}10^{-15}$ / (g/m^3) ¶	$-3.2{\cdot}10^{-10}$ / % §	
NIST [9]	HP323	53.5	$+3.6{\cdot}10^{-15}$ / % ¶	$+7.7{\cdot}10^{-11}$ / % §	Temp. 23 °C $m = 0$
	HP324	53.5	$-6.1{\cdot}10^{-15}$ / % ¶	$-1.3{\cdot}10^{-10}$ / % §	
	HP2165	53.5	$-4.4{\cdot}10^{-15}$ / % ¶	$-9.5{\cdot}10^{-11}$ / % §	
	HP2315	39.0	$-1.6{\cdot}10^{-15}$ / % ¶	$-3.4{\cdot}10^{-11}$ / % §	
	HP1316	53.5	$+3.8{\cdot}10^{-15}$ / % ¶	$+8.2{\cdot}10^{-11}$ / % §	
	HP352	42.8	$-11{\cdot}10^{-15}$ / % ¶	$-3.0{\cdot}10^{-11}$ / % §	
NAOM [10]	HP1315	53.5 (?)	$+5.4{\cdot}10^{-15}$ / % ¶	$+1.1{\cdot}10^{-10}$ / % §	$m = 0$
	HP885	53.5 (?)	$+5.4{\cdot}10^{-16}$ / % ¶	$+1.1{\cdot}10^{-10}$ / % §	
This paper	HP523	53.5	$-5{\cdot}10^{-15}$ / % §	$-1{\cdot}10^{-10}$ / % ¶	Temp. 63 °C $m = 2$
	HP1445	53.5	$+2{\cdot}10^{-16}$ / % §	$+4{\cdot}10^{-12}$ / % ¶	
§inferred ¶measured					

Table 1: Comparison between humidity sensitivities of Cesium beam standards reported in previous papers and humidity sensitivity of C-field measured in this work.

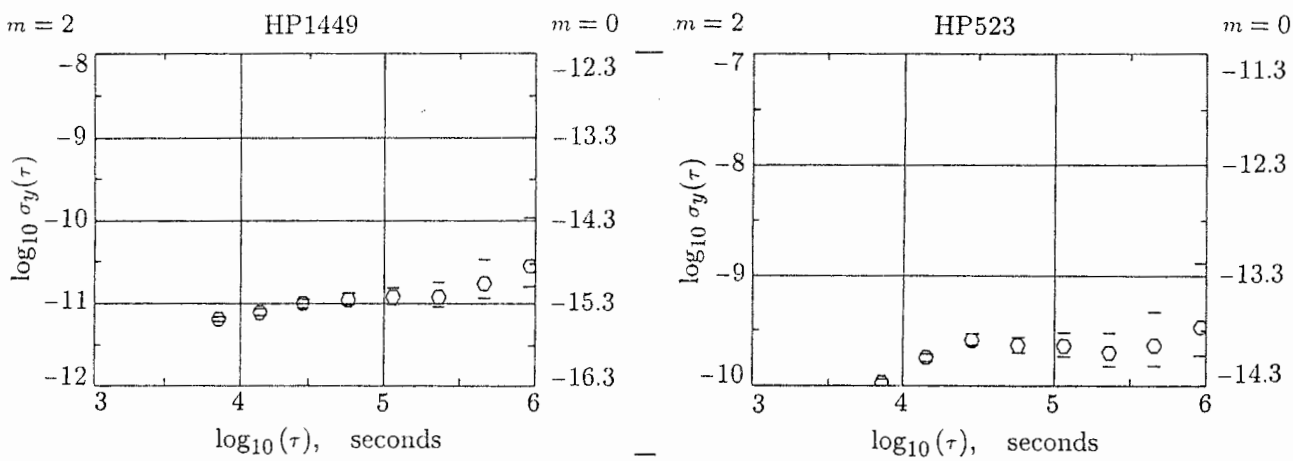


Figure 7: Square root of the Allan variance measured with the standards in the environmental chambers locked to the $m = 2$ transition (left scale) and inferred for the $m = 0$ transition (right scale).

mental conditions in which they are usually kept.

Long term stability limitations observed in standards aligned for optimum microwave power insensitivity are in agreement with what can be induced as an effect of long term C-field instability, and previously unexplained observed sensitivities to humidity can be accounted for by the kind of coefficients measured for this effect.

The conclusions can be drawn that C-field variations are responsible for long term instability when this is not limited by microwave power variations, and that correlated annual terms in the ensemble of commercial clocks which contribute to TAI are most likely driven by C-field sensitivity to the environment, in particular to humidity, as already pointed out [8,9].

References

[1] A. De Marchi, G. D. Rovera, A. Premoli, "Pulling by Neighboring transitions and its effects on the performance of Cesium beam frequency standards," *Metrologia*, vol. 20 (1984).

[2] A. De Marchi, "Understanding environmental sensitivity and ageing of Cesium beam frequency standards," in *Proc. 1st EFTF*, Besançon (1987).

[3] A. De Marchi, "New insights into long term frequency instabilities (drift & noise) of Cesium beam frequency standards," in *Proc. 41st FCS* (1987).

[4] A. De Marchi, "Rabi pulling and long term stability in Cesium beam frequency standards," *IEEE Trans. on UFFC*, vol. UFFC-34 np. 6 (1987).

[5] A. De Marchi, "A first account of long term stability results obtained on various Cesium standards by the power sensitivity minimization technique," in *Proc. 20th Annual PTTI*, Washington DC (1988).

[6] A. De Marchi, "Improving the long term stability of Cesium beam frequency standards," in *Proc. 4th EFTF*, Neuchâtel (1990).

[7] L. S. Cutler, C. Flory, R. P. Giffard, A. De Marchi, "Frequency pulling by hyperfine σ-transitions in Cesium beam atomic frequency standards," *Journal of Appl. Phys.*, March 1 (1990).

[8] E. Bava, F. Cordara, V. Pettiti, P. Tavella, "Improving the long-term stability of a time scale generated by commercial Cesium clocks," in *Proc. 2nd EFTF*, Neuchâtel (1988).

[9] J. E. Gray, H. E. Machlan, D. W. Allan, "The effect of humidity on commercial Cesium beam atomic clocks," in *Proc. 42nd FCS* (1988).

[10] P. Tavella, C. Thomas, "Study of the correlations among the frequency changes of the contributing clocks to TAI," in *Proc. 4th EFTF*, Neuchâtel (1990).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 43

Majorana Transitions: Their Effect on the Velocity Distribution of Cesium Atoms in an Atomic Beam

A. Bauch, T. Heindorff

Physikalisch-Technische Bundesanstalt Braunschweig

Abstract

In a cesium atomic clock, state-selected atoms leaving the polarizer magnet are used as an atomic beam in the resonance experiment. The population of the Zeeman sublevels regarding total number and velocity usually varies, depending on the particular design of the magnet systems used. During the passage through inhomogeneous magnetic fields, transitions among the Zeeman sublevels are induced which entail a change of the velocity distributions of the atoms in a specific quantum state. They can be detected by studying the shape of hyperfine transition lines (F=4)-(F=3). Given the spatial course of the fields, the changes can be calculated analytically or numerically. The results are in agreement with experimental findings.

1. Introduction

The SI unit, second, is realized with cesium atomic beam frequency standards, generally referred to as cesium clocks. The basic design principle of these devices has been well known since the sixties, but there still remain fields of insufficient knowledge in the description of their physical properties. One field which has been identified is the influence of Majorana transitions induced in the atomic beam /1/. When the state-selected atomic beam passes weak inhomogeneous portions of magnetic field, transitions among the Zeeman sublevels of a total angular momentum F state are induced. They are named after Ettore Majorana, who was the first to give a theoretical treatment on this subject in 1932 /2/. Reports on a correlation between Majorana transitions and frequency shifts of the clock transition frequency realized in a cesium clock have been given in the literature /3, 4, 5/. In the PTB this field has been actively researched since 1984. The essential advantage of the experiments done is that the course of the magnetic fields in the device used is sufficiently well calculable. A detailed comparison between theoretical and experimental findings is thus possible.

The subject of this paper is the studies made on the influence of Majorana transitions on the velocity distributions of the atoms in the various F, m substates. A proper understanding of this still limited field helps first to confirm that Majorana transitions do occur in a beam tube and secondly to prove that means to prohibit them are successful. The research is of course targetted on a quantitative understanding of the large frequency shifts which have also been observed during the experiments done at the PTB. The methods of treatment described in section 4 will be useful in this respect /6/.

Some of the material is taken from the author's PhD thesis /7/ which contained the first summary of the work done until then. The experiments and the theoretical treatment have both been refined since then.

2. Outline of the experimental set-up

2.1 The CSX atomic beam apparatus

The experiments described here have been carried out using the CSX apparatus, the PTB's experimental frequency standard. In Fig. 1 a plot of its vertical section is given. A point beam source and a point detector in an axial symmetrical arrangement are used. The properties of the atomic beam are governed by the polarizer and analyzer magnets. They consist of quadrupole and hexapole magnets. The C field is generated by a solenoid surrounded by a magnetic shield. A table of characteristic data of the CSX is given below. A more detailed description can be found in /7, 8/.

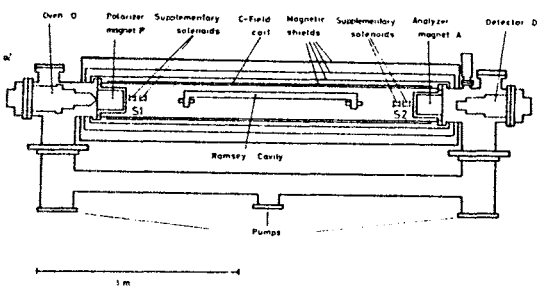


Fig. 1 Vertical section of the experimental frequency standard CSX of the PTB.

Table 1 Characteristic data on the CSX

Total beam current	100 pA
Ramsey peak amplitude (4,0)-(3,0)	14 pA
Fullwidth at half maximum of the line (4,0)-(3,0)	260 Hz
Mean velocity of atoms in the state (4,0)	416 m/s
Halfwidth of velocity distribution	typ. 60 m/s
Mean C field strength B_0	10 to 16 μT
Magnetic field inhomogeneity in the cavity region (relative units)	$5 \cdot 10^{-4}$

Two pairs of small solenoids, one close to the exit of the polarizer magnet and one close to the entrance of the analyzer magnet, were mounted. Thus static magnetic fields can be generated in S1 and S2, and the influence of field inhomogeneities known from the ge-

ometry of the coils and the currents used can be studied.

Three cases can be distinguished:
i) with a suitable choice of the C field and the auxiliary fields, magnetic fields between polarizer and analyzer can be so realized that no indications of Majorana transitions are obvious. This is the CSX's regular mode of operation and used, for example, for the measurement of the initial velocity distributions of the beam atoms.
ii) A combination of the above fields leads to a zero point in the longitudinal component of the magnetic field. This is dealt with in section 3.
iii) It can be arranged that at each point, S1 or S2, two zeros in the longitudinal field occur. This gives rise to obvious changes in the state population which are fully explainable now as described in sections 4 and 5.

2.2 Velocity distributions
A velocity distribution $\varrho_m(v)dv$ of atoms in each quantum state (4,m), m=-3...+3 has been obtained by a Fourier analysis of the Ramsey pattern of the corresponding (4,m)-(3,m) transition, which itself was recorded in the CSX's normal mode of operation.

The shape of $\varrho_4(v)dv$ is assumed to be equal to that of $\varrho_3(v)dv$ with its mean velocity $\langle v\rangle_4$ and the total number n_4 adapted to the experimental findings for m= -3, ... +3. The atoms in the state (4, -4) are perfectly deflected out of the beam, meaning $\varrho_{-4}(v)dv = 0$.

Table 2 Population n_m of the (4,m) levels and mean velocity $\langle v\rangle_m$ of the (4,m) atoms in the CSX atomic beam in the normal mode of operation.

m		-4	-3	-2	-1	0	1	2	3	4
n_m	[%]	0	100	97	91	84	78	73	67	55
$\langle v\rangle_m$	[m/s]		374	391	400	416	429	443	450	459

In table 2 the relative population of the m levels and mean velocities are given as they have been used as input parameters in all calculations. The eight velocity distributions are plotted in Fig. 2. They are calculated with a 4 m/s stepwidth.

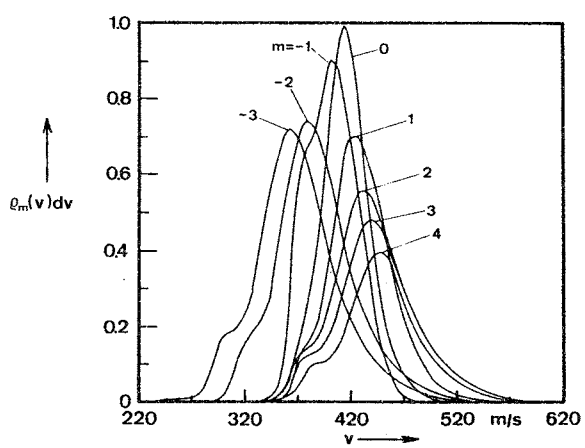


Fig. 2 Velocity distribution $\varrho_m(v)$ dv of atoms in the state (4,m) in the CSX.

The obvious effect of Majorana transitions is to modify these distributions. For quantitative comparisons of value the following assumption must be made:

The distributions $\varrho_m(v)dv$ obtained as described are really those of the atoms leaving the polarizer magnet. Their modification by the transfer function of the analyzer-detector system is neglected. Admittedly this has never been proved as being permissible.

3. Single rotation of the quantization axis in space

3.1 Outline of the analytical treatment
Ettore Majorana treated the passage of a spin 1/2 particle through a region where the direction of the magnetic field changes. The situation is illustrated in Fig. 3. Assuming an atom starting in the state m=+1/2 where m is the spin projection quantum number referred to the local magnetic field, the probability of a spin flip to m'=-1/2, $|T(1/2 \to -1/2)|^2$, is given by /2/

$$|T(1/2 \to -1/2)|^2 = \sin^2(\alpha/2) = \exp\{-\varkappa\cdot\pi/2\}. \quad (1)$$

$\varkappa$ in equation (1) is defined as

$$\varkappa = \omega_L(B_t) \cdot B_t / (dB_z/dt) . \quad (2)$$

Here ω_L is the Larmor angular precession frequency, $\omega_L(B) = g\cdot\mu_0\cdot B/\hbar$, where the g-factor of cesium is 1/4 and μ_0 is the Bohr magneton. The fields are as given in Fig. 3. The expression for $\varkappa$ contains the ratio of the Larmor frequency at point Q divided by the change of B_z in Q with time as seen by the atom passing with a velocity v. To describe an experiment it is necessary to know the atom's velocity v, the field gradient dB_z/dz and B_t. Then we obtain

$$\varkappa = g\mu_0 \cdot B_t^2 / (v \cdot dB_z/dz \cdot \hbar) \quad (3)$$

$\varkappa$ can take values from 0 to ∞. In the non-adiabatic limit, $\varkappa$=0, the spin direction is fixed in space and reversed with respect to the local magnetic field. The adiabatic limit, $\varkappa$=∞, describes the situation in which the spin follows the direction of the magnetic field.

As in equation (1) α is one of the Eulerian angles, it is possible to transform a result obtained for a spin 1/2 system to that of the real system of interest. This is done using equation (4) which has been taken from Bloch and Rabi /8/:

$$T_{mm'} = \left[(F+m)!(F+m')!(F-m)!(F-m')!\right]^{1/2} \cdot$$
$$\cdot i^{2F-m-m'} \exp[i(m'\Phi + m\Psi)] \sin^{2F}(\alpha/2) \cdot \quad (4)$$
$$\cdot \sum_k (-1)^k \frac{\cot an^{m+m'+2k}(\alpha/2)}{(m+m'+k)!(F-m'-k)!(F-m-k)!k!} .$$

The transition probabilities $|T_{mm'}|^2$ used in all calculations throughout this paper are given in /1, 7/.

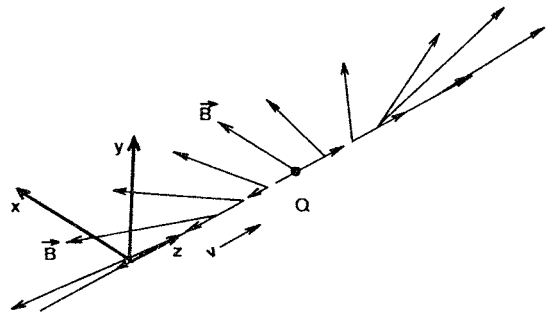


Fig. 3 Single rotation of the quantization field B in space. In point Q the field is purely transversal, B_t, (or zero) and the quantities dB_z/dz and B_t have to be determined for use in equation (3).

3.2 Experimental findings

Using the coils in S1 of the CSX we have generated a field as described above which closely approximates the non-adiabatic limit. In addition in the region of point Q an independent transversal field B_t can be applied, allowing a variation of κ and $\sin^2(\alpha/2)$. When this is done, the total beam signal is reduced and the shape of the Ramsey patterns altered. In detail there are changes in the velocity distributions of the atoms in the state (4,m) and of the linewidth of the corresponding transition. In general the velocity distributions of all atoms are seen to become broader and almost equal if $\sin^2(\alpha/2)$ approaches a value of 0.5. In the experiment $\sin^2(\alpha/2)$ cannot precisely reach the value one, since B_t cannot be zero across the whole beam cross section. Even if it is assumed that the applied fields are purely longitudinal (B_t=0), then the constraint div B=0 imposes

$$\langle B_t^0 \rangle = -0.5 \cdot \langle r \rangle \cdot dB_z/dz \qquad (5)$$

where <r> is the average radial distance of the atoms from the axis of the solenoids (z-axis). A value of <r> of 1 mm is reasonable in the CSX. This leads to a maximum value of $\sin^2(\alpha/2)$=0.99. A minimum value of $\sin^2(\alpha/2)$ of 0.003 was estimated in a similar way.

An example of this is shown in Fig. 4 where a set of five velocity distributions $\varrho_0(v)dv$ is plotted. They were deduced from Ramsey patterns which were recorded with a field configuration described by the indicated value of $\sin^2(\alpha/2)$ present at S1 in the CSX. The broadening is obvious, but the shift of the mean velocity is rather small. In Fig. 5 the result of a measurement and a calculation of the mean velocity $\langle v \rangle_0$ are compared. In the adiabatic and non-adiabatic limit, the two values of $\langle v \rangle_0$ agree, but between these limits distinctive changes in $\langle v \rangle_0$ are found. The discrepancies are probably due to inadequate definition of the initial conditions or the modelling of the fields present in the CSX. In general the calculated velocity distributions are broader than the experimental ones, which indicates the presence of the velocity filtering effect of the analyzer magnet. An inclusion of atoms in state (4,-4) would not be of help. In Fig. 6 the calculated and experimentally derived velocity distribution $\varrho_0(v)dv$ in the case of $\sin^2(\alpha/2)$=0.855 are overlaid. It is informative to see how after the passage through the inhomogeneous fields the distribution $\varrho_0(v)dv$ is composed of atoms which initially started in the other m substates.

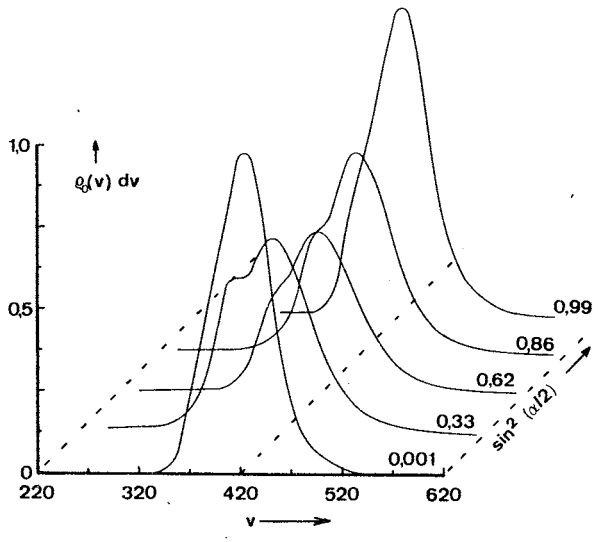


Fig. 4 Velocity distributions $\varrho_0(v)$ dv in the CSX evaluated from the clock transition Ramsey pattern under the condition of a single rotation of the quantization axis being present in S1, characterized by $\sin^2(\alpha/2)$.

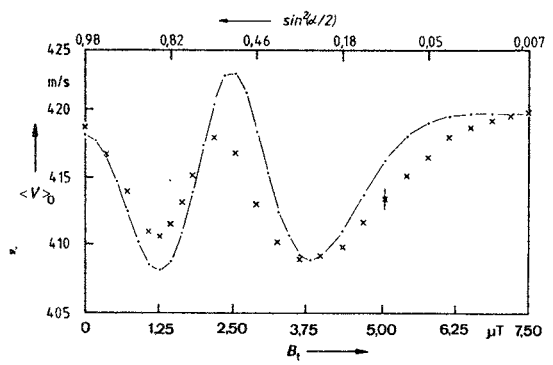


Fig. 5 Mean velocity $\langle v \rangle_0$ of atoms in the state (4,0) under the condition as in Fig. 4
x x deduced from the measured Ramsey linewidth
• • calculated using the distributions from Fig. 2 and $\sin^2(\alpha/2)$ as input parameters
B_t is the transversal magnetic field additionally applied near point Q.

Experiments and calculations have also been done on the atoms with other m values. Here one is again faced with the same problem as above: If $\sin^2(\alpha/2)$ is varied from 0 to 1, then the role of atoms with m and -m is interchanged. For example, unlike the normal mode of operation, (see table 2), in the non-adiabatic case the fastest atoms enter the analyzer magnet in the state m=-3. The analyzer magnet does not now focus all the "converted" atoms onto the detector. This is the main reason why the total beam signal decreases by as much as 26 % in the non-adiabatic mode of operation.

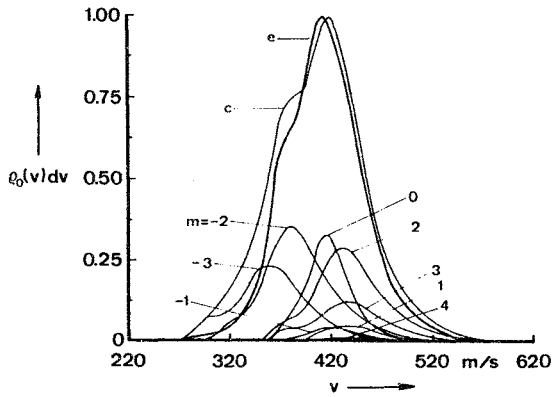


Fig. 6 Velocity distribution $\varrho_0(v)\,dv$ under the condition as in Fig. 4 with $\sin^2(\alpha/2) = 0.855$.
e: experimental curve
c: calculated with the velocity distributions from Fig. 2 and $\sin^2(\alpha/2)$ as input parameters. Curve c is the sum of the individual contribution of atoms having left the polarizer magnet with the indicated quantum number m.

The treatment gets even more complicated when the magnetic field rotation occurs behind the cavity at S2. Then atoms with quantum numbers F=3 and F=4 pass this region and their passage through the analyzer has to be known. However, a general rule has been formulated which seems to be applicable to CSX, at least at first glance /7/:
If a Ramsey pattern R_m of atoms in state m is recorded with the field rotation before the cavity, then the same shape is found for R_{-m} if the field rotation takes place behind the cavity. Fig. 7 illustrates this finding.

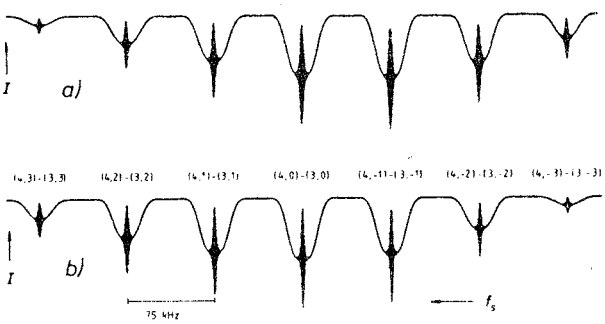


Fig. 7 Spectrum of σ transitions in CSX when point Q lies in S1(a) or in S2 (b). I: beam signal, f_s: probing microwave frequency.

4. Double rotation of the quantization axis in space

4.1 A spin-1/2 particle in a spatially varying magnetic field: numerical calculation

A relation is sought similar to equation (1) where the transition probability $|T(1/2 \to -1/2)|^2$ is related to the fields traversed by the atoms which are now allowed to be of an arbitrary shape. The procedure has been adopted from a paper by Schröder and Baum /9/.

The atom's state is described by

$$|\Psi\rangle = C_+(t)\,|U_+\rangle\exp\{-iE_+t/\hbar\} + C_-(t)\,|U_-\rangle\exp\{-iE_-t/\hbar\} \qquad (6)$$

where the $|U_\pm\rangle$ are the eigenfunctions of the time-independent Hamiltonian, H_0, which describes the basic atomic properties. In the Schrödinger equation

$$i\hbar\,(d/dt)|\Psi\rangle = (H_0+H_1(t))|\Psi\rangle \qquad (7)$$

the interaction term is assumed to be

$$H_1(t) = -g\cdot\mu_0\cdot B(t)\cdot S \qquad (8)$$

where S is the spin vector operator and B(t) is the magnetic field vector. B(t) is a static field which becomes time-dependent by the movement of the atom only. The motion is assumed to be parallel to the z-axis as in Fig. 3 so that $B(t) = (B_x(t),\ 0,\ B_z(t))$ and $B(t) = |B(t)|$.

Insertion of the relations (6) and (8) into (7) finally leads to the coupled differential equations for the evolution of the coefficients $C_\pm(t)$, with the abbreviation $\gamma = g\mu_0/\hbar$:

$$\dot{C}_+(t) = i\gamma\left(B_z(t)\cdot C_+(t)+B_x(t)\cdot C_-(t)\right)/2$$
$$\dot{C}_-(t) = i\gamma\left(-B_z(t)\cdot C_-(t)+B_x(t)\cdot C_+(t)\right)/2 \qquad (9)$$

The fields are given in steps of 0.1 mm. The transversal field $B_x(t)$ is calculated using relation (5). For ease of integration, the transformation to the interaction representation is made by the relation (10).

$$c_\pm(t) = C_\pm(t)\cdot\exp\{(-i)(\mp 1/2)\int_0^t \gamma B_z(t')dt'\} \qquad (10)$$

This leads to a new set of differential equations (11).

$$\dot{c}_+(t) = i\gamma B_x(t)\cdot c_-(t)\cdot\exp\{-i\int_0^t \gamma B_z(t')dt'\}/2$$
$$\dot{c}_-(t) = i\gamma B_x(t)\cdot c_+(t)\cdot\exp\{i\int_0^t \gamma B_z(t')dt'\}/2 \qquad (11)$$

It is numerically integrated with the step-width dt=0.1 mm/v, where v is the atom's velocity.

4.2 Application and results

The procedure has mostly been used in the case of two zeros in B_z close together, a situation which is illustrated in Fig. 8. The atoms start at z=0. At z=200 mm the field is homogeneous and the coefficients c_+, c_- no longer change. Starting with the initial conditions $c_+ = 1$, $c_- = 0$, $|c_-|^2$ is calculated at z = 200 mm and identified as $|T(1/2 \to -1/2)|^2$ in equation (1). An atom with quantum numbers F=4,m will be found with m changed to m' with a certain probability as determined by equation (4).

Before the presentation of results one limitation of this method should be pointed out.

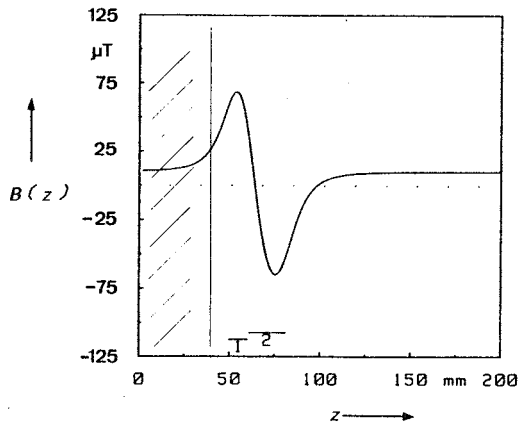


Fig. 8 Calculated magnetic field component $B_z(z)$ as a superposition of the constant C field and the fields produced by the coils 1 and 2. The true field deviates grossly from the calculated one in the dashed region (see text).

The fields as illustrated in Fig. 8 are calculated as the sum of the constant C field and the fields produced by the coils as obtained from the geometry and the currents used. The fact that the ends of the pole tips of the polarizer magnet are located at about z=30mm is neglected, means that the fields there grossly deviate from those shown in Fig. 8. The stray fields of the magnet will also change the fields. The results presented now depend to a very large extent on the integral field, or in a classical picture, the number of precessions of the spin in the magnetic field between the two zeros. Inevitably, no detailed quantitative agreement can be obtained. However, it has been shown that many experimental findings can be qualitatively understood /7/.

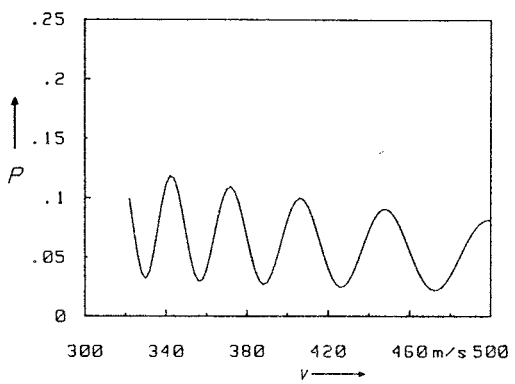


Fig. 9 Calculated transition probability P, $P = |T(1/2 \rightarrow -1/2)|^2$ of a spin 1/2 atom with the g-factor of cesium, 1/4, and with velocity v passing a magnetic field as shown in Fig. 8.

Result 1: If a fixed setting of the fields is considered, then the probability $|T(1/2 \rightarrow -1/2)|^2$ is an oscillating function of the atomic velocity v as depicted in Fig. 9. The period which appears reflects the decrease in the number of precessions of the spin in between the two zeros by 1 going from lower to higher velocities as the time of flight decreases.

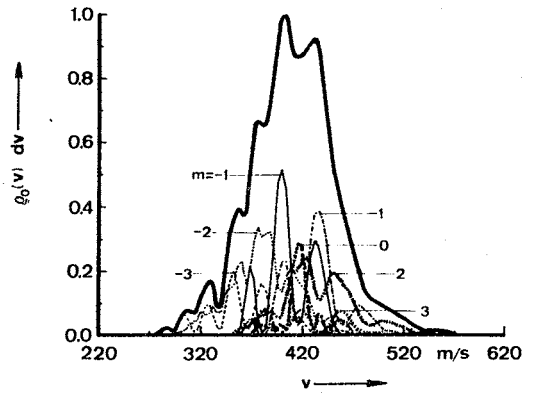


Fig. 10 Calculated velocity distribution $\varrho_0(v)$ dv when the atoms pass a magnetic field as in Fig. 8 with a specific setting of the currents. The thick solid curve is the sum of the individual contributions of atoms having left the polarizer magnet with the indicated quantum number m.

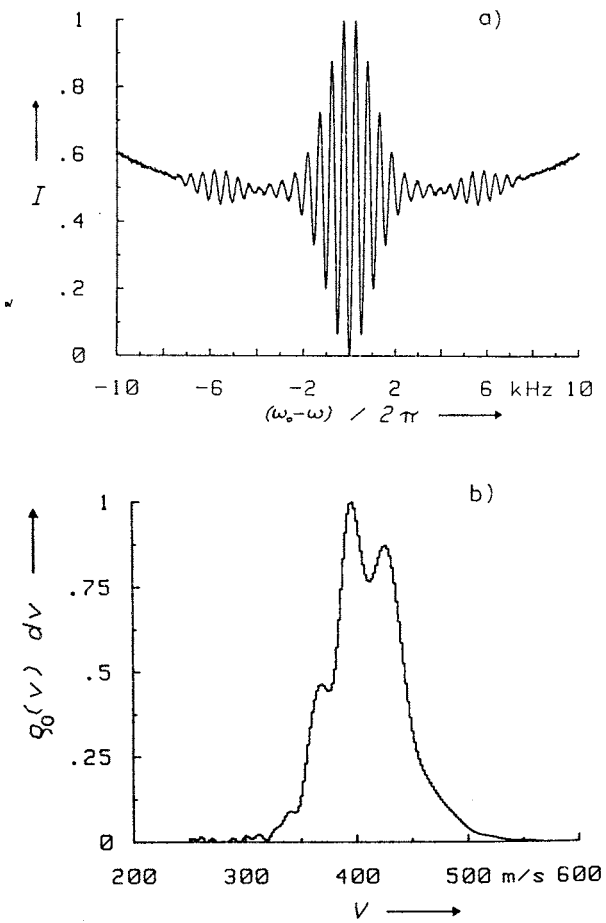


Fig. 11 a) Ramsey pattern of the clock transition in the CSX when the atoms pass a field as in Fig. 8. I: normalized beam signal.
b) Velocity distribution $\varrho_0(v)$ dv evaluated from the Ramsey pattern in Fig. 11a.

Result 2: If a fixed velocity v is considered, then the probability $|T(1/2 \rightarrow -1/2)|^2$ is a perodic function of the setting of the current flowing through coil 2: The period which appears reflects the increase in the integral field in between the two zeros so that the number of precessions increases by 1.

Result 3: As can be forseen from result 1, the shape of the velocity distributions changes considerably. In Fig. 10 a calculated distribution $\varrho_0(v)dv$ is shown. As in Fig. 6 the initial quantum state m of the atoms, which after the passage through the inhomogeneous fields are found in the state m=0, can be identified.

Result 4: At the same field setting which leads to the result shown in Fig. 9, the Ramsey pattern of the clock transition is as given in Fig. 11a). In fact, this was the first experimental finding obtained during the studies made with the CSX and has been briefly explained in /10/. If a Fourier analysis of a Ramsey pattern of this shape is made, the result is a velocity distribution $\varrho_0(v)dv$ as depicted in Fig. 11b) with a striking similarity to the one shown above.

5. Discussion and Outlook

If a state-selected atomic beam in a cesium clock passes inhomogeneous fields, then Majorana transitions are induced among the Zeeman levels. The change in the population of Zeeman levels can be calculated using either analytical or numerical methods. In both cases, however, the shape of the fields has to be known sufficiently well. This condition has been fulfilled in the experiments done with the CSX of the PTB. The shape of the fields referred to as "double rotation of the quantization axis" may seem somewhat artificial, but effects like those reported in section 4 have also been observed elsewhere /11, 12/.

It should be noted that the effects are as obvious as pointed out only for the following reason: The transition probability $|T(1/2 \rightarrow -1/2)|^2$ is a periodic function related to the number of the spin precessions in between the two zeros and in the experiments this number was of the order of 10 - 20. So a 5 % change in velocity or field is needed for a substantial change in the observables, 15 % being the relative width of the velocity distributions. If the number was 100 or more, the effects would be smeared out almost completely.

Assuming a change in the velocity distributions as described above, one may calculate the change of the frequency corrections to be applied during clock operation due to the second-order Doppler effect and the cavity phase difference /13/: The observed frequency shifts are larger by two orders of magnitude /7/. It is suggested /6/ that the Zeeman coherence created by the passage through the inhomogeneous fields is coupled to the coherence between the (4,0) and (3,0) states by the transversal rf field in the cavity which induce $\Delta F=1$, $\Delta m=\pm 1$ transitions. Some of the calculations presented in section 4.1 will be used again: In the numerical calculation full information can be obtained on the complex coefficients $C_+(t)$ and $C_-(t)$. The complete set of Eulerian angles needed in equation (4) can thus be obtained.

6 Bibliography

/1/ J. Vanier, C. Audoin, The Quantum Physics of Atomic Frequency Standards, Adam Hilger, Bristol and Philadelphia, 1989

/2/ E. Majorana, Nuovo Cimento 9 (1932), 43-50.

/3/ G. Becker, IEEE Trans.Instr.Meas. IM-27 (1978), 319-325.

/4/ S. Urabe et al., IEEE Trans.Instr.Meas. IM-29 (1980), 304-310.

/5/ Wang Yi-qiu, Acta Metrologica Sinica 4 (1983), 86-95.

/6/ R. Schröder, A. Bauch, see proceedings of this conference.

/7/ A. Bauch, PhD-thesis, published as PTB-Bericht, PTB-Opt-20, (1986).

/8/ F. Bloch, I.I. Rabi, Rev.Mod.Phys. 45, (1945), 237-244.

/9/ W. Schröder, G. Baum, Journal of Physics E 16 C(1983), 52-56.

/10/ A. Bauch, K. Dorenwendt, T. Heindorff, IEEE Trans.Instr.Meas. IM-34 (1985), 359-360.

/11/ Y. Nakadan, Y. Yoda, IEEE Trans.Instr. Meas. IM-34 (1985), 133-135.

/12/ J.McA. Steele, NPL, Teddington, private communication.

/13/ D.W. Allan et al., Proc. 31st Ann.Symp. on Frequ.Contr. (1977), p. 555-561.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 44

The Expansion of Optical Locking Range of a Spectral-Narrowed Laser Diode System using a Current Feedback Technique

T. Maeno, H. Saito, J. Umezu, Y. Ohta, R. Hayashi and P. Thomann*

Communications Research Laboratory, Koganei, Tokyo, 184, Japan
*Observatoire de Neauchatel, Switzerland

Abstract

Laser diodes are used in an optically pumped cesium frequency standard system. An optical feedback technique utilizing a Fabry-Perot cavity is used to stabilize frequency and reduce spectral linewidth. Linewidth of the spectral-narrowed laser diodes is reduced to 1/200 times that of the free-running condition. The signal to noise ratio of optical detection is improved for the linewidth narrowing. However, the long-term stability of the optical locking is not sufficient for a practical atomic clock. Thus, a new technique utilizing current feedback is developed to expand the optical locking range. The locking range is improved seven times wider than that of nonfeedback systems.

Introduction

While the natural linewidth of absorption of a cesium atom is approximately 5 MHz, the spectral linewidth of a laser diode light is 20-30 MHz. Therefore, the probability of optically pumping cesium atoms can be increased when the spectral linewidth of a laser diode light is reduced to less than the natural linewidth of 5 MHz. Thus, frequency-stabilized and spectral-narrowed laser diode systems were developed for optical pumping to improve the stability of the cesium atomic clock. The frequency stabilization and spectral narrowing were carried out using an optical feedback technique. The signal to noise ratio of optical detection was improved 3.5 times larger than that of non-narrowing systems, though the long-term stability of optical feedback was poor for a practical atomic clock. A current feedback technique was applied to the laser system to expand the locking range and time of optical feedback.

A Laser Diode System

Frequency stabilization and reduction of spectral linewidth were carried out by utilizing the optical feedback from a Fabry-Perot cavity [1][2][3][4][5]. A schematic diagram of the system is shown in Fig.1. A confocal Fabry-Perot cavity is used in this system.

The maximum optical output power of the laser diode is approximately 10 mW. The light wavelength is 852 nm (350 THz). The wavelength changes with the temperature and current drift at the rate of 0.05 nm/K (-20 GHz/K) and 0.01 nm/mA (-3.4 GHz/mA), respectively. The temperature of the laser diode then stabilizes at 300-310 K, and its fluctuation decreases to less than 0.01 K. The drift value of the constant current source

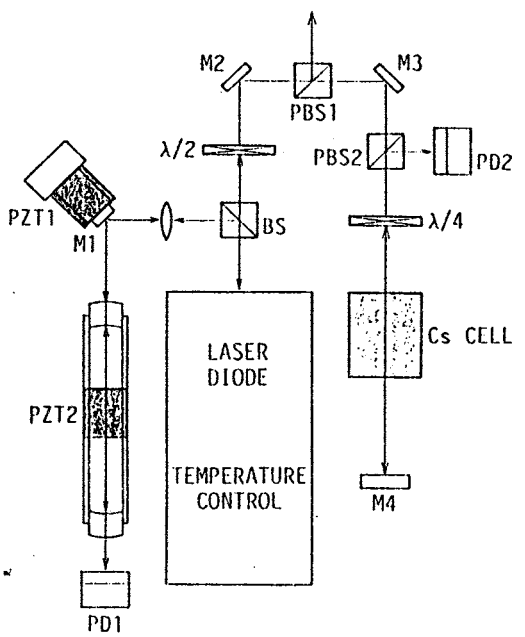


Fig.1 Schematic diagram of the spectral-narrowed laser diode system. Optical output power is 10 mW. Laser wavelength is 852 nm (350 THz). Laser linewidth is less than 100 kHz.

is less than 10 μA at 100 mA. The current noise is 1.8 $nA_{rms}/\ Hz^{1/2}$ at 10 Hz.

The free-spectral range and finesse of the Fabry-Perot cavity are 1 GHz and 140, respectively. Frequency selectivity of the cavity forces the frequency of the laser diode to the cavity resonance frequency. In other words, the optical output frequency of the laser diode system can be tuned by varying the cavity length.

Saturated absorption lines of cesium atoms were used as frequency references. As shown in Fig.1, a part of the laser beam passes a cesium-vapor cell and returns to a photodetector, PD2. The cavity length is controlled by an electrical signal of PD2 to stabilize the laser frequency. PZT1 and PD1 are also used to optimize the optical feedback phase.

The laser linewidth was measured by beat signal between two laser systems. The spectrums of optical beat are shown in Fig.2. The broader spectrum shows the beat of the spectral-narrowed laser system and the non-narrowed one. The narrow spectrum shows the beat of two spectral-narrowed laser systems. A linewidth wider than 20 MHz was reduced to less than 100 kHz by optical feedback.

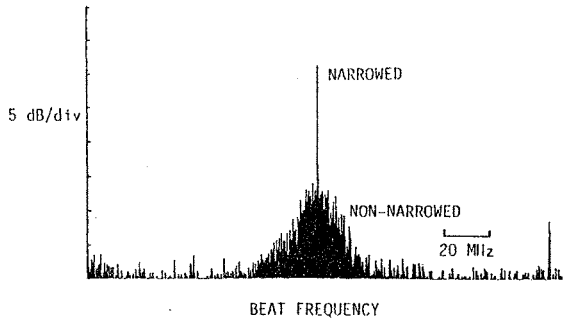


Fig.2 Spectrums of optical beat between two laser diode systems. The beat frequency of two lasers is detected by an avalanche photo diode and analyzed by a spectrum analyzer.

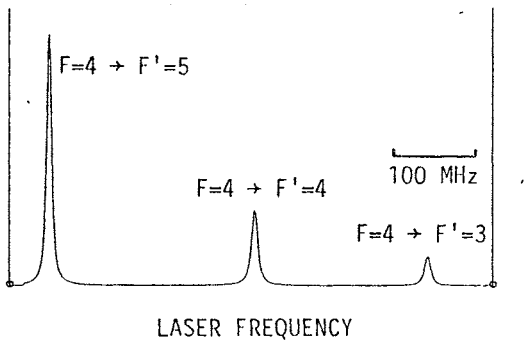


Fig.4 Fluorescence spectroscopy of a cesium beam measured in a vacuum chamber. Three fluorescence lines with high resolution can be observed.

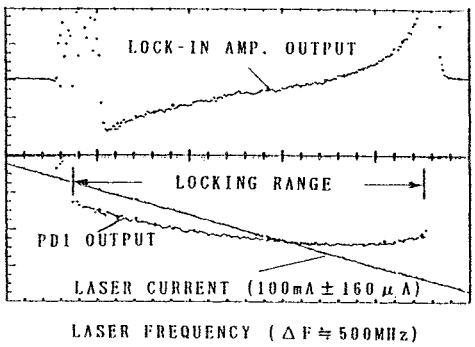


Fig.3 Relationship between the laser current and passed light intensity of a Fabry-Perot cavity. The upper figure shows the derivative of the light intensity.

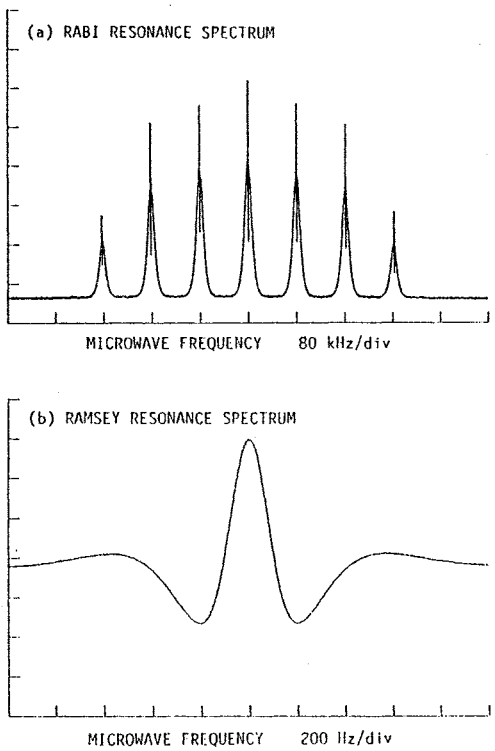


Fig.5 (a) The Rabi and (b) Ramsey resonance spectrums. These spectrums are highly symmetrical. The Ramsey spectral width at half amplitude was approximately 170 Hz. Compared with nonspectral-narrowed laser detection, S/N is improved by a factor of 3.5.

The Current Feedback Technique

The optical feedback is easily unlocked, because the laser frequency is changeable and its drift is larger than the optical locking range of the laser system. The locking range is expanded by increasing the optical feedback value from a Fabry-Perot cavity. However, the large feedback causes multimode oscillation and mode hopping. Therefore, in our system, the optimum feedback value was 10% of the total laser power. The typical locking range was 300 MHz, but the range was not sufficient for long-term operation. The typical locking time was three hours.

The current feedback technique was developed to improve the long-term stability. Output of PD1 is utilized as a current feedback signal. The light passing through the Fabry-Perot cavity is detected by PD1. The relationship between the laser current and the output of PD1 is shown in Fig.3. The upper figure shows the derivative of PD1 output measured by a lock-in amplifier. The current was swept within 100 mA±160 μA. By increasing the current, PD1 output increased and then decreased. The optimum current is indicated by the maximum point of PD1 output. The laser current is kept on the optimum value using the derivative of PD1 output. After appling the current feedback technique, the locking range and time are expanded to 2 GHz and 100 hours, respectively.

Fluorescence and Ramsey resonance spectrum

The fluorescence spectrum of a cesium beam measured in a vacuum chamber is shown in Fig.4. Three lines can be observed in this figure. The left one shows a transition frequency between F=4 and F'=5, where F indicates the hyperfine components of the ground state $6^2S_{1/2}$, and F' indicates the hyperfine components of the excited state $6^2P_{3/2}$.[1][6] The middle one and the right one show the transition frequencies of F=4-F'=4 and F=4-F'=3, respectively.

The Rabi and Ramsey resonance spectrums observed by the optically pumped cesium frequency standard system of CRL are shown in Figs.5 (a) and (b). These spectrums are highly symmetrical. The Ramsey spectral width at half amplitude of peak-to-valley was approximately 170 Hz. The S/N of optical detection was measured using this frequency standard system. The signal levels represent the difference between the detected signal level with the microwave radiation tuned to the center of the Ramsey resonance and well outside the hyperfine spectrum. The noise was measured by a FFT analyzer at the peak of the Ramsey resonance. The S/N was 1400 $Hz^{-1/2}$ at 10 Hz. Compared with nonspectral-narrowed laser detection, the S/N is improved by a factor of 3.5.

The Allan variance $\sigma_y^2(\tau)$ of the frequency standard system is calculated roughly as follows:

$$\sigma_y(\tau) = \frac{W}{\sqrt{2}(S/N)N_0}\tau^{-1/2} = 9\times10^{-12}\tau^{-1/2} \quad (1)$$

where, W is the linewidth of Ramsey resonance, 170 Hz, and N_0 is 9.2 GHz.

Conclusions

Spectral-narrowed laser diode systems were developed for an optically pumped cesium frequency standard system. The spectral linewidths of these laser systems were reduced to 100 kHz utilizing an optical feedback technique. S/N of optical detection was improved by a factor of 3.5 for spectral narrowing.

The current feedback technique was applied to a laser system to improve long-term stability. The locking range of 300 MHz was expanded to 2 GHz. Typical locking time is 100 hours at present. We will determine the frequency accuracy of our system soon.

Acknowledgments

The authors are grateful to Mr. Kuniaki Uchida, director, and Dr. Koji Nakagiri of Standards and Measurement division at CRL, for their extensive advice.

References

[1] S. Ohshima, Y. Nakadan, and Y. Koga, "Development of an Optically Pumped Cs frequency Standard at NRLM", IEEE Trans. Instrumentation and Measurement, Vol.37, No.3, pp.409-413, 1988

[2] Ph. Laurent, A. Clairon, and Ch. Breant, "Frequency Noise Analysis of Optically Self-locked Diode lasers", IEEE Journal Quantum Elec., Vol.25, No.6, pp.1131-1142, 1989

[3] B. Dahmani, L. Hollberg, and R. Drullinger, "Frequency Stabilization of Semiconductor Lasers by Resonant Optical Feedback", Optics Letters, Vol.12, No.11, pp.876-878, 1987

[4] L. Hollberg and M. Ohtsu, "Modulatable Narrow-linewidth Semiconductor Lasers", Appl. Phy. Lett., Vol.53, No.11, pp.944-946, 1988

[5] H. Li and N. B. Abraham, "Analysis of the Noise Spectra of a Laser Diode with Optical Feedback from a High-Finesse Resonator", IEEE J. Quantum Electron., Vol. 25, No. 8, pp.1782-1793, 1989

[6] J. L. Picque', "Hyperfine Optical Pumping of a Cesium Atomic Beam, and Applications", Metrologia, Vol. 13, pp.115-119, 1977

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 45

A NEW KIND OF HYDROGEN-MASER TIME AND FREQUENCY STANDARD FOR CHINESE VLBI NETWORK

Z.C.Zhai, H.X.Huang, C,F,Lin, J.W.He and W.Q.Zhang

Shanghai Observatory, Academia Sinica, Shanghai 200030

Abstract

A new kind of hydrogen-maser time and frequency Standard was developed at shanghai Observatory for very long baseline interferometry (VLBI) experiments.And five,so far,this kind of hgdrogen masers were made at Shanghai Observatory for chinese VLBI network. These standards are readily transportable, rugged and completely integrated frequency standards. The entire system is engineered as a unit, and all electronics were mechanically and electrically integrated with the physical portion into the overall design. This paper describes their design features and structures as well as the measured results of their performance. Its applications and data in VLBI experiment are also included.

INTRODUCTION

The hydrogen maser is generally regarded as the ideal time and frequency standard for VLBI.It has a unique combination of superior short-term and long-term stability.Further,the maser is a primary standard with unsurpassed resettability of frequency.

In the past, the size, weight, and general awkwardness of the equipment have limited the use of the maser to fixed installations. The masers which have been uesd for several years in Shanghai VLBI station are that kind of hydrogen masers,which were made earlier by Shanghai Observatory[1-3]

A hydrogen maser described in this paper is a new kind of maser, designed to be easily transportable and rugged enough to withstand the normal rigors of being moved from one station to another. In the new desigh,we absorbed many design ideas of masers developed by Dr.R.F.C.Vessot at the SAO.[4]

DESIGN FEATURE OF THE NEW MASER

The desigh philosophy for new maser developed from several basic aspects.First, the maser design emphasizes good intrinsically stability with respect to both mechanical and thermal perturbations.It is enclosed in a very carefully controlled thermal and mechanical environment to minimize maser detuning by variations in ambient conditions, and thus to permit stable maser operation over long periods of time without requiring autotuning or other adjustments.

Second, the new maser is a compact, relatively light-weight, rugged, and easily transportable standard. The entire standard was to be silf-contained in a single standard-sized rack cabinet. All electronics including the phase lock receiver were to be mechanically and electrically integrated into the overall design. It is a relatively small instrument weighing 250kg and requiring a source of only 24 volts DC at 200 watts to operate. The third feature is that there is a circulating air system in the rack cabinet. The system, which controis the temperature of the pump, the dissociator, and the upper maser electronics, is an entirely independent system with selt-contained sensors and electronics. In addition, the new maser incorpoorates an extensive monitoring capability. Four front panel mounted meters, each with an eight-position selector switch are located on the monitor panel to provide quick-look monitoring of 32 fuctions. These include all main power-supply voltages, all heater voltages, hydrogen dissociator operating conditions, receiver/synthesizer signal level, and phase-lock control voltages.The monitor panel also includes a small four-digit counter and LED readout for displaying the output frequency of the synthesizer.

Fig. 1 shows the photograph of the complete maser system.

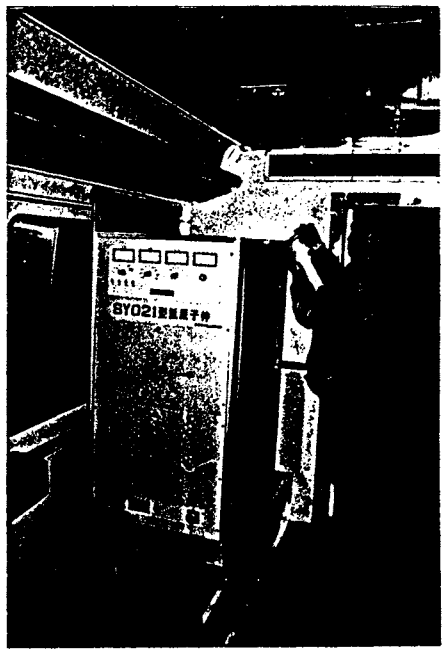


Fig 1 The New Hydrogen Maser Loaded on Vehhicle to be moved to the station

NEW MASER OSCILLATOR AND CLOCK SYSTEM

The design features of the Shanghai Observatory new hydrogen maser were described in detail in earlier paper[5]. The cavity resonator is an extremely rugged CERVIT structure. Longitudinal stresses produced by the thermal expansion coefficient of the resonator hold-down can are absorbed by a large Bellville spring washer. Thermally induced radial stress in the resonator base is relieved by a 40-glass balls quasi-kinematic mount. Barometrically induced stresses on the vacuum tank are isolated from the cavity resonator by a double-base structure. It should be noted that the design philosophy of the maser emphasizes good long-term frequency stability without the necessity for continuous automatic tuning.The tuning

procedure is usually required only when the maser is moved to a new location or before VLBI observations.

The entire maser clock system is shown in block diagram in Figure 2. A 100MHz crystal oscillator

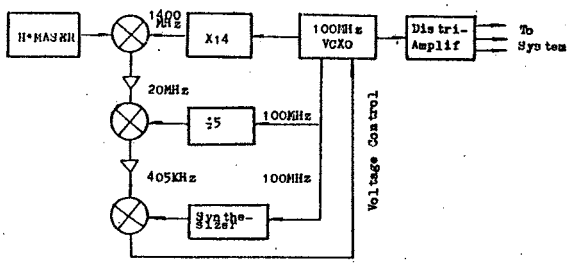


Fig 2 Maser clock system block diagram

is phase locked to the maser signal to provide standard frequencies at levels useful to the clock system. A simple dual-conversion receiver is used to generate the phase locking signals;a synthesizer is also used to set the time scale. 1MHz and 5MHz signals, derived from internal dividers in the crystal oscillator,is used to drive the VLBI timing system.

THE PERFORMANCES OF NEW MASER

A comprehensive series of environmental and short-term stability tests on the new maser has been evaluated before the new maser was moved to the VLBI station. Figure 3 shows the short-term frequency

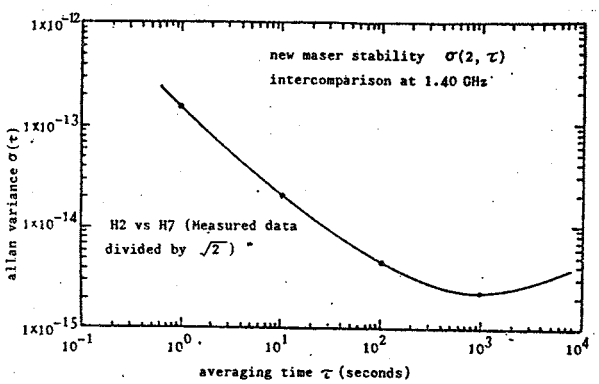


Fig. 3 New maser stability data

Stability of the new maser. It can be seen from the Fig. 3 the stability data keeps E-15 levels for the periods of time beyond 100 seconds.

In the environmental tests, the output frequency was carefully monitored while one of the environmental conditions was varied. The results of each of these environmental tests are itemized below.

1. Output frequency vs temperature tests-The new maser was placed individually in the test chamber and the chamber temperature was cycled between 22°C to 31°C,the rosultant variation in output frequency was plotted.

2. Barometric pressure tests-No output frequency variations were observed as the maser was individually subjected to barometric pressures of 25 mmHg above and below ambient pressure with a two-hour dwell at each extreme.

3.Magnetic field Sensitivity tests-It was measured by placing the maser under test within a 2.4-meter diameter set of Helmhoitz coils and varying the current through the coils. A about 0.4 gauss axial field was used in the tests.

A summary of the environmental sensitivities of the new maser is presented as follows[6]

Magnetic Field	(±0.4G)	1.1E-12/G
Barometric Pressure	(±25mmHg)	< 1.0E-14
Temperature	(22°C-31°C)	
slow coefficient		2.9E-14/°C
fast coefficient		1.2E-14/°C

APPLICATION

1. Tuning the hydrogen maser

The oscillation frequency of a hydrogen maser is "pulled" by the maser cavity resonantor. The amount of pulling is a function of cavity detuning and the line width of the atomic resonance.

The cavity is tuned by modulating the flux of atoms entering the storage bulb.It can be shown that the spin exchange contribution to the line width is directly proportional to the flux, and when the cavity is at the tuning point f_o , then f is independent of line width and therefore also independent of the total hydrogen flux.

The maser can be tuned by using another hydrogen maser as a reference or a frequency reference less stable than that of the maser itself. At Shanghai VLBI station, hydrogen maser is tuned by using a microcomputer system. The end results for tuning point and line Q are displayed on the screen.

2. VLBI data

During VLBI experiments, hydrogen maser performances are monitored by a microcomputer real-time measuring system.The stability data is shown not only by Allan Variance, but also in a form that might be more useful for VLBI experimenters. Fig. 4 shows phase

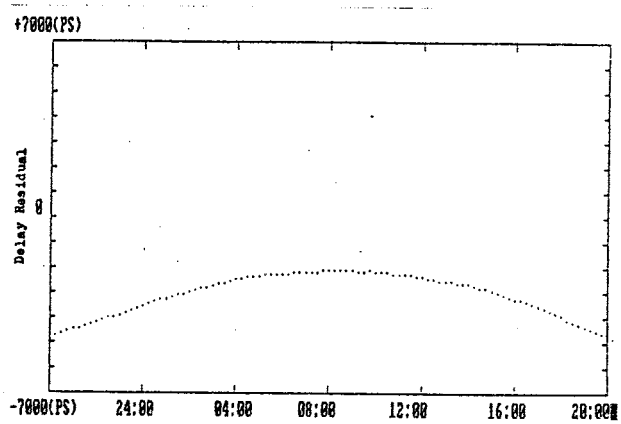


Fig. 4 VLBI phase Comparison With Frequency offset Removed

difference in seconds with only a frequency offset term removed from the data. Notice that the data has the quadratic behavior associated with uniform frequency drift. Fig. 5 shows the same data with a uniform frequency drift term also removed.

CONCLUSION

For VLBI observations, the hydrogen maser is the suitable satisfactory frequency standard. Shanghai Observatory new hydrogen maser has demonstrated that it can be used routinely and reliably for VLBI. Doubtlessly, the features of ease transportation and

convenient to operate make the maser to be also a ideal time and frequency standard for other fields.

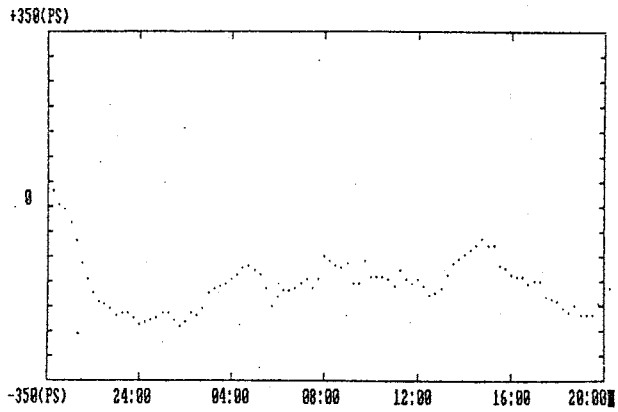


Fig. 5 VLBI Phase Comparison With Frequency Drift Removed

REFERENCES

[1]. Z.C.Zhai, JIETE, 27. 11, 1981;
[2]. Z.C.Zhai etc, proc of 38th annual symposium on Frequency control, 1984;
[3]. Z.C.Zhai etc, proc of 18th PTTI meeting, 1986;
[4]. R.F.C.Vessot, H.H.Levine, E.H.Hattison etc, proc. of 8th PTTI meeting, 1976;
[5]. Z.C.Zhai, H.X.Huang etc, proc, of 41st Annual symposium on Frequency control, 1987;
[6]. Z.C.Zhai, C.F.Lin etc, "Electromagnetic metrology", International Academic Publishers, 1989.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 46

TF-4010A-TFL's COMPACT RUBIDIUM FREQUENCY STANDARD

A. Stern and E. Eltsufin

T.F.L. Time & Frequency Ltd.
P.O.B 1792, Holon, 58117, Israel

ABSTRACT

We report the operation of a recently developed, extremely compact (500c.c.), Rubidium Frequency Standard.
This paper describes the unit which is comprised of an extremely small Physics-Package (87c.c.), a digital LSI synthesizer, hybridized circuitry and a miniature OCXO.
The Physics-Package design includes a simple low-noise lamp driver circuit, a special "C" field design and a special microwave cavity operating in a pseudo TE-101 mode small compact.
This unit exhibits excellent performance. It has an outstanding short-term and long-term stabilities (7xE-12/√tau and 1xE-11/month) and operates over a wide temperature range with a small frequency offset.

INTRODUCTION

In the last decade we have seen a dramatic reduction in size and price of all kinds of precision oscillators such as TCXO's, OCXO's and Rubidium Frequency Standards. TFL has put forward the goal to design a new Rubidium-Frequency-Standard in order to achieve a further substantial reduction in size and price and at the same time to maintain or even to improve the performance.

The R&D effort was placed on both the redesigning of the Physics-Package, where novel ideas were introduced, and on the redesigning and the miniaturization of the electronics circuitry. The miniaturization has enabled us to concentrate all the electronic circuits into only two printed circuits boards. In order to preserve the low phase-noise and short-term-stability we have used a special low price Ovenized-Controlled-Crystal-Oscillator as will be described later.
The final unit, shown in figure 1, has achieved all these goals and exhibits the same performance or better compared to other Rubidium Frequency Standards which are available in the market today.

DESIGN DESCRIPTION

The general electronic design is of a conventional one. The functional block diagram is shown in Fig 2. The 10MHz from the Ovenized-Controlled-Crystal-Oscillator (OCXO) is multiplied by 9 to obtain 90MHz which, together with a synthesized 5.3125MHz, are injected into a Step-Recovery-Diode (SRD) multiplier. The SRD is located on top of the microwave cavity inside the Physics-Package. It generates the nominal 6.834687500GHz, clock transition frequency, from the 76th harmonics of the 90MHz and its 5.3125MHz sideband. The clock transition frequency is frequency modulated at a relatively fast 444Hz, and is being radiated onto the Rb87 atoms in the resonance cell inside the cavity. A Rb87 lamp emits light which is being filtered by a Rb85 filter cell and transmitted through a Rb87 resonance cell to be detected by a photo-diode. The output of the photo-diode is amplified, phase detected, integrated and voltage-controls the OCXO, thus closing the frequency-lock-loop.

The unit is comprised of only 4 modules: the Control-Board, the Power-Supply-Board, the Physics-Package and the OCXO. These modules are packed into a 500cc box of dimensions 93x82.6x65 mm. A general package design is shown in Fig 3.

We have been successful in condensing all the electronics circuitry into only 2 boards. The Control-Board (78x88mm) includes the multiplier, the synthesizer, the preamplifier, the frequency-lock-loop and the alarm circuit. The Power-Supply-Board (78x44mm) includes a DC/DC switching power supply, as well as a switching temperature control for the microwave cavity and the lamp house of the Physics-Package. The Physics-Package (see Fig 4) is extremely small, only 87cc in volume, and comprises a lamp subassembly, a cavity subassembly and a two-fold magnetic shield made out of mu-metal. The oscillator is a miniature Oven-Control-Crystal-Oscillator with an AT-cut, 3rd-overtone, crystal.

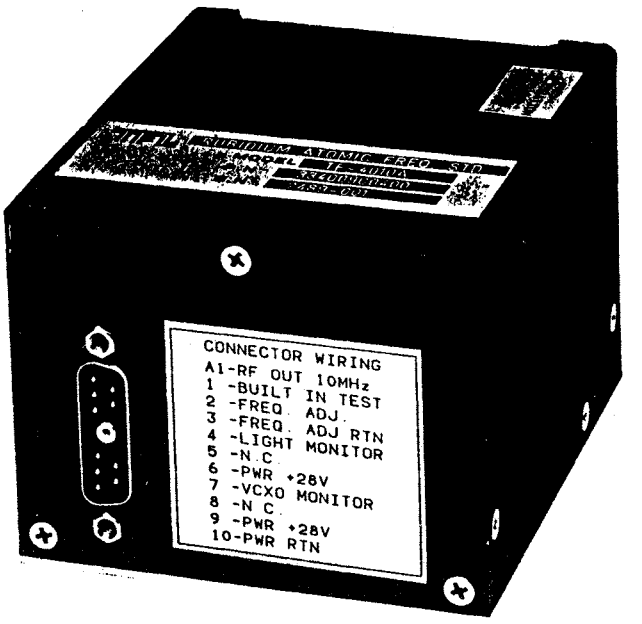


Fig 1: Rubidium Frequency Standard model TF-4010A.

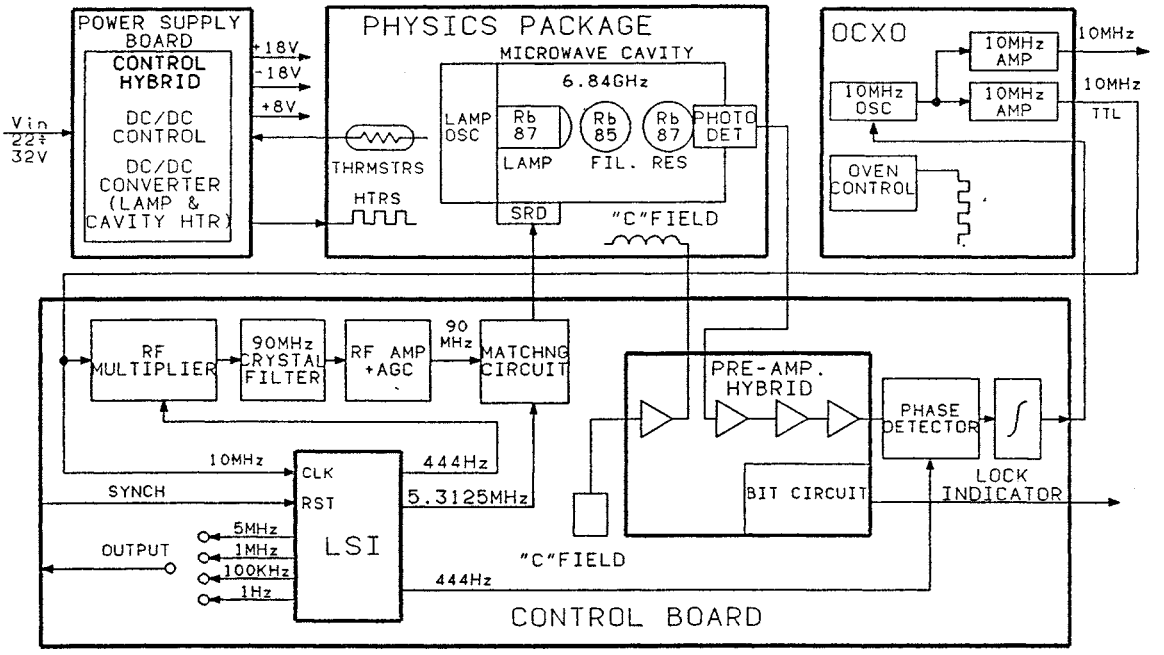


Fig 2: A functional block diagram TF-4010A.

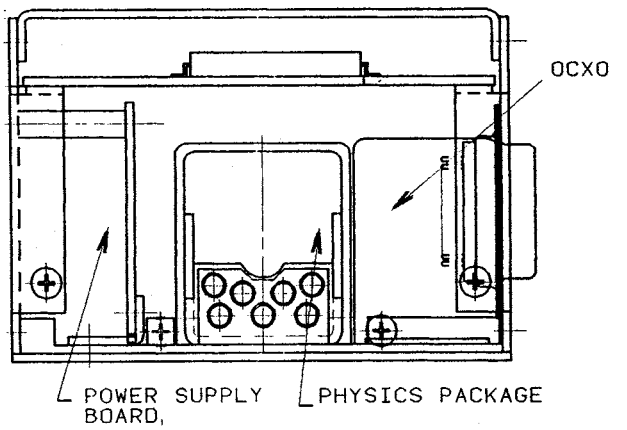


Fig 3: A general package design.

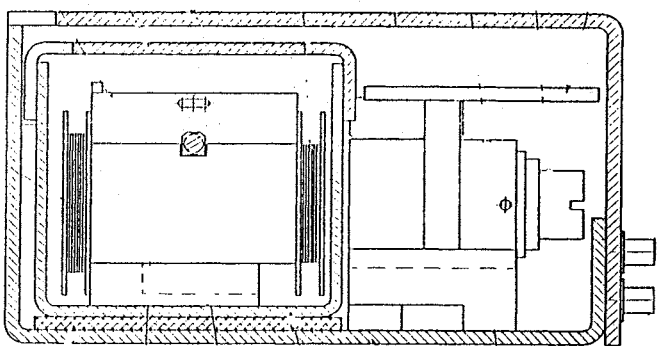


Fig 4: The Physics Package design.

Henceforth we describe the key elements of the design that enable the reduction in size and price and at the same time improve the performance.

THE LAMP OSCILLATOR

We have designed a very simple lamp oscillator which comprises only 15 components and forms an integral part of the lamp house. It is located in a near proximity to the RF excitation coils. It fulfills the purpose of not increasing the noise above the theoretical shot-noise of the light.

THE MICROWAVE CAVITY

We have invented a special cylindrical-rectangular microwave cavity, partially loaded with dielectric which is operating in a pseudo TE-101 mode. Main advantages are: very small volume (about 4.3cc) and a high filling factor for the vapor cells. The design is a modification of a previous rectangular cavity design [1]. A detailed description and analysis of the cavity will be given elsewhere [2].

THE SRD MULTIPLIER CIRCUIT

An SRD multiplier circuit in a chip and wire form is located on top of the microwave cavity and forms an integral part of it. This circuit allows the receptions of the RF input signals to the SRD with no reflection and high efficiency.

THE LSI SYNTHESIZER

A great deal of the miniaturization is achieved due to the development of a standard-cell ASIC that synthesizes the various frequencies used in the frequency-lock-loop as well as additional output frequencies. The chip which is packed in a 16 dual-in-line package, receives a 10MHz input frequency. It generates 5.3125MHz and 444Hz for the frequency lock loop. In addition it provides TTL outputs of 5MHz, 1MHz, 0.1MHz and 1Hz (with synchronization input) that can be used as optional outputs to the unit.

HYBRIDS

A large volume of the circuitry is condensed into 2 hybrids. The photo-diode preamplifier as well as part of the lock loop are included in a hybrid which is located on the Control-Board. Most of the circuitry of the DC to DC switching power supply as well as the switching temperature controller are included in an hybrid which is located on the Power-Supply-Board.

THE OCXO AND THE LOOP

In order to maintain a low phase noise and also a good short-term-stability, we decided to adopt TFL's inexpensive, compact OCXO, model TF-65040A. It is currently being produced in very large quantities. This is a 10MHz, AT-cut, SMD device, packed in a miniature 36cc package.

Generally, the short-term-stability (as usually measured by the "Allan deviation") of a Rubidium Frequency Standard is related to the OCXO, for averaging times shorter than the loop-time-constant, and depends on the Rubidium atoms, for averaging times which are longer than the time constant. This partition is also being manifested in the frequency domain, where the phase-noise is dependent on the oscillator for frequencies (from carrier) higher than the loop-bandwidth and depends on the Rubidium resonator for frequencies which are lower than the loop-bandwidth. We have chosen to use a loop-time-constant around 5ms and loop-bandwidth of around 35Hz.

However, in order not to degrade the frequency and phase stability in regions that are still dependent on the Rubidium, the OCXO phase-noise at 444Hz from carrier should be kept below certain level (444Hz is our modulation frequency as described above). Therefore a special effort was made to achieve a low phase-noise around 444Hz and to improve the oscillator. Typical results that were achieved are -140dBc (at 1Hz bandwidth) around 444Hz.

	Specifications	Typical Measurments
Output	10MHz	
Frequency	5MHz,1MHz,0.1MHz	
	1Hz with synchronization	
"Allan"	2xE-11 at 1sec	8xE-12 at 1sec
Deviation	2xE-12 at 100sec	9xE-13 at 100sec
Phase-noise	-100dBc at 10Hz	-105dBc at 10Hz
(1Hz bw)	-125dBc at 100Hz	-135dBc at 100Hz
	-135dBc at 1KHz	-145dBc at 1KHz
Aging	4xE-11/month	1xE-11/month
Temperature	±3xE-10/	±1.5xE-10/
Stability	-10°C to +65°C	-10°C to +65°C
Warm-up time	5 min to 2xE-10	4 min to 2xE-10
Steady state power	<14W	<13W
Start up power	<28W	<26W
Loop time constant	5 to 10ms	
Size	93x82.6x65mm	

Table 1: Key specifications and typical results measured for first prototypes.

PERFORMANCE

Table 1 summarizes the key specifications and typical values as were measured for the first 5 prototypes.

Fig. 5 shows the frequency drift versus time (aging), Fig. 6 depicts the frequency offset and the power consumption during warm-up and Fig. 7 shows the frequency offset versus ambient temperature between -10°C to 60°C.

As seen from the table and graphs the exhibited performance is comparable or better than the performance usually stated for a higher priced Rubidium Frequency Standards existing in the market today. In particular we would like to emphasize the excellent short-term-stability of 7 to 8xE-12/√tau, the long-term-stability of about 1xE-11/month and the low power consumption at start-up and steady state. The good short-term-stability is achieved due to the low noise lamp driver, the low noise ovenized oscillator and the careful design of the multiplier and SRD, as explained above. The good long-term-stability is achieved by maintaining the same operating criteria, used in bigger higher priced units. e.g., a low light-intensity sensitivity due to the light-shift effect. Low power consumption is achieved due to a careful thermal design of the whole package.

CONCLUSION

We have described a new compact, inexpensive but still a high performance Rubidium-Frequency-Standard that has been developed at TFL. We believe that a further substantial reduction in price and size are still possible. A continuation of the program towards this goal is being undertaken at TFL.

REFERENCES

1. H.E. Williams, T.M. Kwon and T. McClelland, Proc. 37th An. Symp. Freq. Cont. 1983, page 12.
2. E. Eltsufin and A. Stern, to be published.

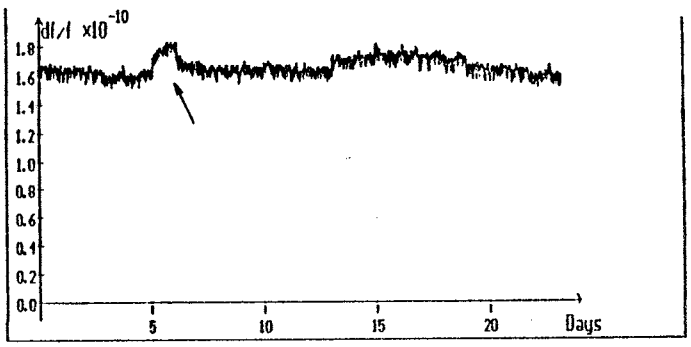
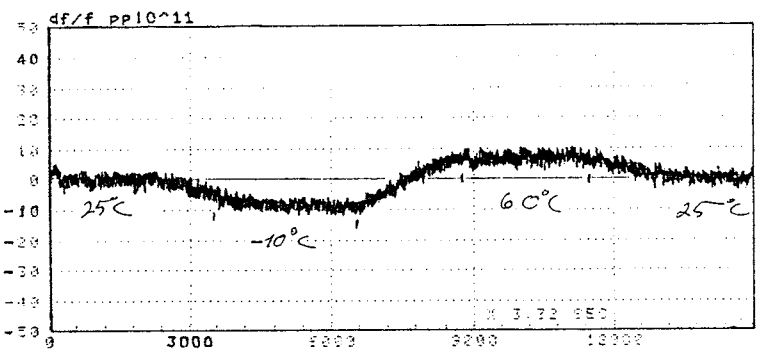


Fig 5:

Frequency drift versus time.
Ambient temperature was kept at 24±3°C, except for the marked region (air-condition failure during weekend).



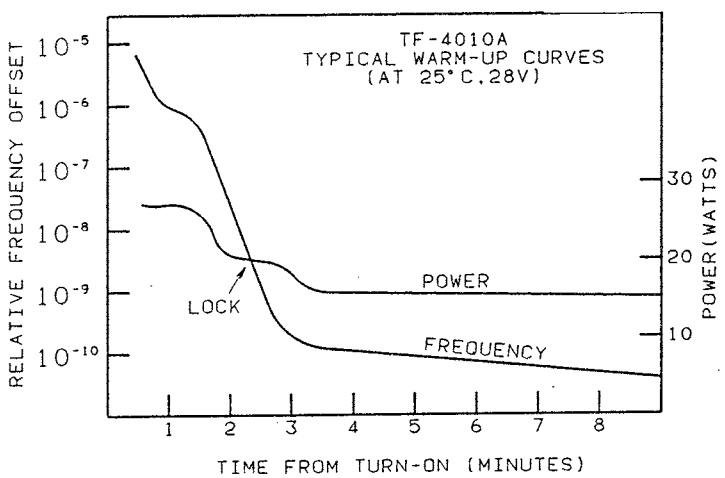


Fig 6: Frequency offset and power consumption during warm-up.

Fig 7:

Frequency offset versus ambient temperature.
Temperature cycle: 25°C, 25°C to -10°C, -10°C, -10°C to +60°C, +60°C, +60°C to -25°C and 25°C.
2 hours at each temperature and 2 hours at the transition periods.

Session POSTER Number 47

Computer controlled short term frequency stability measurement system with the application of covariance method

Barbara Gniewińska

Tele and Radio Reserch Institute
Warszawa, Poland

Abstract

Basing on covariance method a computerized system for short-term frequency stability measurements with the resolution of better than 10^{-13} for τ = 1s, has been developed. The covariance method makes possible the determination of short term frequency stability of a single oscillator.

I Operation Principle

Operation principle is illustred in fig 1

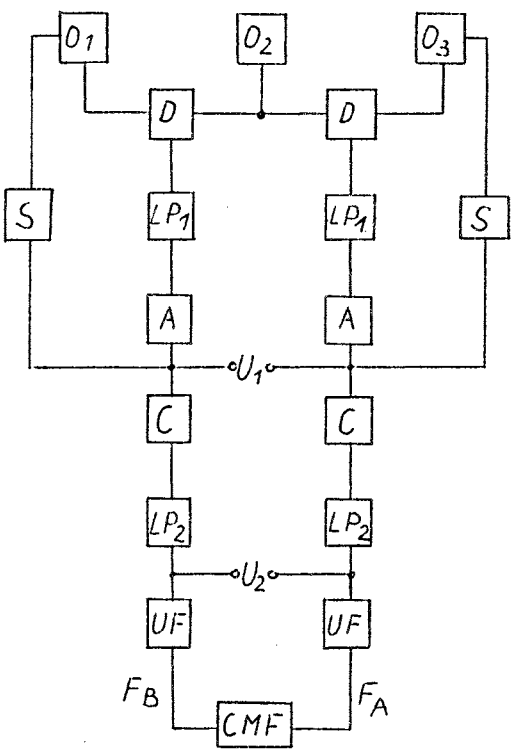


fig 1 Block diagram of short term frequency stability measuring system

where:

- O_2 - Oscillator under test
- O_1, O_3 - Reference oscillators
- D - Balanced phase detector with the slope μ [V/rad]
- LP_1 - Low-pass filter
- A - Low noise amplifier with amplification factor A
- S - Frequency control circuit of slope S [Hz/V]
- C - Correction circuit of $1 + \Omega/2\pi\mu SA$ transmittance
- LP_2 - Low-pass filter defining the limiting frequency of the system
- U/F - Converter of sensitivity P [Hz/V] and frequency F_o
- CMF - Computer-based frequency measuring equipment

According to the denotations

$$U_1(\Omega) = \frac{1}{S} \frac{1}{1 + \frac{\Omega}{2\pi\mu SA}} \frac{\Delta\varphi}{2\pi}$$

and at the transmittance

$$C = 1 + \frac{\Omega}{2\pi\mu SA} \; ; \; U_2 = \frac{1}{S} \frac{\Delta\varphi}{2\pi}$$

then

$$F_A = F_o + \frac{Pf_o}{S}(y_1 - y_2) \; ; \; F_B = F_o + \frac{Pf_o}{S}(y_2 - y_3)$$

where

$$y_{1,2,3} = \frac{\Delta f_{1,2,3}}{f_o}$$

The computer frequency meter designed as a card for IBM Computer PC/XT performing the frequency measurement F^A and F^B assures the:

- measurement resolution of 0,05 Hz
- zero death time between the following measurements
- simmultaneous measurements in both measurement channels

Introducing the covariance function in the form

$$\sigma(1,2)_A \cdot \sigma(2,3)_B =$$

$$\left\langle \left[\left(y^{01}_{k+1} - y^{02}_{k+1} \right) - \left(y^{01}_{k} - y^{02}_{k} \right) \right] \cdot \left[\left(y^{03}_{k+1} - y^{02}_{k+1} \right) - \left(y^{03}_{k} - y^{02}_{k} \right) \right] \right\rangle =$$

$$= \left\langle \left(F^{A}_{k+1} - F^{A}_{k} \right) \left(F^{B}_{k+1} - F^{B}_{k} \right) \right\rangle$$

$$\left\langle \left[\underbrace{\left(y^{01}_{k+1} - y^{01}_{k} \right)}_{\delta^{A}_{01}} - \underbrace{\left(y^{02}_{k+1} - y^{02}_{k} \right)}_{\delta^{A}_{02}} \right] \left[\underbrace{\left(y^{03}_{k+1} - y^{03}_{k} \right)}_{\delta^{B}_{03}} - \underbrace{\left(y^{02}_{k+1} - y^{02}_{k} \right)}_{\delta^{B}_{02}} \right] \right\rangle$$

$$= \delta^{A}_{01} \cdot \delta^{B}_{03} - \delta^{A}_{01} \cdot \delta^{B}_{02} - \delta^{A}_{02} \cdot \delta^{B}_{03} + \delta^{A}_{02} \cdot \delta^{B}_{02}$$

At the lack of correlation between the fluctuations of oscillators under test, all the components except

$$\delta^{A}_{02} \cdot \delta^{B}_{02} = \left\langle \left(y^{02}_{k+1} - y^{02}_{k} \right)^2 \right\rangle = 2\,\sigma^2_{y02}$$

are reducing to zero.
Similar, at the lack of correlations between the fluctuations of measurement channells, it can be proved, that the noise effect of the measurement circuit also reduces to zero.
In practice, due to the impossibility to reach the complete independance of the processes and the finite observation time, only a certain amount of circuit noise effect reduction and the increase of measurement resolution take place.
The measurement process is carried out automatically, according to the program introduced by the operator. There is a possibility for programing the following measurement parameters

- averaging time /from 10^{-4}s to 10^{2}s/
- number of samples /from 1 to 1000/
- number of repeated measurements sets /from 1 to 10/.

The measurements results are presented in the form of a diagram or a table on a monitor screen or printed.

II Measurement Results

In our developed system we achieved the measurement resolution improvement shown in fig 2.

In the system the Hewlett Packard phase detector type 10534 A has been used.
The main parameters of the remaining modules are given in table 1.

Table 1

Low noise amplifier	A=200 V/V noise 2nV/√Hz
U/F Converter	F=1MHz $\sigma_A = 10^{-5}$/s P=0,6 MHz/V
Frequency control circuit	S=0,1 Hz/V
Low pass filter LP_2	cut-off frequence f_h=2KHz

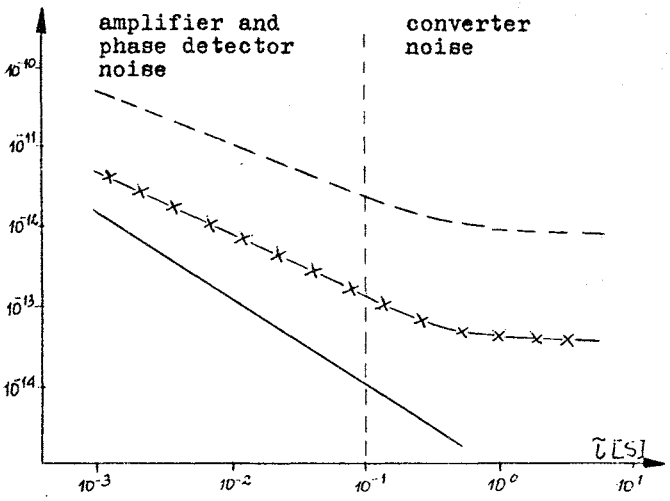


Fig 2. Measurement resolution

- - efect of noise in one channel system
-x- effect of noise in two channel system
--- effect of the F frequency measurement

The resulting measurement resolution as it was anticipated, depends considerably upon the number of samples.
As an example, in table 2, the measurements resolution of the system versus samples number N for $\tau = 10^{-3}$s and =1s are given.

Table 2

N	20	50	100	200	500
$\sigma_A \cdot 10^{-12}$ ($\tau = 10^{-3}$s)	10	9,5	5,5	4,8	4
$\sigma_A \cdot 10^{-14}$ (τ = 1s)	8,3	6	5,5	4	–

Comparison measurements of the OCXO-1C oscillators of ITR production carried out in the developed system are illustrated in fig 3

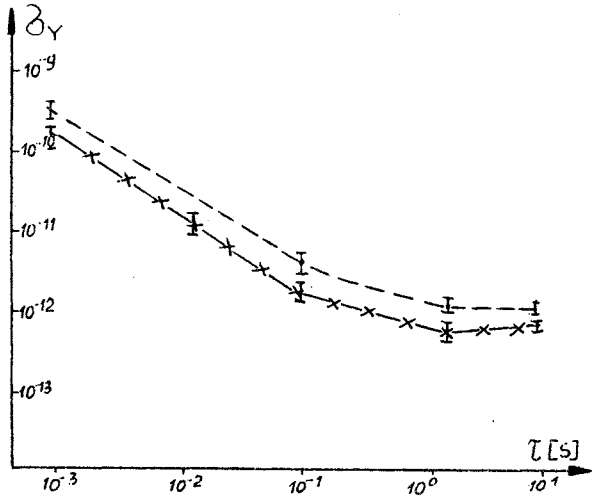


fig 3 Measurements results of 10 oscillators type OCXO-10P
- - one channel system
-x- two channel system

References

1. G.Fest, J.Groslambert, J.J.Gagnepain
Individual Characterization of an Oscillator by Meas of Cross Correlation or Cross-variance Method
IEE Transaction on Instrumentation and Measurement VOL IM No 3, 1983

2. D.Fest
Mesure du comportement a court terme des oscillateurs par une methode d´intercorellation. d´Onde Electrique No 6-1-1983

3. B.Gniewińska
Measurement System for Short-Term Frequency Stability
Publication of ITR - No Z-583, 1989

4. A.Kondarewicz
Computer Based Frequency Meter
Publication of ITR No Z-583, 1989

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 48

Systems of Convolution Equations, Deconvolution, and Gabor and Wavelet Signal Analysis

Stephen D. Casey
Department of Mathematics and Statistics
The American University
4400 Massachusetts Ave., N.W.
Washington, D.C. 20016, U.S.A.

Abstract

The linear, translation invariant model for a sensor is the convolution equation

$$s = \mu * f,$$

where f is the unknown signal, μ is the sensor impulse response function (or, more generally, impulse response distribution), and s is the output signal. In many sensing applications, the output s is an inadequate approximation of f, which motivates solving the convolution equation for f, i.e. deconvolving f from μ. If the function μ is time limited, then this deconvolution problem is ill-posed.

A powerful theory of solving such equations has been developed by Berenstein et al. It circumvents ill-posedness by using multiple sensors. If we overdetermine the signal f by using a system of convolution equations,

$$s_i = \mu_i * f,\ i = 1, \ldots, n,$$

the problem of solving for f is well-posed if the set of convolvers $\{\mu_i\}$ satisfies a condition which we call strongly coprime. In this case, there exist compactly supported distributions

$$\nu_i,\ i = 1, \ldots, n$$

such that

$$\delta = \nu_1 * \mu_1 + \ldots + \nu_n * \mu_n,$$

which in turn gives

$$f = \nu_1 * s_1 + \ldots + \nu_n * s_n.$$

We describe the strongly coprime condition and we give examples of sets of strongly coprime sensors and their deconvolutions. We then discuss several applications of the theory, and in particular, show how this method of solution may be used to gain a subpixel resolution of an image.

Finally, since the purpose of deconvolution is greater signal resolution, it is shown how data compression and analysis using the Gabor and wavelet transforms are implemented in coordination with the deconvolution methods described above. Explicit formulae for doing simultaneous signal recovery via deconvolution and data compression and analysis via the Gabor and wavelet transforms are derived.

1 Convolution Equations and Remote Sensing

Let s be a function which models a signal received by an active or passive remote sensor. We assume that this function has finite energy, i.e., s is square integrable. The linear, translation invariant model for the signal s is the convolution equation

$$s(\tau) = f * \mu(\tau) = \int_{\mathbf{R}^k} \mu(t) \cdot f(\tau - t) dt,$$

where μ is the sensor impulse response function (or, more generally, sensor impulse response distribution), and f is the data function or actual signal we want to recover. The equation $f * \mu$ is a weighted averaging of f with μ in which certain information about f is lost.

Convolution is an operation that has properties that make it useful in both theory and application. When we are working with integrable functions on $\mathbf{R}^k$, convolution is associative, commutative, and linear, that is

- (associativity) : $f * (g * h) = (f * g) * h$,
- (commutativity) : $f * g = g * f$, and
- (linearity) : $f * (\alpha \cdot g + \beta \cdot h) = \alpha \cdot (f * g) + \beta \cdot (f * h)$, where α and β are constants.

It is also a smoothing operation, a property which manifests itself in different ways. For example, if g is an infinitely smooth function, then $f * g$ is infinitely smooth, with

$$D^\alpha (f * g) = f * D^\alpha g.$$

If g is the characteristic function of a region, then $f * g$ is smoother than f by one derivation. And, finally, there is the Central Limit Theorem, which tells us that infinite convolutions of continuous functions which have both bounded variance and variance bounded away from zero yields an infinitely smooth function (which is, in fact, a linear translation of an unnormalized normal distribution, or Gaussian function).

There is, however, no function which is the identity under convolution, that is, there exists no g so that $f * g = f$. Mathematically, an identity can be constructed - but the resultant object is not a function, but rather a functional, the Dirac delta δ. The delta was developed within the framework of the theory of distributions. It acts on a continuous function as follows :

$$\langle \delta, f \rangle = \int_{\mathbf{R}^k} f(t) \cdot \delta(t) dt = f(0).$$

The delta is constructed as the limit of approximate identity functions, compactly supported smooth functions $h(t)$, $h(t) \geq 0$, with $\int_{\mathbf{R}^k} h(t) dt = 1$. Then,

$$\delta(t) = \lim_{\lambda \to \infty} \lambda^n h(\lambda t), \lambda \in \mathbf{R}.$$

(We define the support of a function on $\mathbf{R}^k$ is the smallest closed subset of $\mathbf{R}^k$ on which the function is nonzero, i.e. for $h : \mathbf{R}^k \longrightarrow \mathbf{R}^m$, $\text{support}(h) = \text{closure}(h^{-1}(\mathbf{R}^m - \{0\}))$, and we say that h is compactly supported if the support of h is a closed and bounded set. Note, from this definition, we can see why it is impossible for δ to be a function, because it has support at only a single point, yet its integral equals one.) The action of the delta on continuous functions gives it the property that

$$f * \delta(\tau) = \int_{\mathbf{R}^k} f(t) \cdot \delta(\tau - t) dt = f(\tau),$$

or, in other words, δ is the identity under convolution.

The problem of extracting information about f from $s = f * \mu$ is fundamental in signal and image processing. It can be addressed either partially or completely. In the situation in which some *a priori* knowledge or assumptions are available about the signal, techniques that may solve only a specific type of

problem or a specific part of the general problem may be applied. For example, if it is known that the signal has a certain signature and will be in certain frequency and amplitude ranges, sensors may be chosen so that a correlation with an appropriate filter and thresholding of this signal will recover a filtered and smoothed version of the function, even in a relatively noisy environment.

A much more difficult problem is the complete reconstruction, or deconvolution, of f from μ. This deconvolution may be thought of as the construction of the Dirac delta within the functional (or distributional) class of the sensors. Physical characteristics of all sensors cause modification and/or loss of signal information. However, the problem of signal recovery in this situation is tractable, because, for example, sensor impulse response can be modeled in a laboratory, and the linear model of the sensor as the convolution of the signal with this impulse response gives a very close approximation of the received signal.

A powerful theory of solving convolution equations for time limited sensors has been developed by Berenstein et al, and is discussed in section 2. It circumvents the "ill-posedness" inherent in these equations (a concept defined in the next section) by using multiple sensors. We present solutions of these types of problems by solving for "deconvolvers" that allow us to construct the δ using convolution equations. This in turn allows us to reconstruct the original signal f. We then discuss several applications of the theory, and in particular, show how this method of solution may be used to gain a subpixel resolution of an image.

Also, since the purpose of deconvolution is greater signal resolution, it is shown in section 3 how data compression and analysis using the Gabor and wavelet transforms are implemented in coordination with the deconvolution methods described above. Explicit formulae for doing simultaneous signal recovery via deconvolution and data compression and analysis via the Gabor and wavelet transforms are derived.

2 Systems of Convolution Equations and the Deconvolution of the Remotely Sensed Signals

In this section, we shall assume that we want to recover an signal f from the data s received from a remote sensor. The linear, translation invariant model for this remotely sensed signal is the convolution equation

$$s = f * \mu,$$

where μ is the sensor impulse response function (or, more generally, sensor impulse response distribution). We want to solve this equation for f, or deconvolve f from μ. The problem of solving this equation is said to be well-posed whenever :

- (existence) : there always is some f that will produce the measured s,
- (uniqueness) : there is at most one f that will produce the measured s, and
- (continuity) : small changes in s cause only small changes in f.

The physical limitations of all time-limited sensors make the problem of solving the convolution equation for f ill-posed (see the Appendix of [5]). In order to circumvent the ill-posedness of the problem, a multi-sensor approach is used to recover f.

The theory of deconvolution discussed in the following paragraphs has been developed by Berenstein et al (see [1] - [6]). These methods of solution are both linear (convolution with "deconvolvers") and realizable (the support of the deconvolvers being contained in the bounded support of the kernels of the convolution equations).

If a fixed signal f is overdetermined by using a system of convolution equations,

$$s_i = f * \mu_i,\ i = 1, \ldots, n,$$

then the problem of solving for f is well-posed if the set of convolvers $\{\mu_i\}$ satisfies a condition which is called *strongly coprime*. In this case, there exist compactly supported distributions

$$\nu_i,\ i = 1, \ldots, n$$

such that

$$\delta = \mu_1 * \nu_1 + \ldots + \mu_n * \nu_n,$$

where δ denotes the Dirac delta. This in turn gives

$$\begin{aligned} & s_1 * \nu_1 + \ldots + s_n * \nu_n \\ &= (f * \mu_1) * \nu_1 + \ldots + (f * \mu_n) * \nu_n \\ &= f * (\mu_1 * \nu_1) + \ldots + f * (\mu_n * \nu_n) \\ &= f * (\mu_1 * \nu_1 + \ldots + \mu_n * \nu_n) \\ &= f * \delta \\ &= f. \end{aligned}$$

The convolvers with which the theory works with are in the space of distributions of compact support, or, in the language of applications, time-limited sensors.

Let $\widehat{\mu}(\zeta)$ denote the Fourier-Laplace transform of $\mu(t)$, for $t \in \mathbf{R}^k$, $\zeta \in \mathbf{C}^k$. For notation, recall that if μ is an integrable function, then

$$\widehat{\mu}(\zeta) = \int_{\mathbf{R}^k} \mu(t) e^{-i(x \cdot \zeta)} dt,$$

gives the Fourier-Laplace transform of μ. The transform is linear. If $\omega = \Re\zeta$, then $\widehat{\mu}(\omega)$ is the Fourier transform of μ. If μ is compactly supported, $\widehat{\mu}$ is an analytic function for all $\zeta \in \mathbf{C}^k$. And, as is well known,

$$\widehat{(f * g)}(\zeta) = \widehat{f}(\zeta) \cdot \widehat{g}(\zeta).$$

Theorem ([1] - [5]) : For the compactly supported distributions

$$\mu_i,\ i = 1, \ldots, n$$

on $\mathbf{R}^k$, there exists compactly supported distributions

$$\nu_i,\ i = 1, \ldots, n$$

such that

$$\delta = \mu_1 * \nu_1 + \ldots + \mu_n * \nu_n$$

if and only if there exist positive constants A and B and a positive integer N such that

$$\left(\sum_{i=1}^{n} (\widehat{\mu}(\zeta))^2 \right)^{\frac{1}{2}} \geq A e^{-B|\Im\zeta|} \cdot (1 + |\zeta|)^{-N},\ \zeta \in \mathbf{C}^k.$$

Distributions (convolvers) μ_i, $i = 1, \ldots, n$ that satisfy the inequality in the theorem are referred to as *strongly coprime.* (A more general, but more technical, version of this theorem may be found in [1].) The strongly coprime condition guarantees not only that the transforms of the convolvers have no common zeroes, but also that these zeroes do not cluster too quickly as $|\zeta| \to \infty$. We note that the only way a single compactly supported convolver can be strongly coprime is for it to be a translate of the identity convolver; that is, for it to be the Dirac delta or a translation thereof (see the Appendix of [5]). This result is one way to state the general ill-posedness of single convolution equations under the constraints imposed by the conditions of the theorem.

Construction of the deconvolvers requires techniques from multi-variable harmonic and complex analysis. The transform of

$$\delta = \mu_1 * \nu_1 + \ldots + \mu_n * \nu_n,$$

is

$$\sum_{i=1}^{n} (\widehat{\mu}_i(\zeta)) \cdot (\widehat{\nu}_i(\zeta)) = 1.$$

The *Bezout problem* is to solve for transformed deconvolvers $\widehat{\nu}_i(\zeta)$, $i = 1, \ldots, n$, for $\zeta \in \mathbf{C}^k$. The construction of the transformed deconvolvers is achieved by solving a set of partial differential equations (the $\overline{\partial}$ or Cauchy-Riemann equations) using the known values of on the zero sets of the $\widehat{\nu}_i$. These may be thought of as interpolation formulae from several complex variables as they reconstruct analytic functions based on distinct values at a discrete set of isolated points (see [1] for the most general version of this construction). The ν_i are constructed simultaneously by taking the inverse Fourier-Laplace transform of the construction formulae.

Finding strongly coprime sets and demonstrating strong coprimeness are nontrivial tasks. The following subsets of $\mathbf{R}^k$ can be chosen so that their characteristic functions (1 on the set, 0 off the set) are strongly coprime : two intervals on $\mathbf{R}$; two disks on $\mathbf{R}^2$; three rotations of a square in $\mathbf{R}^2$; $n+1$ cubes in $\mathbf{R}^k$. Explicit deconvolution formulae for each of these have also been calculated (see [1]).

Generally, the deconvolvers can be chosen so that the support of each is contained in the largest support of the convolvers. This means that deconvolution at a point $t \in \mathbf{R}^k$ depends only on data near t. In particular, we have precise knowledge about the domain of a digital version of the deconvolver, although the spacing or pitch of the samples over that domain still depends on the bandwidth of the reconstruction. Also, for digital deconvolution, the computational complexity of the compactly supported ν_i is essentially optimal. The total number of multiplies and adds is the minimum possible.

The results on deconvolution are for distributions with compact support. The transforms of these distributions are entire functions and hence have at most isolated zeroes. Further, these results are about a set of these distributions such that no isolated zero is common to all in the set. Obviously, these deconvolution results are not applicable to convolvers b that are bandlimited. They are, however, applicable to the compactly supported convolvers in a system : if the μ_i, $i = 1, \ldots, n$ are strongly coprime and each μ_i is the composition of all of the compactly supported convolvers of the ith linear system, and if b is the composition of the bandlimited convolvers for each system, then, with deconvolvers ν_i, $i = 1, \ldots, n$,

$$\sum_{i=1}^{n} b * \mu_i * \nu_i = b.$$

That is, we can reconstruct up to the bandlimited response.

An extensive discussion of the applications of these deconvolution techniques is given in [6]. For this paper, we give only a brief overview, referring the reader to [6] for the details.

A primary class of sensor components that are modeled by bandlimited convolvers is the class described by an aperture response function for electromagnetic or acoustic waves. The beam pattern for acoustic transducers and for radar antennas, as well as for the real and synthetic arrays of either, is a convolver whose Fourier transform is bandlimited. (The transform rather than the convolver has compact support.) The same is true for the point spread function of an optical lens. The bandlimit for all of these is the familiar diffraction limit. Here we seek compactly supported strongly coprime convolvers in a system as above.

On the other hand, the sensor components which are modeled by compactly supported convolvers nearly all of the various pulses used in radar and sonar. In radar and sonar, the convolvers describe the modulation of the transmitted carrier frequency and the deconvolvers describe the filters used to process the demodulated return signal. The deconvolution theorem says in this context that if a set of strongly coprime pulses are transmitted and if each pulse of the set is separately processed, then the strongly coprime pulses can be completely compressed, without sidelobes, in the case of zero (or known) doppler shift. The types of pulses that can be used to make up a strongly coprime set include any of the pulse types used in radar and sonar. Pairs of strongly coprime chirps is one such example. This feature suggests that it is possible to consider the enhancement of the range resolution performance of pulse types that perform optimally relative to some other performance measure.

It also should be noted that the deconvolvers are not unique. This non-uniqueness allows one to develop deconvolvers that are optimal with respect to a given condition. In [5] and [6], deconvolvers that optimize

the signal to noise ratio of the signal relative to white Gaussian noise were constructed. Other types of noise could be dealt with in similar fashion.

The deconvolution theory is also applicable to the acoustic signals $f(x,y,z,t)$ which change in both space and time because of local atmospheric effects. (Low frequency signals are particularly sensitive to these effects.) The output signals $s(x,y,z,t)$ vary with both the position of the sensors and the time at which the signal is received. The theory has to be suitably modified to take this problem into account. Ideally, we would like to have a Green's function $G(x,y,z,t;x_0,y_0,z_0,t_0)$ for our model of the acoustic wave. Then, $f(x,y,z,t)$ is represented by an integral equation with terms G, the initial signal information g and $\frac{\partial f}{\partial t} = h$, atmospheric inhomogeneity A, and boundary conditions – namely,

$$f = \frac{\partial G}{\partial t} * g + G * h + \int^t G * A + \text{boundary terms.}$$

If the boundary effects can be repressed, then

$$\begin{aligned} & f(x,y,z,t) \\ &= \frac{\partial G(x,y,z,t)}{\partial t} * g(x,y,z) + G(x,y,z,t) * h(x,y,z) \\ &+ \int^t G(x,y,z,\tau) * A(x,y,z,\tau). \end{aligned}$$

If f is overdetermined by a strongly coprime set of sensors μ_i as $s_i = f * \mu_i$, convolution with the sensors is over the spatial variables x,y,z. Then for ν_i as above, and as convolution is over x,y,z,

$$\begin{aligned} & \sum_{i=1}^n s_i * \nu_i \\ = & \sum_{i=1}^n f * \mu_i * \nu_i \\ = & \sum_{i=1}^n \left(\frac{\partial G}{\partial t} * g + G * h + \int^t G * A \right) * \mu_i * \nu_i \\ = & \left(\frac{\partial G}{\partial t} * g + G * h + \int^t G * A \right) * \sum_{i=1}^n (\mu_i * \nu_i) \\ = & \frac{\partial G}{\partial t} * g + G * h + \int^t G * A \\ = & f. \end{aligned}$$

We close this section by looking at a problem of getting a subpixel resolution of a pixel image. The mathematical model for this image is the integrator of an image f over a time interval in $\mathbf{R}^k, k = 1,2,3$. Thus, this model is $s = f * \mu$, where μ is the characteristic function χ of some bounded region in $\mathbf{R}^k$ (and therefore is compactly supported). We consider such an equation in one dimension, letting $\mu = \chi([-1,1])$.

First, it is easy to see that solving $s = f * \mu$ for f is an ill-posed problem. Let $f_k = \sin(k\pi t)$, $g_k = \cos(k\pi t)$. Then, $\mu * f_k = 0$ and $\mu * g_k = 0$. Moreover, if g equals any linear combination of f_k and g_k, then $\mu * g = 0$. To see the validity of these equations, simply note that support$(\widehat{f_k}, \widehat{g_k})$ is contained in the zero set of $\widehat{\mu}$ and that the Fourier-Laplace transform is linear and transforms convolution into multiplication.

To circumvent this ill-posedness, overdetermine the signal f by convolving it with the characteristic functions of the intervals $[-1,1]$ and $[-\sqrt{q}, \sqrt{q}]$. For any integer q that is not a perfect square,

$$\mu_1(t) = \chi([-1,1])(t)\ ,\ \mu_2(t) = \chi([-\sqrt{q}, \sqrt{q}])(t)$$

are a strongly coprime pair. The Fourier-Laplace transforms of these functions are

$$2\frac{\sin(\zeta)}{\zeta}\ ,\ 2\frac{\sin(\sqrt{q}\zeta)}{\zeta},$$

respectively. Using elementary methods, one can reduce the proof that these characteristic functions are strongly coprime to estimates in number theory (see [3]).

Two constructions of deconvolvers ν_1, ν_2 are presented. We want these ν_i to satisfy

$$\nu_1 * \mu_1 + \nu_2 * \mu_2 = \delta.$$

By taking the Fourier-Laplace transform of both sides of this equation, we get the Bezout equation

$$\widehat{\nu_1} \cdot \widehat{\mu}_1 + \widehat{\nu_2} \cdot \widehat{\mu}_2 = 1.$$

One solution to this second equation is

$$\widehat{\nu_i}(\zeta) = \frac{\overline{\widehat{\mu_i}(\zeta)}}{\sum_{i=1}^{2} |\widehat{\mu_i}(\zeta)|^2}, \ i = 1, 2.$$

Then, $\nu_i(t)$ are solved for by taking the inverse transform. These ν_i are not compactly supported.

In the second construction, let $\mathcal{Z}$ be the set of zeroes of $\widehat{\mu_1}$, i.e., $\mathcal{Z} = \{\pm n\pi, \ n \in \mathbf{N}\}$. Note, $\frac{d}{dz}\widehat{\mu_1}(z) \neq 0$ for $z \in \mathcal{Z}$. Then, we let

$$\widehat{\nu_1}(\zeta) = \sum_{z \in \mathcal{Z}} \left[\frac{1}{\frac{d}{d\zeta}\widehat{\mu_1}(z) \cdot \widehat{\mu_2}(z)} \right] \left[\frac{\widehat{\mu_2}(\zeta) - \widehat{\mu_2}(z)}{\zeta - z} \right],$$

$$\widehat{\nu_2}(\zeta) = \sum_{z \in \mathcal{Z}} \left[\frac{1}{\frac{d}{d\zeta}\widehat{\mu_1}(z) \cdot \widehat{\mu_2}(z)} \right] \left[\frac{\widehat{\mu_1}(\zeta) - \widehat{\mu_1}(z)}{\zeta - z} \right].$$

A straightforward calculation shows that these satisfy the Bezout equation. Again, ν_i are solved for by taking the inverse transform. Unlike the deconvolvers above, these ν_i are compactly supported.

A simulation which summarizes these results is presented in figure 1. In this simulation, we let the value of q equal 2.

3 Deconvolution and the Gabor and Wavelet Transforms

Given that it is possible to reconstruct remotely sensed signals by the methods outlined in the previous section, we are now lead to consider the problem of signal analysis and data compression. As an example, consider image information from a relatively simple object. In uncompressed form, this can be a large piece of information. Assuming only 8 grey scales and a pixel resolution of only 300 x 300 pixels per square inch, a 10 x 10 inch image requires 3.375 megabytes of memory. If the methods described in section 2 are implemented to perform the simplest subpixel resolution of the image, the resultant image would now take up 20.25 megabytes.

An effective method of producing a simultaneous analysis and data compression of signals lies in the localized Fourier methods introduced by Gabor, and recently developed by Meyer, Daubechies, Grossmann, etc. (see [7], [9]). These methods, know as Gabor or wavelet analysis, produce a representation of a signal as a sequence of coefficients, from which the signal can be reconstructed by expansion with a set of basis functions. The coefficients serve not only to compress the signal, but also represent localized Fourier information about the signal. Each coefficient represents the function for a certain region in both time and frequency.

The Gabor and wavelet transforms have a large advantage over the traditional Fourier transform as analysis tools for signal processing. Let f be a function for which we already have a Fourier and localized Fourier analysis, and let g be a function which differs from f in only one small region of phase space. The calculation of the Gabor and/or wavelet coefficients of g requires that we calculate new coefficients for g only in the region in which g differs from f. In contrast, the calculation of the Fourier transform of g requires that the integral be calculated over the entire time domain.

We assume that the class of functions we are interested in working with have finite energy in time, i.e., $\mathbf{L}^2(\mathbf{R}^k)$. The Fourier transform maps this space onto functions of finite energy in frequency, $\mathbf{L}^2(\widehat{\mathbf{R}^k})$.

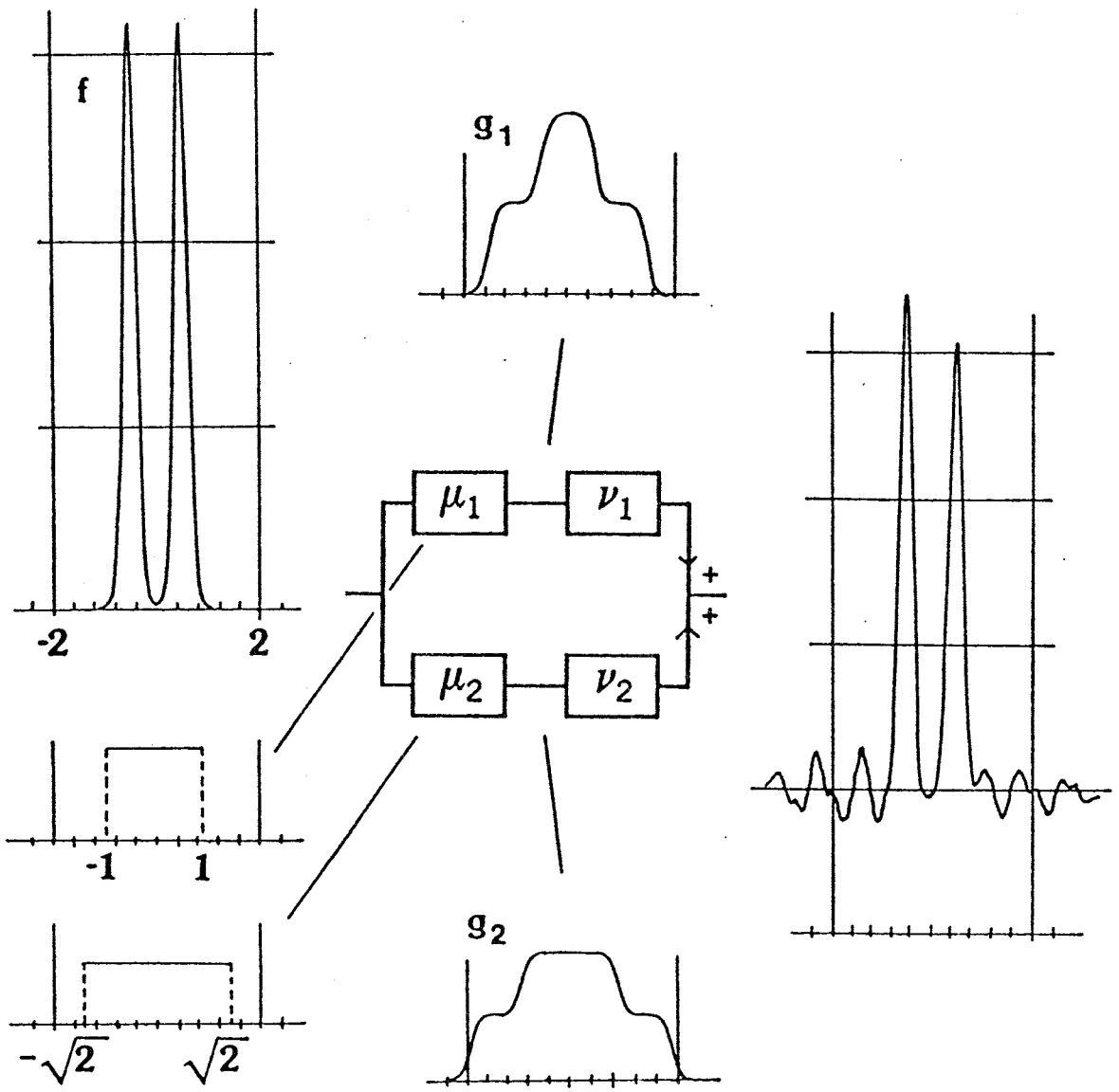


Figure 1

Properties of this transform guarantee that it is impossible to have a function that is compactly supported in both time and frequency. Moreover, the more concentrated a function is in time, the more spread out the function is in frequency, as is seen by considering the relationship between second moments of f and $\widehat{f}$. The amount of information that we can get about a given function on a compact region of phase space $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$ is limited by the classically known uncertainty principle, namely,

$$\| tf(t) \|_2^2 \cdot \| \omega \widehat{f}(\omega) \|_2^2 \geq \left(\frac{1}{16\pi^2} \right) \cdot \| f(t) \|_2^2,$$

where $\| \cdot \|_2$ denotes the $\mathbf{L}^2$ norm.

This uncertainty imposes constraints on any functions one would choose as a basis in combined time and frequency space, e.g., the choice of a singular function, supported at a single point in time, transforms to the entire frequency axis. We want to choose our basis so that each element in the basis is "limited" (as can be made precise by considering second moments) to a finite region in $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$. The idea of trying to do this localized analysis goes back to Gabor (see [7] and [9]).

Choose a function g of finite energy, normalized so that it has norm equal to one. We define the Gabor transformation of a function f of finite energy by

$$\Gamma_g f(\alpha, \beta) = \int_{\mathbf{R}^k} f(t) \, \overline{g(t-\alpha)} \, e^{(-2\pi i \beta t)} \, dt,$$

where $\overline{\zeta}$ denotes the complex conjugate of ζ. If $g \times \widehat{g}$ is concentrated near zero in $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$, then $\Gamma_g f(\alpha, \beta)$ will give a picture of f around time α and frequency β. The transform $\Gamma_g f$ is a unique representation of local Fourier information about the function, and we can completely recover f via the inverse transform

$$f(t) = \int_{\mathbf{R}^k} \int_{\widehat{\mathbf{R}^k}} \Gamma_g f(\alpha, \beta) \, g(t-\alpha) \, e^{(2\pi i \beta t)} \, d\alpha \, d\beta.$$

If we partition $\mathbf{R}^k$ into identical regions, the choice of g will lead to a tessellation of $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$ where the corresponding regions of $\widehat{\mathbf{R}^k}$ are determined by the effective spread of $\widehat{g}$, as determined by is second moment. We look to get localized Fourier information in the $(m,n)^{th}$ region of phase space as follows. Define the Gabor elementary function (GEF) of order (m,n) by

$$\phi_{mn}(t) := g(t - mT) \cdot e^{(2\pi i n \Omega t)},$$

and refer to g as a generating function. The function ϕ is centered at position $(t = nT, \omega - m\Omega)$ in $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$. Gabor chose g to be an shifted and scaled Gaussian, as this minimizes the combined effective spread in $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$. For $\Omega \cdot T = 1$, the set $\{\phi_{mn}\}$ is complete [9], and so for appropriate coefficients a_{mn}, we can write

$$f(t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} \, \phi_{mn}(t)$$

for an arbitrary function f of finite energy. But, as the set of GEF's is not orthogonal, calculation of the coefficients a requires solving for an auxiliary function $\gamma(t)$ so that

$$\int_{\mathbf{R}^k} \gamma(t) \, \overline{g(t-mT)} \, e^{(-2\pi i n \Omega t)} \, dt = \langle \gamma(t) \, , \, g(t - mT) \, e^{(2\pi i n \Omega t)} \rangle = \delta_m \, \delta_n.$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\delta_m \delta_n$ is the product of two Kronecker delta functions. Then,

$$a_{mn} = \langle f(t) \, , \, \gamma(t - mT) \, e^{(-2\pi i n \Omega t)} \rangle.$$

For ease of notation, we write this as

$$a_{mn} = \langle f \, , \, \gamma_m e_n \rangle.$$

Each coefficient a_{mn} is a value uniquely representing the function $f(t)$ in the $(nT, m\Omega)$ region of $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$.

The wavelet transform of a function is also an operator that works locally in phase space. It is generated by taking both translations and dilations of a normalized generating function. We define the wavelet transformation of a function f of finite energy by

$$\Psi_g f(\alpha, \beta) = \int_{\mathbf{R}^k} f(t)\, e^{-\beta/2}\, \overline{g(e^{-\beta}t - \alpha)}\, dt.$$

Again, the transform $\Psi_g f$ is a unique representation of local Fourier information about the function, and we can completely recover f via the inverse transform

$$f(t) = \int_{\mathbf{R}^k} \int_{\widehat{\mathbf{R}^k}} \Psi_g f(\alpha, \beta)\, e^{-\beta/2}\, g(e^{-\beta}t - \alpha)\, d\alpha\, d\beta.$$

If β is a large negative number, then $e^{-\beta/2}g(e^{-\beta}t)$ is highly concentrated about $t = 0$ in $\mathbf{R}^k$, yet has the same energy as g. As β goes to $-\infty$, $e^{-\beta/2}g(e^{-\beta}t)$ becomes increasingly concentrated at $t = 0$. Therefore, for each fixed β and varying α, $\Psi_g f(\alpha, \beta)$ displays the information about f that can be found in a given frequency band. As β goes to $-\infty$, $\Psi_g f(\alpha, \beta)$ displays the small-scale, high-frequency information in the signal f. As β goes to ∞, $\Psi_g f(\alpha, \beta)$ displays the large-scale, low-frequency information in the signal f.

In the development of the discrete version of the wavelet transform, we need a generalized notion of a basis called a "frame". We can think of a frame as a "basis plus", in that we can represent functions in terms of the frame, but that this representation is not unique nor do we have that the frame functions are orthogonal (see [7]).

We define the wavelet frame function of order (m, n) by

$$\psi_{mn}(t) := \alpha^{-n/2}\, g(\alpha^{-n/2}t - m\beta),$$

and refer to g again as a generating function. For appropriate coefficients b_{mn}, we can write

$$f(t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} b_{mn}\, \psi_{mn}(t)$$

for an arbitrary function f of finite energy. The coefficient b_{mn} are not unique. These coefficients can be recovered by

$$\int_{\mathbf{R}^k} f(t)\, \frac{1}{\alpha^2}\psi_{mn}(t)\, dt = \langle f(t)\, ,\, \frac{1}{\alpha^2}\psi_{mn}(t)\rangle = b_{mn}.$$

Each coefficient b_{mn} is a value representing the function $f(t)$ in a region of $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$.

Now, we assume that we have a fixed signal f overdetermined by a system of convolution equations,

$$s_i = f * \mu_i,\ i = 1, \ldots, k,$$

for a strongly coprime set of convolvers. Thus, there exist compactly supported distributions

$$\nu_i,\ i = 1, \ldots, k$$

such that

$$\delta = \mu_1 * \nu_1 + \ldots + \mu_k * \nu_k,$$

where δ denotes the Dirac delta. This in turn gives

$$\begin{aligned} & s_1 * \nu_1 + \ldots + s_k * \nu_k \\ &= (f * \mu_1) * \nu_1 + \ldots + (f * \mu_k) * \nu_k \\ &= f * (\mu_1 * \nu_1 + \ldots + \mu_k * \nu_k) \\ &= f * \delta \\ &= f. \end{aligned}$$

If we have a system of GEF's for a sufficiently fine tessellation of $\mathbf{R}^k \times \widehat{\mathbf{R}^k}$ and corresponding auxiliary function $\gamma(t)$, we may calculate the Gabor coefficients of the original signal $f(t)$ as follows. Using the notation $a_{mn} = \langle f, \gamma_m e_n \rangle$, and if $\tilde{\nu}_i(t) = \nu_i(-t)$, we have

$$\begin{aligned}
&\sum_{i=1}^{k} \langle s_i \ , \ \tilde{\nu}_i * \gamma_m e_n \rangle \\
&= \sum_{i=1}^{k} \langle (f * \mu_i) \ , \ \tilde{\nu}_i * \gamma_m e_n \rangle \\
&= \sum_{i=1}^{k} \langle (f * \mu_i * \nu_i) \ , \ \gamma_m e_n \rangle \\
&= \langle \sum_{i=1}^{k} (f * \mu_i * \nu_i) \ , \ \gamma_m e_n \rangle \\
&= \langle f * \left(\sum_{i=1}^{k} \mu_i * \nu_i \right) \ , \ \gamma_m e_n \rangle \\
&= \langle f * \delta \ , \ \gamma_m e_n \rangle \\
&= \langle f \ , \ \gamma_m e_n \rangle \\
&= a_{mn}.
\end{aligned}$$

Similarly, if we have a frame of wavelet functions of sufficiently fine resolution, we may calculate the wavelet coefficients of the original signal $f(t)$.

$$\begin{aligned}
&\sum_{i=1}^{k} \langle s_i \ , \ \tilde{\nu}_i * \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= \sum_{i=1}^{k} \langle (f * \mu_i) \ , \ \tilde{\nu}_i * \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= \sum_{i=1}^{k} \langle (f * \mu_i * \nu_i) \ , \ \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= \langle f * \left(\sum_{i=1}^{k} \mu_i * \nu_i \right) \ , \ \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= \langle f * \delta \ , \ \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= \langle f \ , \ \frac{1}{\alpha^2} \psi_{mn} \rangle \\
&= b_{mn}.
\end{aligned}$$

Thus, by combining the "linearity" of each procedure, it is possible to do a signal recovery and data compression simultaneously. This is also the natural way to turn the analog deconvolution process we have described into a digital one.

Finally, it should be noted that in a very recent paper - *"Inversion of the local Pompeiu transform,"* by C. Berenstein, R. Gay, and A. Yger [2], a local version of deconvolution has been worked out. The result is as follows :

> Suppose we know the averages of f on disks of radii r and r only when these disks lie in some fixed disk $B(0, R)$. Then we can solve for ν_1 , ν_2 such that
>
> $$\chi_{disk\,1} * \nu_1(t) \ + \ \chi_{disk\,2} * \nu_2(t) \ = \ \delta(t).$$

This then allows one to reconstruct f on $B(0,R)$, even if f is obscured in certain subregions of $B(0,R)$. In that paper, the ν_i were expressed in terms of Fourier series. They are currently being recalculated in terms of the wavelet expansion. Then, if η_i are deconvolvers which differ from the ν_i in only one small region of phase space, the calculation of the wavelet coefficients of η_i requires that we calculate new coefficients for η_i only in the region in which the η_i differ from the ν_i. This in turn will make the result more useful in signal and image processing.

References

[1] Berenstein, C. A. and Yger, A., "Analytic Bezout identities," *Advances in Mathematics* 10, 51-74 (1989).

[2] Berenstein, Carlos A., Roger Gay, and Alain Yger, "Inversion of the local Pompeiu transform," *J. Analyse Math.* 54, 259-70 (1990).

[3] Berenstein C. A., Yger, A., and Taylor, B. A., "Sur quelques formules explicites de deconvolution," *Journal of Optics (Paris)* 14, 75-82 (1983).

[4] Berenstein, Carlos A., and Alain Yger, "Le probleme de la deconvolution," *J. Funct. Anal.* 54, 113-60 (1983).

[5] Berenstein C. A., and Patrick, E. V., "Exact deconvolution for multiple convolution sensors – an overview plus performance characterizations for imaging sensors," *Proceedings of the IEEE (Special Issue on Multidimensional Signal Processing)*, Vol. 78, No. 4, 723-34 (1990).

[6] Berenstein, C. A., Casey, S. D., and Patrick, E. V., "Systems of convolution equations, deconvolution, and wavelet analysis," U of MD Systems Research Center whitepaper (1990).

[7] Daubechies, I., "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. on Information Theory*, Vol. 36, No. 5 (1990).

[8] Morozov, V. A., *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York (1984).

[9] Porat, Moshe, and Zeevi, Yehoshua, "The generalized Gabor scheme of image representation in biological and machine vision," *IEEE Tr. on Pattern Anal. and Machine Intelligence* 10, No. 4, 452-68 (1988).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 49

Realization of the Belgian Time Scale.

F. Collin and P. Pâquet

Royal Observatory of Belgium

Abstract

Till 1990 the Time scale of the Royal Observatory of Belgium UTC(ORB) was derived from one of the three atomic clocks in operation at the Royal Observatory.

To reduce the drift of UTC(ORB) with respect to UTC, a permanent control of the evolution of the three atomic clocks has been developed.

In a first step, the main components of the control are a PC and a Time Interval Meter ; the software estimates the drift of each atomic clock with respect to UTC or to GPS and derives in real time a mathematical clock which is in agreement, at the level of 10^{-13}, with the choosen external reference.

In a second step the mathematical clock is materialised by slaving a rubidium oscillator. The rubidium oscillator is automatically monitored by changing the voltage applied to the C-field with a monitored voltage regulator; the frequency output of the rubidium is used to generate UTC(ORB).

Windows used for the estimation of the drift parameters are typically of the order of 50 to 100 days.

Influence of the environment parameters, humidity and pressure, are also given.

1 Configuration

The Time scale of the Royal Observatory of Belgium is experimentaly generated on the basis of three Cesium standard of frequency (HP-205, HP-804, OSA-312), one Rubidium Oscillator, one scanner, a Time Interval Meter, a Voltregulator, a GPS-receiver and, of course a PC which control most of the components above, especially the scanner, the Time Interval Meter and the Voltregulator.

A schematic representation of the configuration is given in figure 1.

2 Computation of the mathematical clock.

Be $Hmath$, the mathematical clock, close to an external reference choosen here as the GPS-time scale; between the two time scales we have, by definition :

$$GPS - Hmath(t) = \frac{\sum_{i=1}^{N}[(GPS - Ces(i))(t) - pred(GPS - Ces(i))(t)]}{N}$$

where:

- N is the number of Cesium clocks.
- $(GPS - Ces(i))(t)$ is the measurement of the difference between the clocks $Ces(i)$ and GPS at epoch t.
- $pred(GPS - Ces(i))(t)$ is the prediction, of the measurement above, computed with a linear regression based on a set of M data; including the measurement at epoch t.

$GPS - Hmath(t)$ appears as a mean of residuals and thus must be near zero. It means that $Hmath(t)$ is near $GPStime(t)$.

With the first results (table 1), we note that the stability of $Hmath$ with respect to $GPStime$ is $1.0698\,10^{-13}$.

The behavior of $Hmath - GPS$ is given from mjd 48233 to mjd 48315 in figure 2.

Clock	dates	mean of mean freq.	Allan variance
$Hmath - GPS$	7/12/90 to 28/02/91	$-5.9758\,10^{-16}$	$8.8743\,10^{-14}$

Table 1: result of $Hmath - GPS$

3 Realization of the mathematical clock with a Rubidium oscillator.

The mathematical clock is materialised by using a Rubidium Oscillator, $Horb$, which is slaved by changing the voltage applied to the C-field with a monitored voltage control regulator. The voltage to be applied is computed by the PC according the values $Hmath - Horb$.

Once per hour, the PC computes the drift of $Horb$ with respect of $Hmath$.

Twice per day the C-field of the Rubidium is corrected. So the drift of $Horb$ is close to the drift of $Hmath$ and thus close to the drift of GPStime.

In figure 3 the measurements of $Horb - GPS$ are given.

4 Results

The generation of UTC(ORB) with the system descrited above has been experimented in 1990 and is ready to become operational.

The actual results are presented in table 2.

Clocks	dates	mean of mean freq.	Allan variance
$HP\ 205 - GPS$	7/12/90 to 28/02/91	$2.0817\,10^{-13}$	$2.0079\,10^{-13}$
$HP\ 804 - GPS$	7/12/90 to 28/02/91	$8.4467\,10^{-14}$	$1.6746\,10^{-13}$
$OSA\ 312 - GPS$	7/12/90 to 28/02/91	$-3.4291\,10^{-13}$	$7.1919\,10^{-14}$
$Horb - GPS$	7/12/90 to 28/02/91	$5.8155\,10^{-14}$	$1.6267\,10^{-13}$

Table 2: actual results.

5 Correlation of clocks with humidity, temperature and pressure.

The atomic clocks are influenced by the variation of humidity, temperature and pressure.

We can see in figure 4 the measurements of the atomic clocks with respect of GPS; in figure 5, the temperature, the humidity and the atmospheric pressure of the cellar.

The correlation between atmospheric elements and the behaviour of the clocks is given by the cross-correlation coefficent $\mu_{ij}(k) = \frac{\sum(x_i(t)-\overline{x_i})(x_j(t+kr)-\overline{x_j})}{N\sigma_i\sigma_j}$ where

- $\overline{x_i}$ is the mean value of the time series (x_i).
- σ_i is the standard deviation of the time series (x_i).
- N is the number of elements in the time series (x_i).

In Table 3 the cross-correlation coefficients between the frequencies of the clocks and local parameters are given.

Clock		Temperature	humidity	pressure
HP205	measures	0.08	0.37	-0.19
	frequency	0.01	0.05	-0.06
	residuals	0.09	-0.04	-0.12
HP804	measures	-0.05	0.33	-0.49
	frequency	0.06	-0.10	0.11
	residuals	0.09	-0.04	-0.12
OSA312	measures	0.08	0.43	0.34
	frequency	-0.20	0.29	-0.24
	residuals	-0.19	-0.17	-0.08
Hmath	measures	0.10	-0.30	0.30
	frequency	0.07	0.10	0.15

The temperature of the cellar being constant on a long term period. We observe a correlation at the level of 0.35 only with ths humidity. It is not really clear because during the test period the humidity of the cellar decreased more or less linearly (figure 6); in consequence the effect of the humidity on the clock cannot be separated from its natural drift.

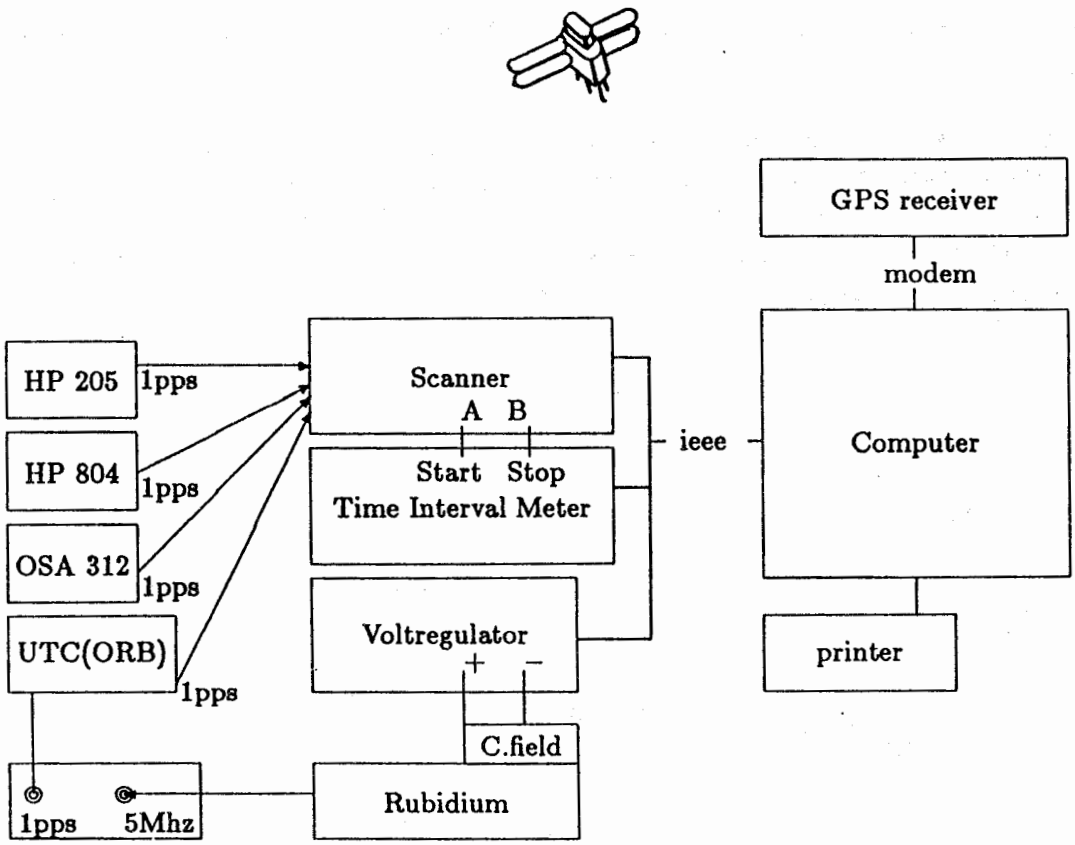


Figure 1: representation of the configuration used to generate UTC(ORB).

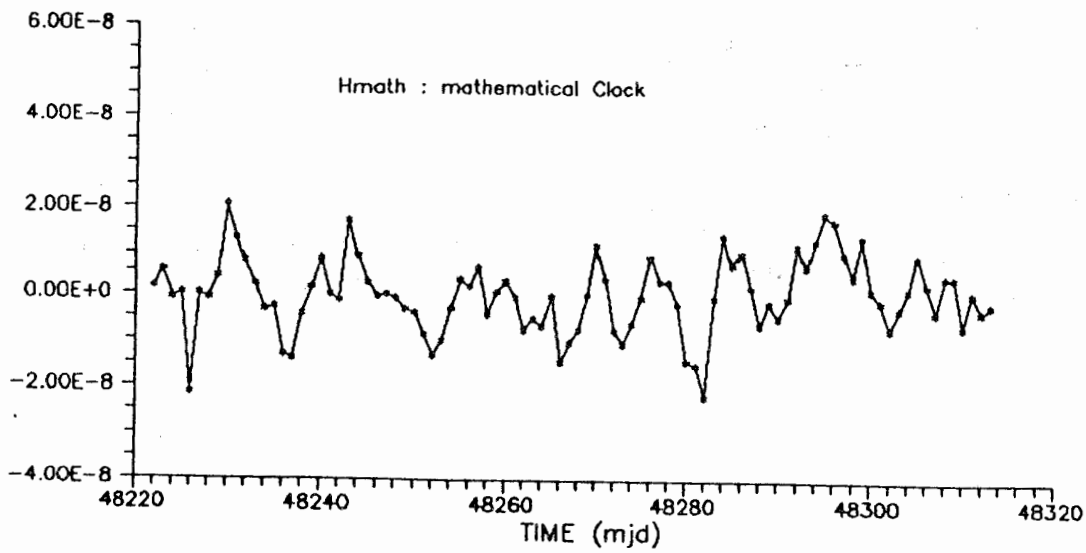


Figure 2: *GPS − Hmath* :the mathematical clock, computed as a mean of the three atomic clocks

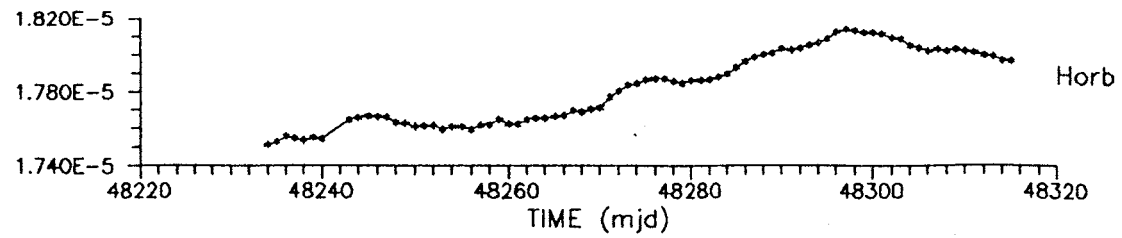


Figure 3: *Horb* − *GPS* (seconds): drift of the physical clock slaved by the mathematical clock

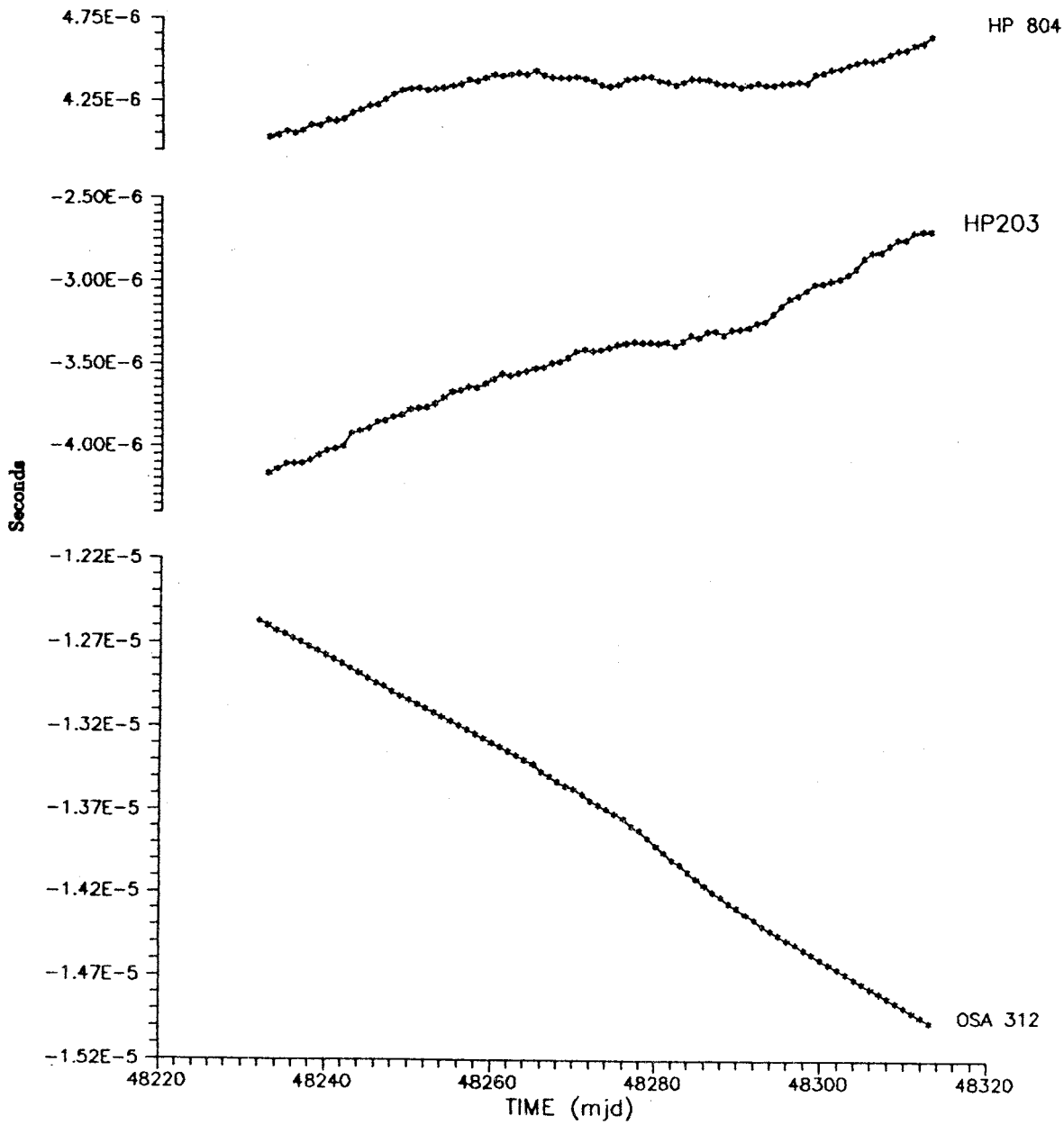


Figure 4: Drift of the three Cesium clocks, with respect to GPS

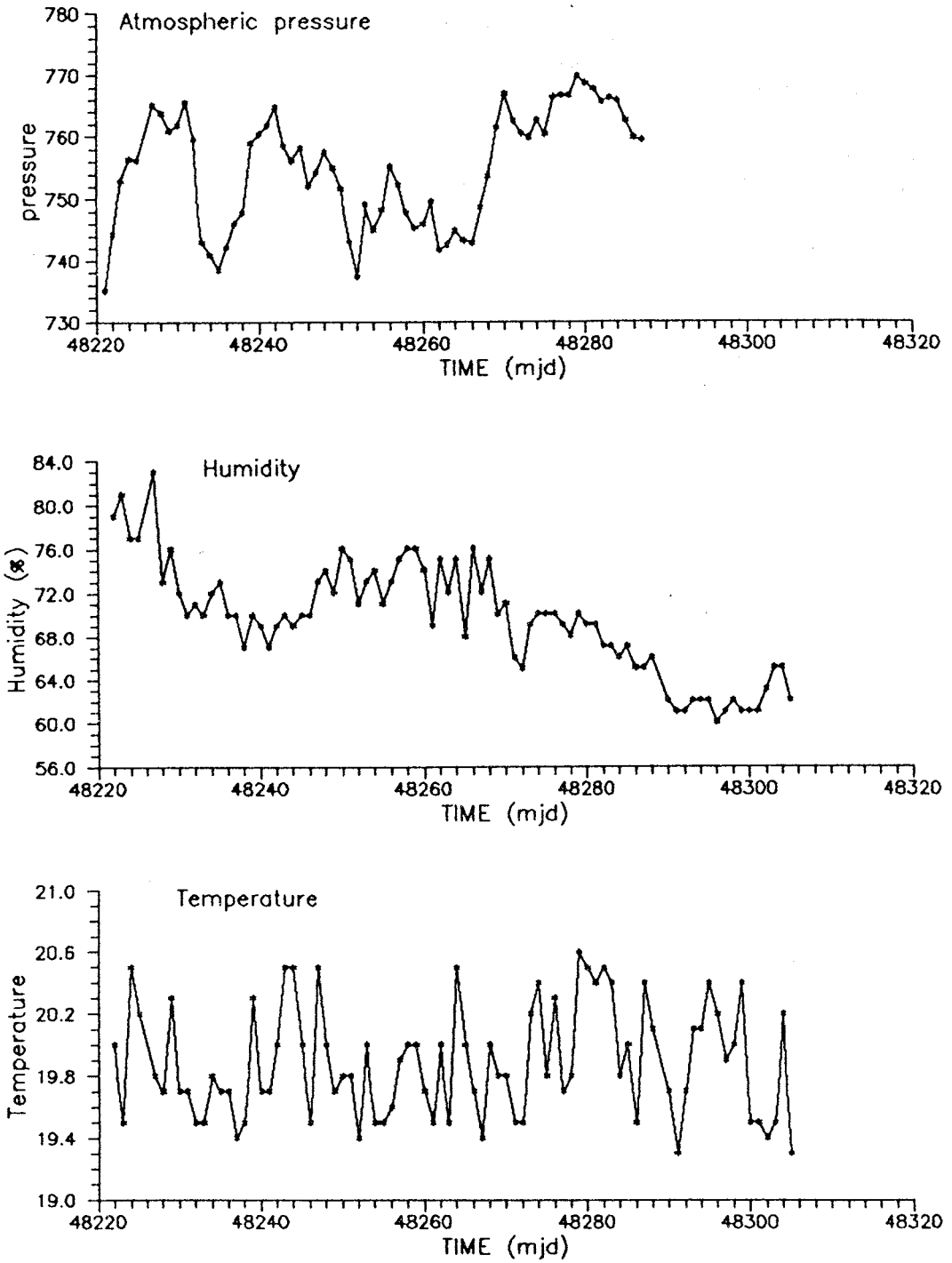


Figure 5: atmospheric coefficients

Session POSTER Number 50

COMPARISON OF SOVIET RADIONAVIGATION SYSTEM AND SYLT-LORAN WITH GPS AS A REFERENCE

Tapio Mansten (1), Kalevi Kalliomäki (1,2), Bengt Selin (3)

(1) Technical Research Centre of Finland, Metrology Research Institute (MTI), Po. Box 34, SF-02151 Espoo, FINLAND
(2) University of Oulu, Department of Electrical Engineering, SF-90570 Oulu, FINLAND
(3) Swedish Telecom Radio, Radiolaboratory (STA), Rlm D 108, S-13680 Haninge, SWEDEN

Abstract

The Metrology Research Institute (MTI) uses Loran-C and a Soviet radionavigation system (RNS) as a secondary means of time and frequency maintenance by comparing our measurement results with those of our neighbouring laboratories. The results show that by common view measurements of Loran-C Sylt national time can be maintained at 100 ns level/10 days. By RNS common view measurements about 500 ns/10 day level is achievable.

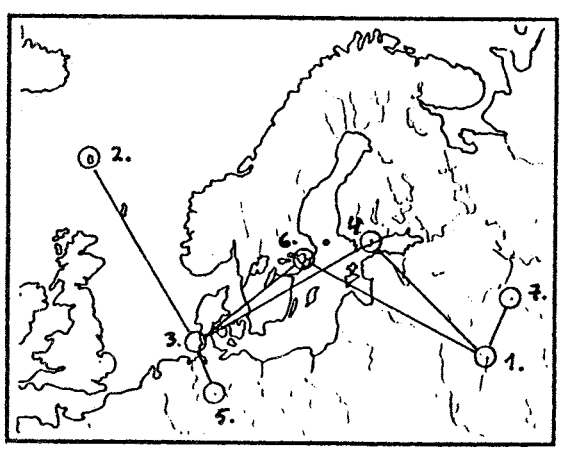
1. Introduction

The phase of Loran-C transmitter Sylt is monitored and reported by several time and frequency laboratories. The master station, Ejde, situates on Faeroe Islands and is monitored and reported by USNO.

The Soviet Union has a radionavigation system, RNS, that closely resembles the Loran-C system. The master station is equipped with a hydrogen maser and situates at Bryansk. The National Scientific and Research Institute for Physical Technical and Radiotechnical Measurements, VNIIFTRI, near Moscow monitors the phase of the master station and publishes the results in time information letters.

The Metrology Research Institute (MTI) contains, amongst others, the National Measurement Centre of Time and Frequency in Finland. MTI time and frequency are based on USNO MC time via a GPS satellite receiver. The time and frequency are traceable to BIPM via atomic clock transports with SU and STA and via time reports of BIBM and USNO. MTI also monitors the Loran-C Sylt transmitter as a back-up and compares the results with those of STA and PTB. In addition MTI monitors the RNS Bryansk master station phase together with STA and Soviet Union and uses the results as a secondary back-up for time and frequency. Both results are reported monthly to VNIIFTRI and STA.

Swedish Telecom Radio (STA) maintains the time and frequency in Sweden. Its time is traced to UTC via the GPS common view measurement program coordinated by BIPM, Paris. STA monitors and reports the phases of Loran-C Sylt and Bryansk RNS transmitters with MTI as an agreed Euromet project.



Transmitters: 1) Bryansk, 2) Ejde, 3) Sylt
Laboratories: 4) MTI, 5) PTB, 6) STA, 7) VNIIFTRI

Figure 1. The situations of transmitters and laboratories concerned.

The positions of transmitters and receivers concerned can be seen from the map in figure 1. Some of the distances are represented in table 1.

Table 1. Approximate distances between the transmitters and receivers concerned.

Stations	Distance km	Over sea %
Bryansk - STA	1200	35
Bryansk - MTI	970	6
Bryansk - VNIIFTRI	360	0
Sylt - STA	770	30
Sylt - MTI	1155	60
Sylt - PTB	330	30

2. Measurements

We have compared continuous data of the Sylt and Bryansk measurements for most of the year 1990. Comparison was done by calculating all the measurement results back to UTC by using appropriate time information letters according to table 2.

Table 2. Time information calculation paths:

MTI:	UTC(MTI)	=	UTC(USNO) via	GPS		
UTC-UTC(Rx1Sylt)	= UTC-UTC(USNO)	+	UTC(MTI)	-	UTC(Rx1Sylt)	+ c1
UTC-UTC(Rx1RNS)	= UTC-UTC(SU)	+	UTC(MTI)	-	UTC(Rx1RNS)	+ c2
STA:						
UTC-UTC(Rx2Sylt)	= UTC-UTC(STA)	+	UTC(STA)	-	UTC(Rx2Sylt)	+ c3
UTC-UTC(Rx2RNS)	= UTC-UTC(STA)	+	UTC(STA)	-	UTC(Rx2RNS)	+ c4
PTB:						
UTC-UTC(Rx3Sylt)	= UTC-UTC(PTB)	+	UTC(PTB)	-	UTC(Rx3Sylt)	+ c5
USNO:						
UTC-UTC(Rx4Ejde)	= UTC-UTC(USNO)	+	UTC(USNO)	-	UTC(Rx4Ejde)	+ c6
VNIIFTRI:						
UTC-UTC(Rx4RNS)	= UTC-UTC(SU)	+	UTC(SU)	-	UTC(Rx4RNS)	+ c7

c1	+	0.5 us
c2	-	70 us
c3	+	1.25 us
c4	-	4000 us
c5	-	0.9 us
c6	+	0.6 us
c7	-	1202 us

The arbitrary constants c1..c7 were fixed in the beginning of the measurement period. Their purpose is to keep the average values of the signals near zero for mutual comparison.

The results were averaged in 11 day groups for 9-ending MJD's. Only the days that all of the laboratories had reported were included. The 11-day average values were then compared to each other by calculating their mutual differences.

3. Results

The Sylt result differences between MTI, PTB and STA stayed well within +/-100 ns during almost all of the 240 day measurement period. The difference averages were near zero and only small trends were observed. The USNO measurement results of Ejde differ strongly from the Sylt results, having a statistically significant trend compared to them. The differences are represented in figures 2 and 3 and the statistics of the results in table 3.

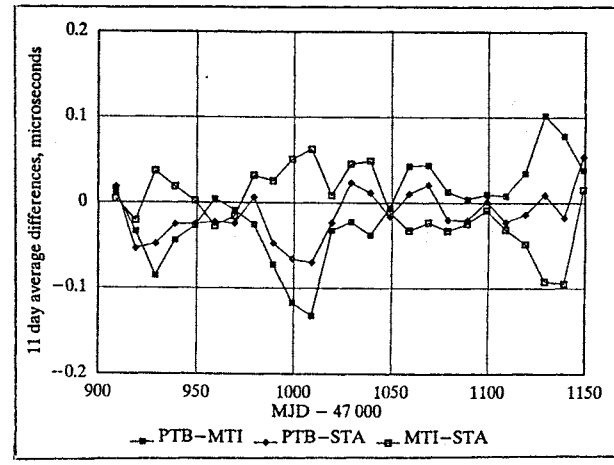


Figure 2. Mutual differences between the 11-day averages of Sylt measurement results form PTB, STA and MTI.

Table 3. Statistical information of differences in Sylt measurement results.

Difference	Average	SD ns	Trend ns/day
PTB - MTI	-10	55	0.5
PTB - STA	-15	30	0.2
MTI - STA	-5	40	-0.3
PTB - Ejde	200	190	2.4
MTI - Ejde	210	160	2
STA - Ejde	210	180	2.2

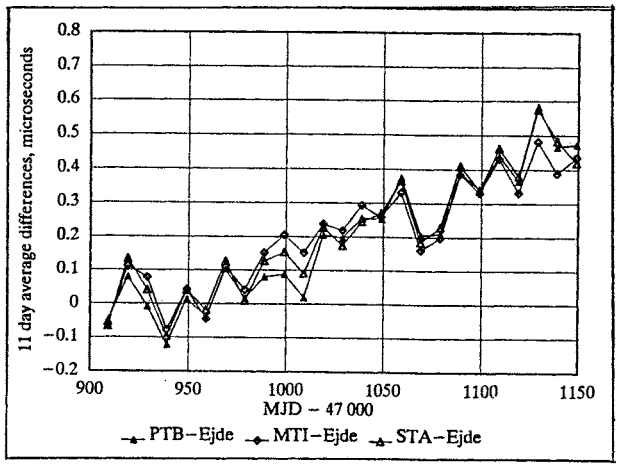


Figure 3. Mutual differences between the 11-day averages of Sylt measurement results from Ejde and PTB, STA and MTI.

The noise in the RNS result differences was almost a decade higher than in the Norwegian Sea measurements although the distances are about the same. The differences can be seen in figure 4 and the statistics of them in table 4.

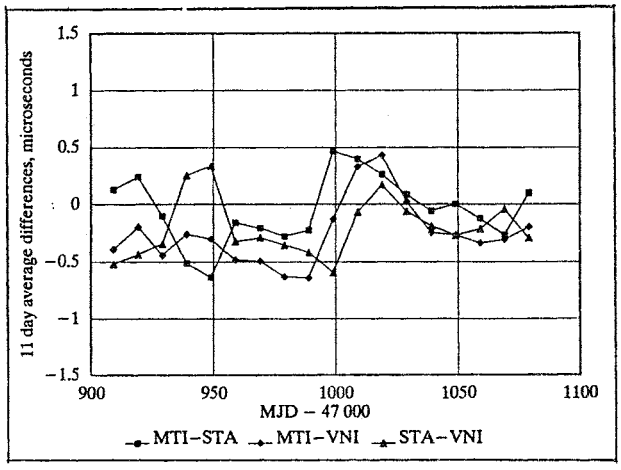


Figure 4. Mutual differences between the 11-day averages of RNS measurement results from MTI, STA and VNIIFTRI.

Table 4. Statistical information of the differences in RNS measurement results.

Difference	Average ns	SD ns	Trend ns/day
MTI - STA	-50	285	0
MTI - VNIIFTRI	-250	280	0
STA - VNIIFTRI	-200	250	0

4. Conclusions

The Norvegian Sea "common view"-measurement offers a good way of comparing MTI's time with that of other laboratories on a 10-day basis. Achievable level of accuracy is about +/-100 ns. Because PTB's master clock is very stable and their Sylt reports come fast (usually in two weeks), MTI can relatively quicly verify the national time maintained via our GPS receiver.

STA can also use the results as a back-up, especially if the still "not fully-operational" GPS system sometimes will be unoperational or restricted to military purposes only.

The Soviet Union measures Sylt-Loran, too, and uses the reports of MTI and STA for comparison purposes.

Sylt is phase-locked to Ejde master station, so one would not expect the observed long-time deviation trends between the two stations. Probably they are due to some slow weather patterns in the North Atlantic propagation path. Clearly it is not advisable to use USNO reports of Ejde for maintaining time with a Sylt receiver, but even then the error would seem to lie inside +/-1 us boundaries.

The RNS signal phase is more noisy and subject to frequent phase jumps and relatively fast drift. Because of higher phase scatter no long-time trends can be seen. The common view time comparison accuracy is limited to +/-500 ns level, but the signal is strong in the whole Finland, so it can be used for time-keeping purposes far from the coast. Earlier the Soviet Union probably used the measurement results as a tertiary reference to UTC Loran-C Sylt and Cs-clock transportation being the main methods, but now the Glonass-GPS-comparison work of professor Daly at The University of Leeds has diminished the significance of our reports in that respect.

Session POSTER Number 51

EUROPEAN COMPUTER TIME SERVICE

Erland Brännström

Swedish Telecom Radio
Swedish Telecommunications Administration (STA)

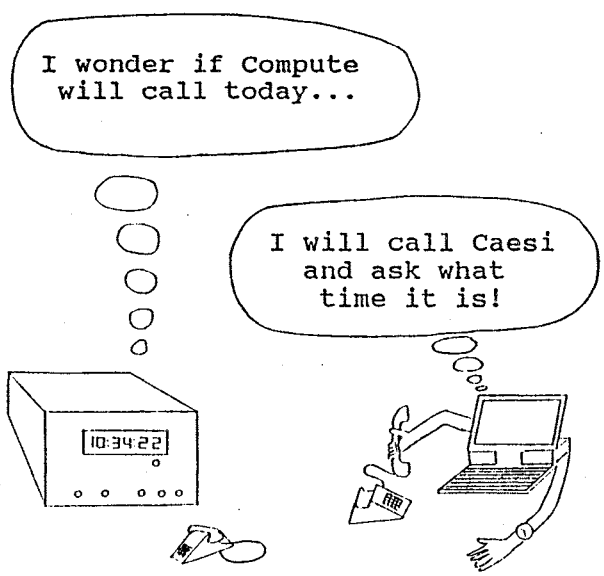
ABSTRACT

There is a growing demand for the automatic and accurate setting of computer clocks.

Four European time laboratories have been cooperating to define a common format for the dissemination of date and time by telephone.

This format is sent as a string by modems to the user. This unique string contains the most complete time and date information, even the announcement of daylight saving time changes.

Since november 1990 this format has been disseminated from a time code generator in Sweden.



INTRODUCTION

An increasing interest exists in having correct time and date in computer systems. Especially the possibility to automatically set clock and calendar is desirable.

The format described in this paper is the result of a cooperation between the following time laboratories:

- Technical University of Graz (TUG) in Austria
- Van Swinden Laboratory (VSL) in the Netherlands
- Istituto Elettrotecnico Nazionale Galileo Ferraris (IEN) in Italy
- Swedish Telecommunications Administration (STA) in Sweden

After examining the already existing formats, see references (1), (2), (3), (4) and (5), the conclusion was that none of these could meet all requirements.

THE TIME CODE FORMAT

The time code is a line of 80 characters transmitted each second. First 77 characters are transmitted. This takes about 670 ms.

Then, after a pause, the time markers "*", <CR> and <LF> are transmitted. The baud rate is 1200.

The picture on next page (Figure 1) shows the code. At the end of this paper, figure 2 shows an example of some lines received from the Swedish generator.

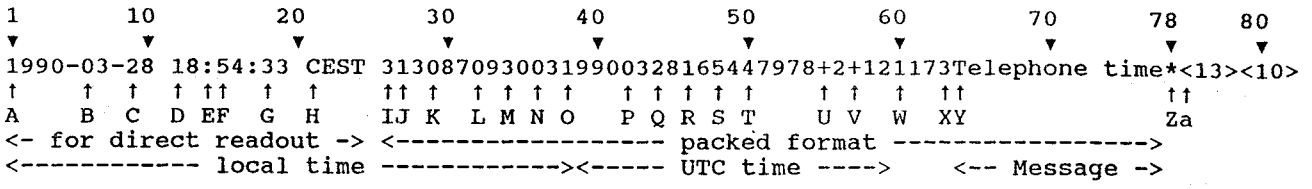


Figure 1

"▼" Marks the character position given by the numbers above the code characters.

APPLIES TO LOCAL TIME

A year
B month
C day
D hour
E normally this is ":", see OPTIONS below
F minute
G second
H local time identifier and time zone identifier
I day of the week
J week of the year
K day of the year
L month ┐
M day ├ Next change of local time from Standard time to Daylight Saving time or vice versa.
N hour ┘

APPLIES TO UTC TIME

O year
P month
Q day
R hour
S minute, see UTC-MINUTES below
T MJD, Modified Julian Date
U DUT1, multiply the number by 0.1 second to get DUT1
V announcement of a leap second

NOT TIME SCALE RELATED

W measured delay
X message sequence number
Y message string
Z visible time marker, either "*" or "#", see MARKER below
a <CR> <LF> ASCII values 13 and 10

The TIME REFERENCE is at the end of <CR>

FURTHER EXPLANATIONS

H: LOCAL TIME IDENTIFIER
This position is clear readable. It is expressed in local language. It shows if daylight saving time is in use.

S: UTC MINUTES
The UTC minutes have to be shown because some countries have a local time that differs from UTC by for instance 15, 20, 45 or 50 minutes.

U: DUT1
This is the difference between the astronomic time scale UT1 and UTC.

V: LEAP SECOND
The announcement is a sign and a number.
+ = Positive leap second (extra second)
- = Negative leap second (delete one second)
Number = The leap second occurs at the end of this month.

W: DELAY MEASUREMENT
This position may contain information about delay measurement (= letters) or may contain the measured delay (= numbers).

X: MESSAGE SEQUENCE NUMBER
This number (0 - 9) identifies the part of the message string transmitted in field Y.

Z: MARKER
This Position is a visible time marker. Normally it will be "*", but changes to "#" if delay measurement is done and the delay is compensated by advancing the code.

a: TIME REFERENCE
The time reference is the leading edge between the stop bit of CR and the start bit of LF.

OPTIONS

In position "E" is usually ":". In Austria this will change to "A" or "B", when switching from Daylight Saving Time to Standard Time is in progress.

In position "H" Sweden transmits three readable characters meaning "summer time" or "winter time". Following is a space and the difference between Local Time and UTC Time in multiples of half hours.

In position "X" Sweden transmits a space. The first message line is transmitted when the unit number of the seconds is zero.

DELAY MEASUREMENTS

Space is reserved in the code for the delay value if delay measurements will be done in the future. First estimate is that delay measurements will not improve the accuracy very much in Europe, except if the call would be linked by a satellite. This is unlikely to happen within a small European country.

In Sweden a transmission delay measurement on the telephone network gave a result of about 4.5 ms for 1000 km.

Using modems following the standard V22, the delay in the modems is much larger but is compensated for. Left is a jitter from the modems, typically about 10 ms.

Total timing accuracy from STA would normally be within ± 10 ms in Sweden without delay measurement. This is valid at the output of the receiving modem.

STA TIME CODE GENERATOR

The STA time code generator (figure 3) consists of a "hardware" clock and a computer unit. It is equipped with rechargeable batteries.

The "hardware" clock is possible to adjust in time, without interrupting the time code. It has three 5 Mhz inputs, and is operating if there is a signal on one of these. It has some alarm outputs intended to be monitored by a computer in the time laboratory.

The phone number to the Swedish time code generator is +46 8 741 0809. When calling, use 1200 baud V22, 8 bits, 1 stop bit and no parity. At the transmitting site a Hayes modem supporting V22 is used.

USER PROGRAM DEVELOPMENT - HINTS

Modems sometimes buffer characters. After connection is established, skip some lines because of buffer effects, perhaps five lines. Use 1200 baud V22 also between your computer and modem.

Use more than one line to calculate the time difference between your own clock and the received code. Use the obtained difference to set your clock. One also can average the differences to reduce effects of jitter (10 ms), if you have a very good clock.

Do not trust only one received line to set time and date. There may be a transmission error on that line. Compare the message and the one received last time. If there is a change, send the message to your printer or display it on the monitor. Make it possible to change the telephone number.

CONCLUSIONS

A computer time service exists, that has the following features:

- Common time code format will be used in several countries.
- The string gives complete time and date information.
- The need of accurate time and date in computers is met.

REFERENCES

(1) "Time Comparison Using Telephone Line", M. Aida, T. Sato and S. Yamamori, Rev. Radio Res. Lab. (Japan) 31, pp. 115-124, 1985.

(2) "A telephone Based Time Dissemination System", D. Jackson, R. J. Douglas, 18th PTTI Meeting (USA), pp. 541-552, 1986.

(3) "Genaue Zeit für Rechner über Telephonmodems", D. Kirchner, in Berichte der Informationstagung Mikroelektronik 89, Springer-Verlag, Wien New York, pp.103-108, 1989.

(4) "The NIST Digital Time Service", J. Levine, M. Weiss, D. D. Davis, D. W. Allan and D. B. Sullivan, 21st PTTI Meeting (USA), pp. 181-190, 1989.

(5) "Methods For The Transfer and Dissemination of Time and Standard Frequencies", CCIR report 363, modified 1989.


```
1991-03-31 01:59:55 VIN 2713090033102199103310059483 46+4000000 För info ring *
1991-03-31 01:59:56 VIN 2713090033102199103310059483 46+4000000 08-707 51 97  *
1991-03-31 01:59:57 VIN 2713090033102199103310059483 46+4000000                *
1991-03-31 01:59:58 VIN 2713090033102199103310059483 46+4000000                *
1991-03-31 01:59:59 VIN 2713090033102199103310059483 46+4000000                *
1991-03-31 03:00:00 SOM 4713090092903199103310010048346+4000000 Rätt tid från *
1991-03-31 03:00:01 SOM 4713090092903199103310010048346+4000000 Riksmätplatsen*
1991-03-31 03:00:02 SOM 4713090092903199103310010048346+4000000 Tid & frekvens*
1991-03-31 03:00:03 SOM 4713090092903199103310010048346+4000000 Telev. Radio  *
1991-03-31 03:00:04 SOM 4713090092903199103310010048346+4000000 136 80 Haninge*
1991-03-31 03:00:05 SOM 4713090092903199103310010048346+4000000 För info ring *
1991-03-31 03:00:06 SOM 4713090092903199103310010048346+4000000 08-707 51 97  *
```

Figure 2. An example of the code format, when using a terminal program to receive the code. One line is received each second. This example is from the Swedish time code generator, and shows a change from winter time to summer time.

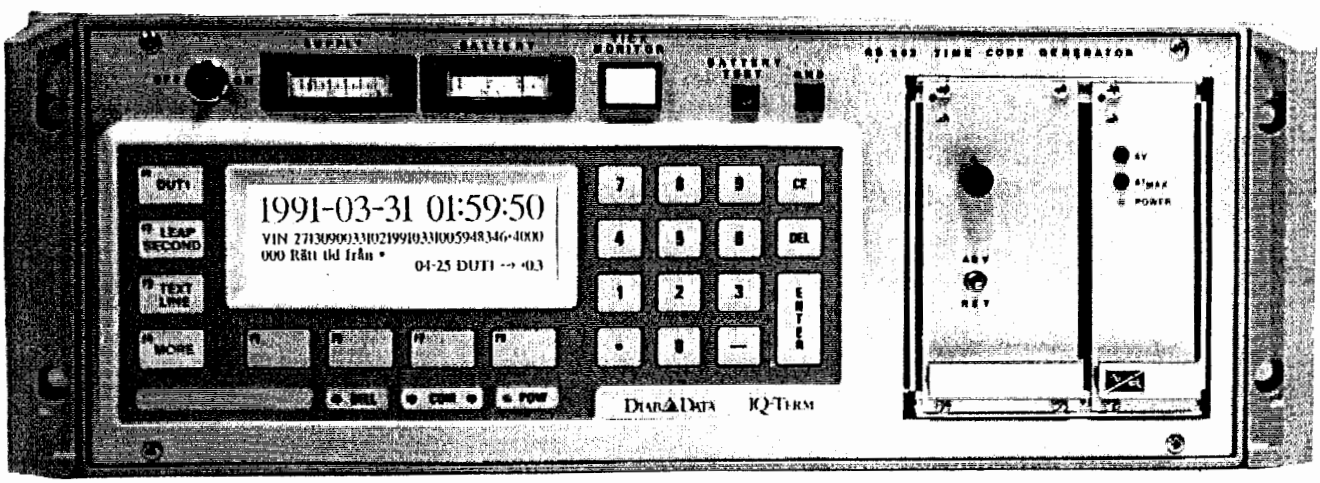


Figure 3. The Swedish time code generator.

Session POSTER Number 52

Optical Pumping of the Rubidium Maser by the Use of a Frequency-Modulated Laser

Pierre Tremblay, Alain Michaud*, Jean Beaubien, Eric Cauchon

COPL, Département de génie électrique, Université Laval, Québec, Canada, G1K 7P4
* present address: Bureau International des Poids et Mesures, Sèvres, France, F-92312

Abstract

Operation of an oscillating rubidium maser has been achieved through the use of a frequency-modulated semiconductor laser as the pumping source. Experimental measurements has shown that the output power is a function of the modulation waveform and frequency. Two theoretical approaches are used to understand this behavior: the rate equations of populations taking into account the power spectrum of the modulated laser and the density operator to calculate the time evolution of a three level atom pumped with a time dependent monochromatic laser. The results foresee the usefullness of frequency spreading when the laser power is sufficiently high. Variation of the modulation depth can be used to compensate for light intensity saturation. The influence of the modulation frequency is also predicted.

I. Introduction

The rubidium maser relies on optical pumping to achieve population inversion [1, ch. 7]. Recently, an oscillating rubidium maser pumped by a semiconductor laser has been successfully operated through the frequency modulation of the laser [2]. Experimental results have shown that spreading the laser spectrum greatly increased the maser output power. The unmodulated laser exciting only atoms of a small class of velocity, broadening of the laser spectrum ensures the contribution of a large fraction of the atoms present in the vapor [3].

In order to understand the observed phenomena, two different aspects must be considered, leading to two different models which will be presented here. The stimulated emission involved in the maser action relies on the population inversion of the atoms present in the vapor. The first approach concerns the population inversion of a vapor using rate equations of populations and taking into account the velocity distribution of the atoms and the laser spectrum. One other aspect to study when involved with optical pumping is the relaxation due to light broadening. This problem can only be solved by the use of the density operator, which gives the emitted power of a steady atom while it is submitted to a frequency-varying monochromatic laser.

II. Maser output power

The experimental setup is depicted in Fig. 1. It represents a conventionnal rubidium maser [4]. A quartz bulb containing ^{87}Rb and N_2 as the buffer gas is enclosed in a cylindrical microwave cavity operating in the TE_{021} mode at 6.835 GHz with a Q around 27 000. The cavity is surrounded by thermal and magnetic shields and by a solenoid used to create a homogeneous axial static magnetic field. The pumping light is obtained from a semiconductor laser whose nominal wavelength (780 nm) is stabilized on a linear absorption line observed in a rubidium cell (D_2-line). The laser power was 16 mW and the light was injected in the cavity in an arrangement maximizing the output power [5]. Unfortunately its polarization state was unknown due to the multiple reflections occurring inside the cavity.

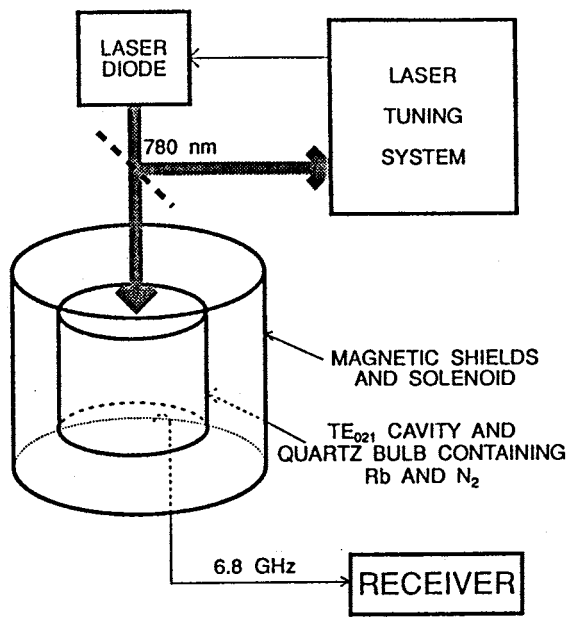


Figure 1: Experimental setup of the laser diode pumped rubidium maser.

It is not possible to obtain the self-sustained oscillation regime with an unmodulated laser. However the maser oscillates when the frequency of the pumping laser is modulated [5]. Various periodic modulation waveforms have been tried, such as sinusoidal, triangular and square. Filtered Gaussian noise was also tried. Figure 2 illustrates the typical behavior of the maser output power when the modulation depth is varied. The maser does not oscillate at low modulation depth while it ceases to oscillate when the laser frequency is swept on a too large interval. It thus showed that there always exists a modulation depth optimizing the maser power. Similar results were obtained for different light intensities.

The maser also exhibits a peculiar behavior regarding the speed of the frequency modulation. The effect of the modulation frequency (or the noise bandwidth) is given in Fig. 3. Each point of the curves has been measured with the modulation depth set at optimum. This shows the evidence of a threshold frequency under which no oscillation is possible. This clearly demonstrates that the modulation must be sufficiently fast in order to obtain a self-oscillating maser.

III. Populations approach

The first approach to understand the measured behavior is based on the rate equations of populations in an atomic vapor and takes into account the calculated spectrum of the modulated laser. This approach will give information on the population inversion achievable in a vapor optically pumped with a frequency-

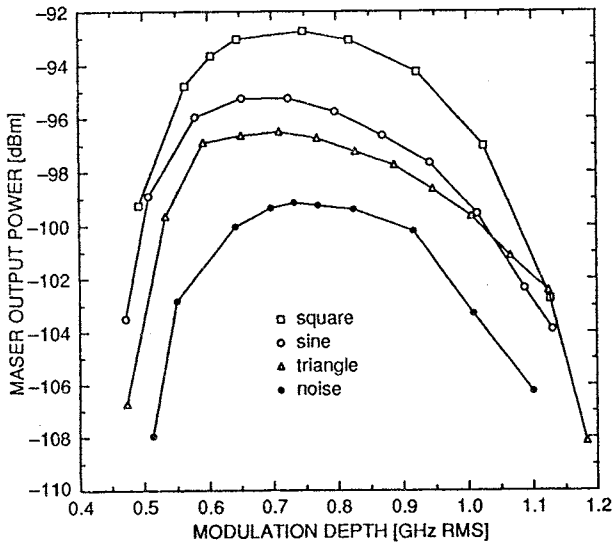


Figure 2: Measured output power as a function of the modulation depth for various modulation waveforms.

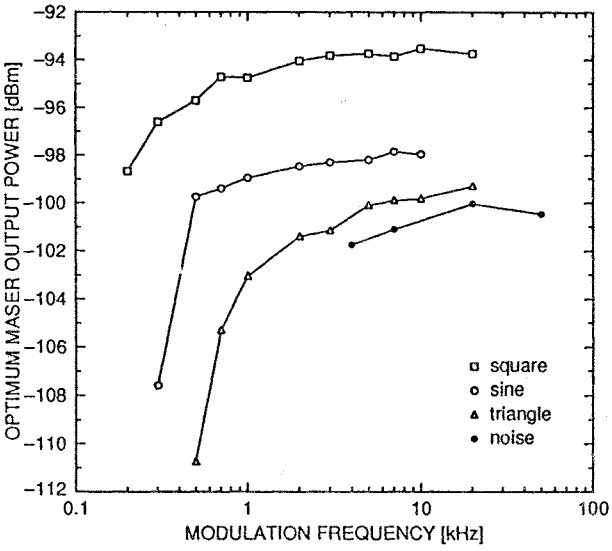


Figure 3: Measured output power as a function of the modulation frequency (or the noise bandwidth).

modulated source, the population inversion being the first necessary condition to obtain the maser action.

A. Model

In this model three-level atoms present in a vapor are studied through the rate equations of their populations. An overview of the parameters considered is given in Fig. 4. Spontaneous (A_{ei}) and stimulated (W_{ei}) emission, absorption (W_{ie}) and ground state relaxation (γ_1) are considered.

The stimulated emission and the absorption rates are related to the power spectral density of the pumping source [6, ch. 13]. We consider a laser with a Lorentzian spectrum, whose frequency is modulated. Figure 5 illustrates the shape of the laser spectrum for the various modulation waveforms considered. For low modulation frequencies, the laser spectrum is independent of the modulation frequency. It is closely related to the probability density function of the modulation signal, which is the standard wideband FM case.

Finally the ground state population inversion obtained for a given atom is averaged over all possible atomic velocities which results in the mean population inversion.

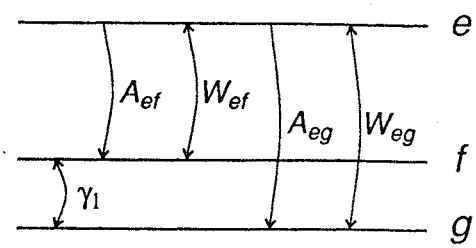


Figure 4: Interaction scheme considered for the populations approach.

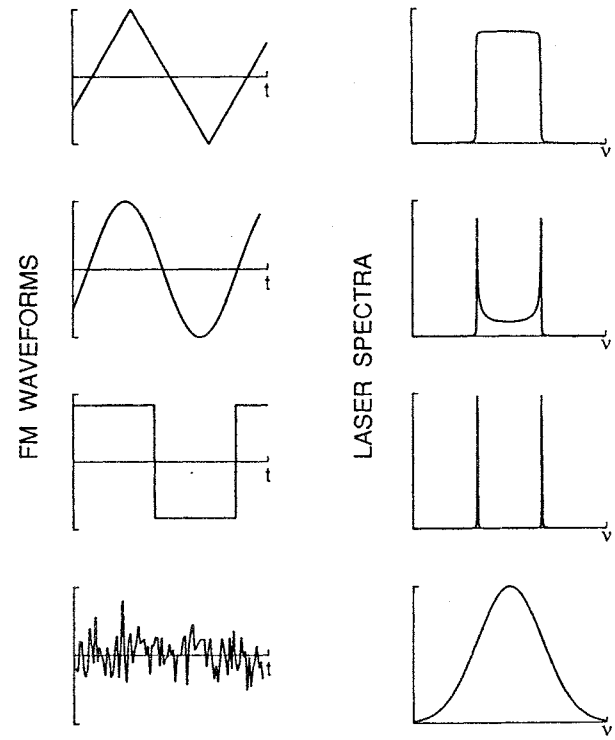


Figure 5: Modulation waveforms considered and the resulting laser spectra.

B. Results

We numerically evaluate the mean population inversion of the atomic vapor pumped by the frequency-modulated laser field. As an example we present the variation of the mean population inversion when the laser is detuned (Fig. 6). It shows that the maximum population inversion quickly increases as the modulation depth grows since the laser interacts with a larger number of atoms. However if the modulation depth becomes too large the maximum population inversion vanishes due to the spreading of the laser intensity over a wider spectrum than the Doppler one.

Figure 7 illustrates the behavior of the maximum population inversion when the modulation depth is varied. For each type of modulation, an optimum modulation depth appears, as it was observed experimentally (Fig. 2). The modulation depth optimizing the population inversion depends on the laser intensity, as it is shown in Fig. 8. Accordingly the maximum population inversion increases with the laser intensity.

These results foresee the usefullness of the frequency spreading when sufficiently high power is available. This model can not consider the hyperfine coherence responsible of the maser action, so it fails to predict the observed effect of the modulation frequency and waveform.

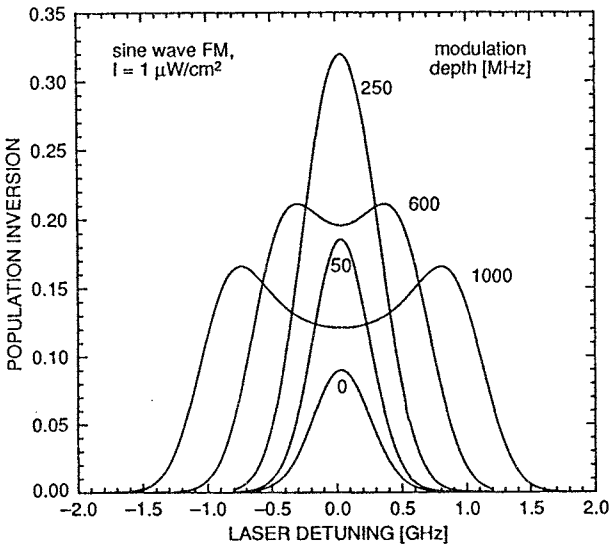


Figure 6: Mean population inversion as a function of the laser detuning for various modulation depths.

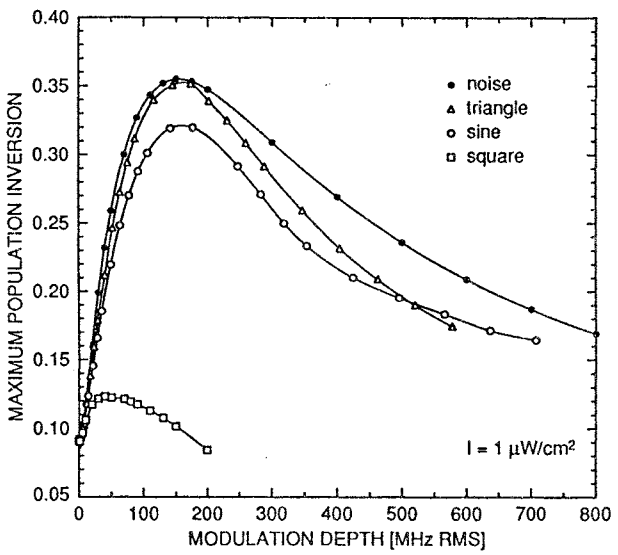


Figure 7: Maximum population inversion as a function of the modulation depth for various modulation waveforms.

IV. Density operator approach

The emitted power in the cavity is related to the hyperfine coherence, which in turn depends on the population inversion. This could only be calculated through the density operator approach [1].

A. Model

A three-level atom is interacting with a time-varying monochromatic laser and a monochromatic microwave magnetic field (as represented in Fig. 9). The atom is interacting simultaneously with two electromagnetic fields, namely the laser and the microwave fields. The model includes the electric and magnetic dipole perturbations (V_{eg}, V_{fg}), and the relaxation of the ground state populations and coherence (γ_1, γ_2). This model differs slightly from that given in ref. [1, §4.3.2] for monochromatic perturbations in a sense that we neglect the direct excitation of the $f \leftrightarrow e$ transition by the laser.

The laser frequency is modulated and the time evolution of the laser field is evaluated. The time evolution of the atomic populations and coherences is then obtained. Within this model

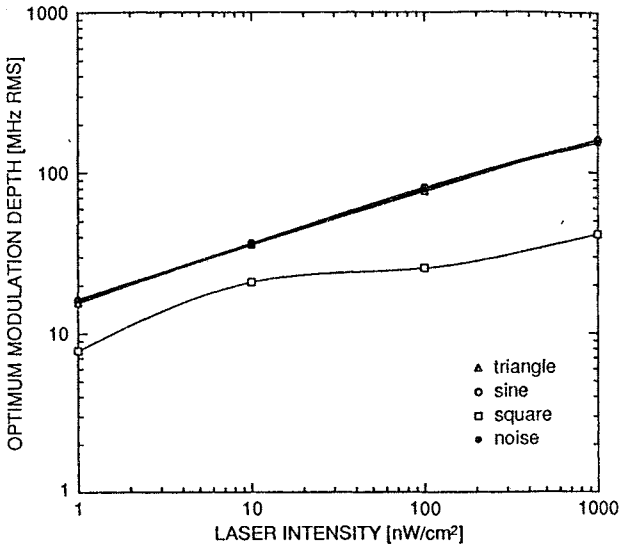


Figure 8: Optimum modulation depth as a function of the laser intensity.

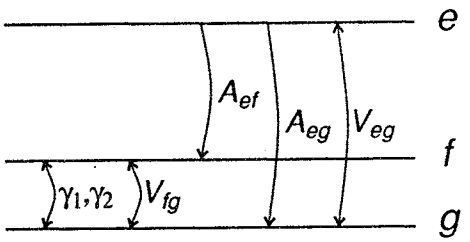


Figure 9: Interaction scheme considered for the density operator approach.

we used an ensemble of atoms having the same velocity, since many important conlusions can be drawn without considering the velocity distribution.

B. Results

Typical time evolution of population inversion and stimulated emission (proportionnal to the hyperfine coherence component in quadrature with the microwave field) is presented in Fig. 10. It is given for a sine wave modulation. The interval between the two dashed lines represents half of a cycle of the modulation waveform. When the laser frequency passes through the transition, the population inversion rapidly increases since the pumping effect is maximum. At the same time the hyperfine coherence decreases due to the light broadening. This example clearly illustrates the competition between population inversion and hyperfine coherence. No matter how large is the population inversion, if the light broadening is too important the emitted power will vanish.

In the following, we will present results obtained in the stationary state for a typical magnetic field intensity (maser power of 10^{-9} W). The emitted microwave power, which is proportionnal to the stimulated emission, will be averaged over one cycle of the modulation waveform.

Figure 11 shows the dependence of the mean emitted power upon the modulation frequency, for different modulation depths. Again we see an increase in the emitted power with the modulation depth. Meanwhile it does not result from the velocity distribution as for the previous model, but it is a characteristic of each atom. The emitted power rapidly drops when the modulation is not sufficiently fast, as observed experimentally (Fig. 3). This phenomenon is due to the disappearance of the constructed hyperfine coherence over time intervals of the order of the transversal relaxation rate.

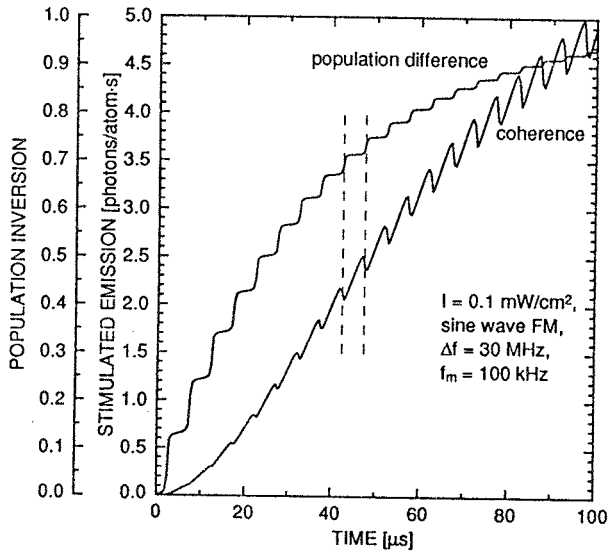


Figure 10: Typical time evolution of population inversion and stimulated emission for a sine wave frequency modulation.

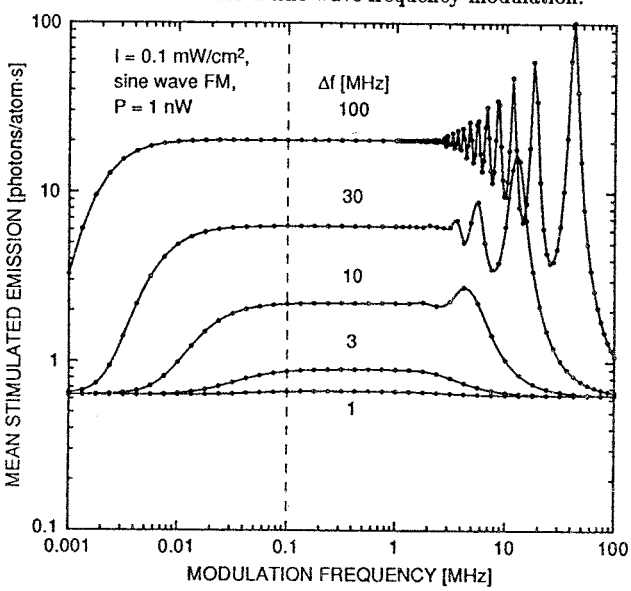


Figure 11: Mean stimulated emission as a function of the modulation frequency for various modulation depths.

At very high modulation frequencies the emitted power approaches the unmodulated value since the maximum phase excursion seen by the atom vanishes as the modulation frequency is increased, thus leading to a monochromatic pumping source. An interesting interference effect is predicted for modulation frequencies around 10 MHz. It is related to the phase shift appearing in the optical coherence, $\rho_{eg}(t)$, which does not follow instantaneously the exciting field.

In order to compare the different modulation waveforms, we will consider a given modulation frequency corresponding to the plateau obtained for all modulation depths. It is illustrated by a dashed line in Fig. 11. The variation of the emitted power with the modulation depth is given in Fig. 12 for the different modulation waveforms. It shows that the square wave seems to be the most efficient way to maximize the emitted power since it requires a smaller modulation depth. This behavior was observed for a large variety of conditions and it corresponds to the experiment.

It has not been possible to evaluate the total emitted power in a vapor, by averaging over all possible velocities, due to the too large amount of computer time required. However, in order to estimate the emitted power of the entire vapor, we calculated

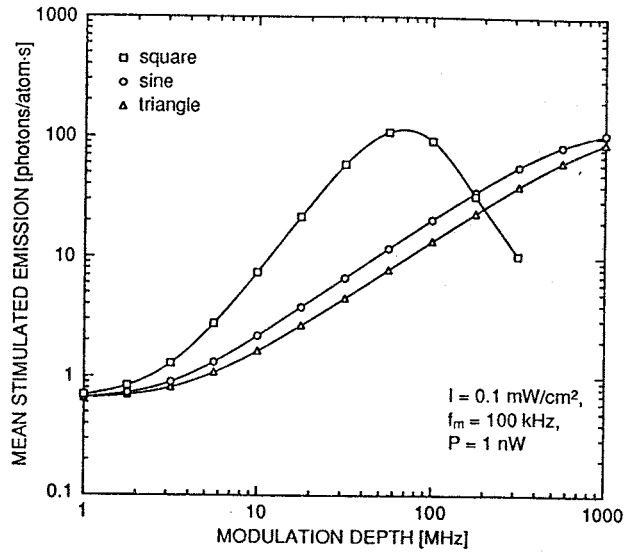


Figure 12: Mean stimulated emission as a function of the modulation depth for various modulation waveforms.

the variation of the atomic emitted power with the laser detuning for different intensities (Fig. 13). For each intensity the optimum modulation depth for the laser tuned was chosen. The continuous curve represents the best monochromatic condition, which occurs for fairly low intensities. The total contribution of the atoms in the vapor will increase as the intensity is increased. The Doppler spreading of the atomic resonance frequencies in the vapor has a width around 500 MHz, so the contribution of the wings of the curves presented in Fig. 13 plays an essential role. There seems to be an optimum intensity for which the total emitted power will be maximized. At higher intensities the behavior approaches the results obtained for the monochromatic case.

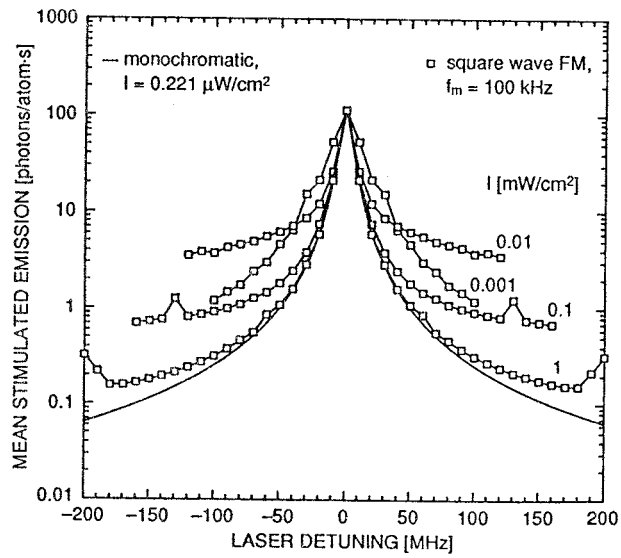


Figure 13: Mean stimulated emission as a function of the laser detuning for various laser intensities, the optimum modulation depth being used for each case.

The efficiency of the frequency-modulated laser to produce large emitted power is then not related to an increase in the maximum power for a given atom, but to the possibility of creating large hyperfine coherence for atoms distributed over a wider velocity range.

V. Conclusions

The rate equations of populations have shown that a wide laser increases the population inversion achievable in an atomic vapor. It results in an optimum modulation depth, which depends on the modulation waveform and which increases with the laser intensity.

Moreover the density operator approach reveals a better understanding of the maser action. Not only it shows the optimum modulation depth, but it also reveals the effect of the modulation frequency. It predicts a threshold frequency and an oscillatory behavior at high frequencies.

In conclusion frequency modulation of the laser enables high optical pumping efficiency while maintaining low relaxation due to light broadening.

Acknowledgement

The authors wish to thank colleagues from COPL for their help. This work has been done under the financial support of the C.R.S.N.G. (Ottawa, Canada) and Fonds F.C.A.R. (Québec, Canada).

References

[1] J. Vanier and C. Audoin, *The quantum physics of atomic frequency standards*, Adam-Hilger, Bristol, 1989.

[2] A. Michaud, P. Tremblay, M. Têtu, "Experimental study of the laser diode pumped rubidium maser", *IEEE Trans. Instrum. Meas.*, vol. IM-40(2), april 1991.

[3] M. Elbel, C. Larat, P.J. Nacher, M. Leduc, "Optical pumping of helium-3 with a frequency electromodulated laser", *J. Phys. France*, vol. 51(1), pp. 39-46, january 1990.

[4] M. Têtu, G. Busca, J. Vanier, "Short-term frequency stability of the ^{87}Rb maser", *IEEE Trans. Instrum. Meas.*, vol. IM-22(3), september 1973.

[5] A. Michaud, *Réalisation d'un maser à rubidium pompé optiquement à l'aide d'un laser à semi-conducteur*, Ph.D. thesis, Université Laval, Québec, Canada, 1991.

[6] C. Cohen-Tannoudji, B. Diu, F. Laloë, *Mécanique quantique*, Paris, Hermann, 1977.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 53

Portable atomic clock on the basis of an active hydrogen maser,"Sapphir".

B.A.Gaygerov, F.S.Rusin, V.P.Sysoev.

NPO VNIIFTRI, Mendeleevo, Moscow region, 141570, USSR

Application of an active hydrogen masers as a portable atomic clock is limited by their large size and considerable weight due to cylindrical microvawe cavities working on oscillation H_{011}. Such a cavity has relatively big sizes: its diameter and height are near to 300mm at a hydrogen maser frequency.

One of the ways to significantly reduce the size of the hydrogen generator is the use of a resonant cavity partially filled with a high permittivity and low losses dielectric. The cavity may be of various shapes but for using in the hydrogen maser it must have oscillations with closed electrical lines. The filling area of high frequency electric field not joined to cavity walls may considerably reduce the resonant system sizes without decreasing the quality factor. Moreover its some enhancement compared with the traditional hollow cavity sometimes may be obtained. The hydrogen maser of such a type was proposed and developed in VNIIFTRI, the USSR [2,3]. The cylindrical microvawe cavity of a diameter and height 160mm and 180mm respectively was used. Inside the cavity there was placed a hollow sapphire cylinder with an inner surface coated with Teflon, which simultaneously plays a role of a storage bulb (Fig.1). Its outer diameter is 95mm, inner - 81.9, the storage bulb value 790ml, the input channel diameter - 4mm, its length - 50mm.The cavity quality factor is about 50,000 and it's sufficient for stable operation of the maser in an active mode. The temperature coefficient of the resonant frequency is determined by temperature properties of sapphire and is about $5 \cdot 10^{-5}$. This is considerably worse than the corresponding characteristics of properly designed large size cavities. To provide the temperature drift of the hydrogen generator frequency due to cavity frequency offset less than $1 \cdot 10^{-14}$ it's necessary to maintain the cavity temperature with an accuracy better than $2 \cdot 10^{-6}$K. Such a requirements cannot be satisfied in practice. The resonant frequency may be maintained by means of a stabilization servo system. Under such a condition the requirements to temperature maintaining can be significantly reduced.

The resonant cavity frequency stabilization servo system uses the amplitude criteria with a "fast modulation" method, developed in VNIIFTRI [4]. The cavity frequency is modulated by a symmetrical rectangular signal with a frequency essentially greater than a spectral line width. When the cavity is tuned out from the line of hydrogen this modulation causes an amplitude modulation of the output signal of the generator. The servo system minimizes this modulation by tuning the resonant cavity. In the given design the resonant cavity frequency modulated by means of a switching diode closing and opening the loop placed onto the cavity top. The depth of the modulation was chosen of the order of 8kHz.

For the spin exchange shift compensation several methods were tested: by additional quality factor modulation concurrently with the frequency modulation; by modulation the RF amplifier gain using nonsymmetrical rectangular pulses modulating the cavity resonant frequency. The last of the listed methods is the most convenient as it allows to easy control the value of the offset bringing in.

The hydrogen generator resonant cavity is placed into a thermostat and surrounded by four magnetic shields of total a coefficient of the order of 10^5. To reduce the interference of a magnetic field nonuniformity to the maser frequency the periodic inversion of populations at levels $m_f = \pm 1$ is being performed by means of a special wire coil. For pumping off a hydrogen the porous titan is used and in addition to the vacuum system has an ion pump of productivity about 0.15 litr/sec.

The portable hydrogen frequency standard was being investigated at the USSR Main Metrological Center of the State Time and Frequency Service continuously during more than three years. The relative mean square value of a frequency variation $2 \cdot 10^{-14}$ at measuring time intervals $10^3 \div 10^5$ was obtained. The comparisons between the USSR National Time and Frequency Standard in Moscow and remote standards situated in Irkutsk, Charkov, Havana, Paris, London were fulfilled. The comparison error was estimated by value of the order of 10 ns/day when an air or railway transport is used.

The hydrogen standard developed has dimensions 550 x 480 x 300mm, weight is 62 kg with the built-in battery and 50 kg without it. The power requirements are 220V AC, 127V AC, 27V DC or 12V DC. The builtin battery provides independent continuous operation during 5 hours. The standard provides output signals 100MHz, 5MHz and 1Hz.

Fig.2 presents general view of the developed hydrogen atomic standard.

Literature.

1. F. S. Rusin, B. A. Gaygerov, S. F. Ponomarenko. "Calculation of the parameters of the cylindrical microvawe cavity partially filled with the dielectric". "Izmeritelnaya tehnika", USSR, 1985, N5, p. 15-16.

2. B. A. Gaygerov. "Hydrogen generator with a small size microwave cavity". "Izmeritelnaya tehnika", USSR, 1986, N6, p. 8-9.

3. B. A. Gaygerov, F. S. Rusin, N. I. Polnikov, V. P. Sysoev, S. N. Ovchinnikov. "Portable atomic clock based on the hydrogen maser with a small size microwave cavity". "Izmeritelnaya tehnika", USSR, 1989, N4, p. 17-18.

4. B. A. Gaygerov, G. A. Yelkin. "Automatic tuning the hydrogen frequency standard". "Izmeritelnaya tehnika", USSR, 1968, N6, p. 90.

Figure captures.

Fig. 1. View of the microwave cavity and sapphire bulb.

Fig. 2. General view of the hydrogen atomic clock.

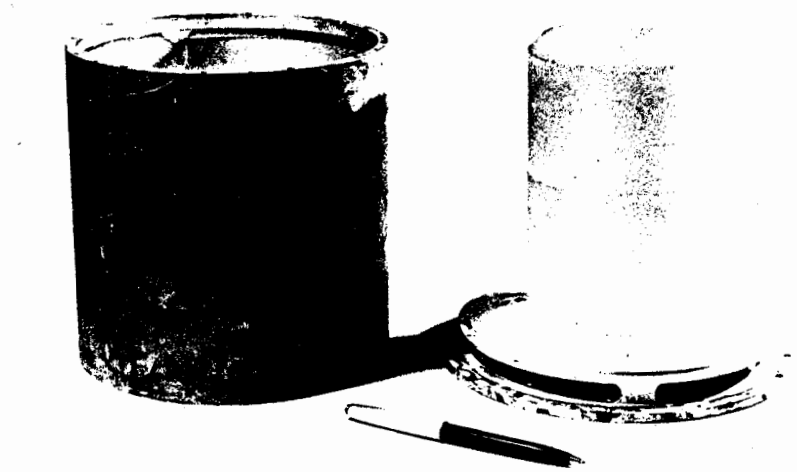


Figure 1.

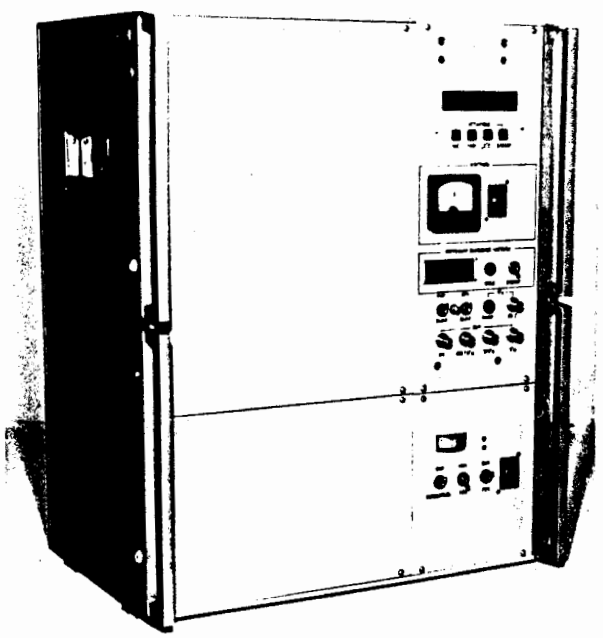


Figure 2.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session POSTER Number 54

Hydrogen Time and Frequency Standard "Sphere".

A. I. Boyko, B. A. Gaygerov, G. A. Yelkin, N. D. Jestkova, G. P. Kurnikov, V. P. Sysoev.

NPO "VNIIFTRI", Mendeleevo, Moscow region, 191570, USSR.

The frequency standard "Sphere" employs the usual quartz crystal oscillator phase locked to an active H-maser. The preamplified 20.405751 MHz signal is being supplied to the 1-st IF amplifier then transformed by the 2-nd mixer into the 2-nd IF frequency 405.751 kHz. Then it is fed to the phase detector (PD). The signal from the synthesizer is supplied to another PD input. The PD output drives the VCO's varactor. The standard's output signals are 1V RMS 5MHz signal and TTL compatible 1 PPS. An overall view of the hydrogen maser is shown at Fig.1,2; the views of its separate units - at Fig.3.

The striking feature of the hydrogen maser is as following: a microwave cavity, a thermostat, an upper vacuum chamber and magnetic shields are of the spherical shape. The metallic semispheres are made of sheet materials by hydroexplosive stamping technique and the cavity quartz semispheres - using gas deposition method. Such a design satisfies high demands for a purity, size accuracy and cavity repeatability.

The molecular hydrogen accumulator is made on the basis of the intermetallic $LaNi_5$ combination put into a cylindrical container of about $50cm^3$, the leakage is made of a thin-wall (0.05 mm) nickel pipe twined into a spiral. An atomic hydrogen source is a small spherical quartz bulb 30mm in a diameter. The incoming flux of hydrogen atoms goes through the channel bored in the crystal flange. The output hole of the bulb has a multichannel collimator. The lower vacuum chamber is pumped out by means of a getter pump (titanium absorbers) and an ion pump of 0.2 l/s. The storage bulb is made of fused quartz and has a diameter 130mm, the input channel has a diameter 5mm and a length 50mm. The inner surface of bulb is coated with a Teflon F10. Two quartz suspenders fix the bulb is fixed into the center of m.v.cavity. The cavity is a quartz sphere of 280 mm diameter which outer surface is coated with a silver film.

The cavity temperature frequency coefficient of is $2 \cdot 10^{-7}$ 1/K/. The one stage thermostat provides thermal stability of output maser frequency better than $5 \cdot 10^{-14}$ 1/K. The Q-factor is 50,000.

Three spherical magnetic shields placed into the upper vacuum chamber and the external cylindrical shield playing a role of a device case provide the coefficient of shielding about 15,000 (see Fig.1).

To addition to there was a magnetic field stabilization servo system improving the coefficient of shielding 20-50 times. The upper vacuum system is a titanium envelope 360mm in a diameter consisting of 2 semispheres vacuum-tightly joined with an indium seal and mechanical clamping of the flanges. The upper chamber is evacuated by means of two ion pumps each of productivity 0.2 l/s.

The dimensions of the standard are 750 height and 400mm diameter. The weight is about 60 kg, power requirements: 27V DC, 35 Wt. The internal battery for 8-hour standby operation is available.

The Q-factor of a radiation line is about $2 \cdot 10^9$ for low beam intensity and $0.6 \cdot 10^9$ - for high one.

Frequency instability (Allan's variance) not exceed $1 \cdot 10^{-14}$ for 1000s and $3 \cdot 10^{-14}$ for 100 s and 1 day, under condition environment temperature maintenance in the $\pm 1^{\circ}C$ limits.

The frequency drift does not exceed $2 \cdot 10^{-14}$ per day. The longterm characteristics of H-maser may be extracted from data presented at Fig.4, where one can see the results the time scale comparison between UTC(SU) and H-maser "Sphere" via TV-link for the period from July 1990 to January 1991.

Figure Captures.

Fig.1. General view of the Hydrogen Time and Frequency Standard "Sphere".

Fig.2. General view of standard without external shield and battery.

Fig.3. Storage bulb, microwave cavity, hexapole magnet, RF dissociator, purifier, hydrogen H-accumulator, ion pump.

Fig.4. Results of the comparisons between the standard "Sphere" and National Time and Frequency Standard via TV link from July 1990 to January 1991.

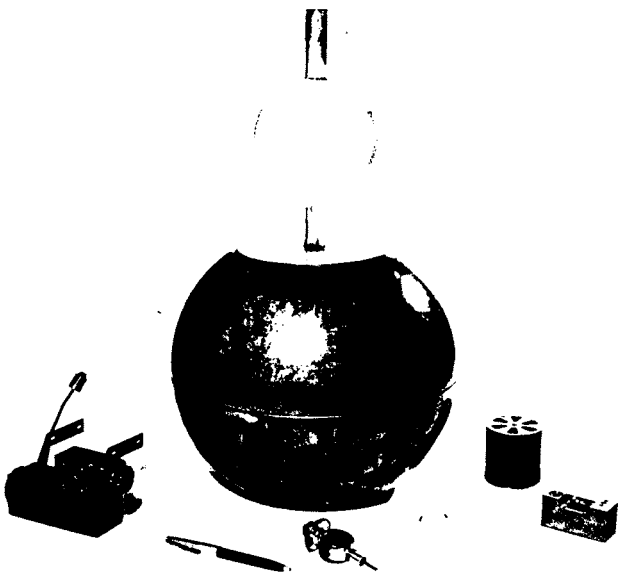


Fig.3



Fig.1

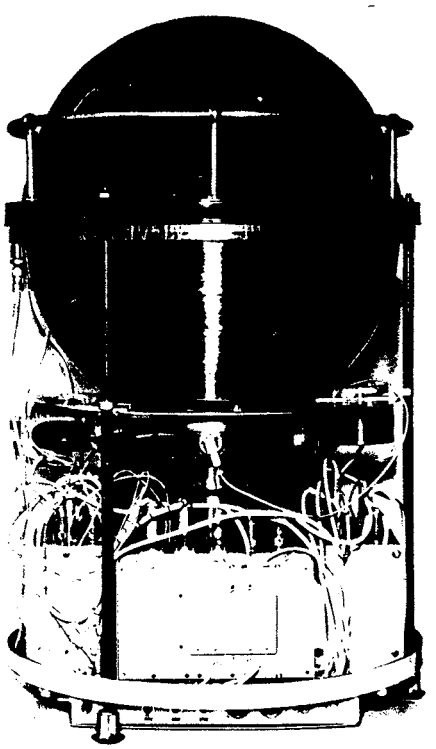


Fig.2

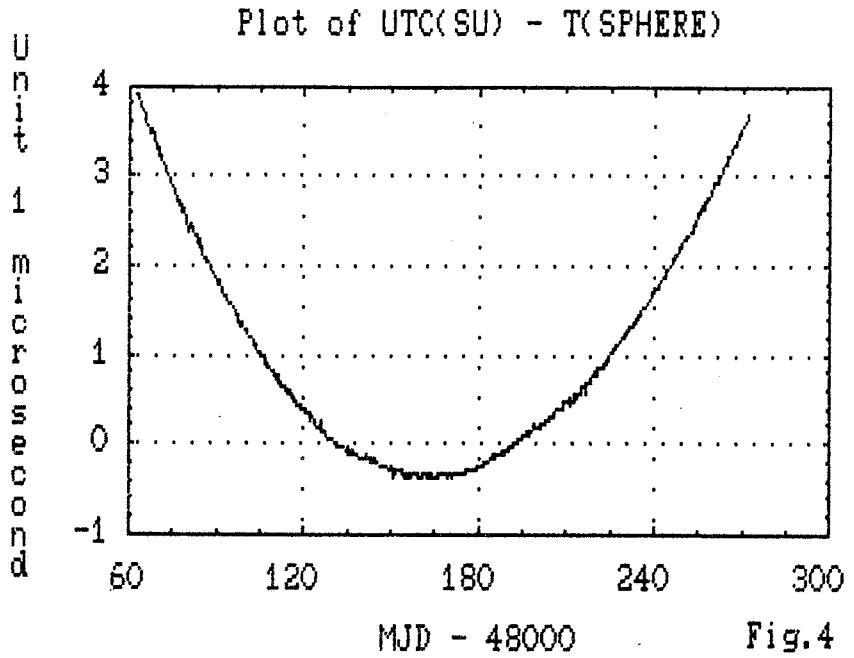

Plot of UTC(SU) - T(SPHERE)
Unit 1 microsecond
4
3
2
1
0
-1
60
120
180
240
300
MJD - 48000

Fig.4

Session POSTER Number 55

STATE OF THE ART POLISHING AND LAPPING FOR VHF QUARTZ BLANKS

BY LECH MORAWIEC

QUARTZ WORKS USA

Initial considerations which should be given regarding Quartz Crystal Blank processing are the control of the angle, final frequency, and surface finishes. Blank geometry is paramount when selecting lapping or polishing equipment. New technics for innovating VHF processing will be the combination of 4-Way lapping and chemical milling. I would also like to describe some trouble shooting situations and what to be observant of.

Ever since the mid 1970's, Frequency Symposiums and Crystal Conferences have been dealing with the production of high frequency blanks in a high volume enviroment.

Being part of the crystal industry for the past nineteen years in the U.S., I have participated very actively in search of these goals trying to learn from all possible sources and develop my own concept of cost, equipment and processes necessary to reproduce such crystal units. This aim has also aided me in adopting some useful innovations from semiconductor and other related industries by studying the mechanics chemical milling and quartz surface behavior under different extreme conditions of manufacturing.

All of these experiments allowed me to define perhaps not the ultimate but certainly workable solutions that can be applied in even a small company production environment without expensive equipment and tooling.

The first and most important factor in building high frequency crystal units is the quartz bar selection. A lot of unsuccessful attempts by various manufacturers are directly related to overlooking this simple fact by trying to use an average quality quartz blank without a closer examination of the original source of the bar and its growing conditions.

Over the past years, I have taken a very closer look at the different quartz growing operations and have noticed various quartz quality coming out of different autoclaves indicating relatively poor process control for even premium Q slow-grown bars. Some of the bars exhibited stresses even visible to the naked eye in the form of fine microfractures which eliminate this kind of material for any further use.

Apparently reusing the same seed material several times does not improve grown bars but obviously degrades the quality of the quartz. Quite a costly lesson to learn.

After careful microscpic examination of the quartz bar an extra insurance would be highly recommended for a successful completions of the tasks. This extra step is a thorough sweeping of the bars to further improve the crystalline structure of the bars. D-C current drop and its steady level is a good way to judge that the process has been carried out properly and was completed.

In various cases, I have noticed that motion is confused with action and a turn-over in operators is blamed for defective results. Like every step of crystal processing, one of the most demanding is total process control with all possible safeguards in place. Mistakes made in this area can be costly.

Standard bar slicing and manufacturing processing are well known to all but one very important fact to be pointed out is that premium Q quartz is very succeptable to electrical twinning. Therefore rounding of the blanks should be done slowly and by expert employees. However, twinning will not effect the final product for some applications. Yet, in others which I will mention later, it will mean useless blanks. One has to keep in mind that quartz α and β even though they may not behave differently in inital lapping, their etch rate is totally different in producing unwanted blank geometry.

Since the most desirable of the majority of manufacturers is to work with "A-T" cut quartz crystal units, this paper will deal with methods and chemicals used to build such crystals. One has to note that other cuts will require different manufacturing approaches available at the specific customer's request. Basic blank preparation is identical to any frequency built already by every crystal operation. Observing a safe "cut-off" of greater grit size and giving enough room for total subsurface damage removal is the key to certain percentage of greater particles than the one marked by the supplier of the abrasives.

(See Fig 1)

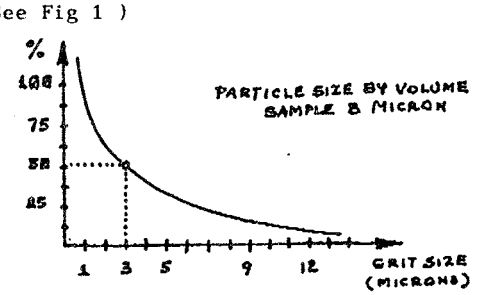


FIG. 1.

If this is the case, it is good to remember that every step of lapping is inevitably connected with dicing of lapped blanks and no matter how fine an abrasive that we use the depth of the scratches will be the only variable factor. Everyone knows about the adverse effects on electrical parameters of a scratched surface on inharmonic responses (which are critical for crystal filter applications) and resistance degradation, not to mention frequent breakage in further processing. This is why one should be careful in choosing an abrasive vendor and check the track record with other manufacturers for shortcomings. The price should be a secondary consideration because the gain in yields will compensate for the cost of the necessary supplies.

It is a good practice to thoroughly clean the blanks in between using the different grit sizes. A light etch will remove all leftover larger particles and microfractured quartz chips which will contribute frequently to scratches on the surface of the balnks.

Once all precautions are taken under consideration, one should be able to produce high quality crystal blanks no matter what the target frequency is.

High frequency has a different meaning depending on the specific goals of each operation. Oscillator houses building crystal units in the range of 30 to 60 Mhz will have more relaxed requirements for its crystal than those of filter houses working in the range of 100 to 150 Mhz with very stringent inharmonic specs and a "Q" of units making or breaking a whole design concept. Since the last kind is much more difficult to build, we will concentrate solely on these very crystals, giving the most practical way to avoid common problems. It is true that blank preparation in an initial stage does not vary from any special technique that we should already know. It is common knowledge that a good crystal unit must start with a nearly perfect blank.

Three years ago during a min-session sponsored by Crystal Components Inc. at the Kansas City Conference, I described several different ways to lap and polish high frequency fundamentals up to 50 to 60 Mhz range. Today, after experimenting with various methods, my recipe would have to be different for assuring repeatability and very consistant results on the production basis. the modern tools used presently have to combine mechanical lapping and chemical crystal processing providing the quartz blanks with exceptional mechanical strength and being able to resist to all of the disadvantages of being very thin wafers.

For various manufacturers lapping blanks up to 40 Mhz on fundamental mode by using "pin-laps" with ring supported stretched mylar carriers is nothing new. This technique is also used to polish a crystal blank. One definite minus in this type of processing is blank geometry, mostly bi-convexed in its nature. This bi-convexity can be easily identified with the use of a spectrum analyzer, or network analyzer by demostrating two strong inharmonic responses.

Bi-convex geometry is useful for some specific applications, however very useless for all others.

The production equipment which is gaining more and more recognition for their merits are the 4-Way laps which will deliver consistent parallel blank geometry and as it has proven itself time after time again to hold the geometry of the blank all the way through the lapping process. The final angle examination shows very constant over the temperature performance and no significant angular shifts. One of the disadvantages is a limited output per machine, but as we learn of various manufacturers, including CCI, who is one of the pioneers in 4-Way lapping, there are newer larger machines which allows one to fill the gap between intermediate lapping and blank finishing.

Returning to high frequency blank processing, there are basically two ways to produce such blanks:

One in the range of 40 to 70 Mhz will employ the flat blank geometry.

Second way will be in the 70 to 180 Mhz range by inverted mesa geometry. A common denominator for both ways is a fully polished quartz crystal free from scratches and other visual defects, such as pits, microfractures and chips. As we all know, it is not an easy task to produce a polished scratch-free surfaces. this is where we utilize the chemical processes. Initial polishing used by most as the only process has to be divided into a two-step operation.

First, initial polishing removes most of the surface defects, however at the same time will mask some of the fine scratches by imbedding some quartz particles also making them "invisible" at the same time. These imbedded particles can be removed by a light etching process or by the use of fluoroboric acid bath which can easily remove not only isomorphic layers of quartz but at the same time get rid of left over cerium oxide from the polish blanks.

We all know the light etch of polished blanks will expose a true picture of the polished surface and is frequently used as an examination tool for surface quality. Very rarely a full polish can be accomplished by using a single step, which will have another disadvantage e.g. a change in blank geometry and by creating frequency spread beyond practical limits.

Prepolish and etch is the only assurance that all loose quartz particles and previous grit particles are totally removed from the blank surface and a second polish combined with random transpositioning of the blanks will produce the highest quality of polished surfaces.

The desired geometry of the blanks can be obtained by the use of different hardness of the padding material, or P.H. adjustment of the polishing slurry, or both. The second light etch will allow you to examine the blank quality and selection of defect free pieces for further processing. The examination should be performed with the use of a stereo-microscope with minimum 15x magnification.

So far we have not mentioned frequency of the polished material. This decision shall be made at the individual company level. Each manufacturer has its own limits or comfort range for lapping and polishing, controlling flatness and QC on a consistent basis. It is my recommendation that for the time being we will stay with the of thickness we can reproduce and where the thickness of a used carrier does not interfere with the thickness of the blanks. One should remember that all carriers during lapping and polishing processes are under certain stresses and their warpage exhibit apparent thickness increase to about plus 25 to 40 % of their measured thickness.

Another practical consideration is a proper blank diameter choice for the high frequency process, not more than 5mm for a flat parallel geometry in the 40 to 70 Mhz range. Because the larger diameters are more difficult to handle and are more succeptable to external mount stresses and will exhibit apparent angle shifts.

It is extremely important when dealing with etch, even a light one, between lapping and especially in polishing do make sure that all blanks are thoroughly clean. Agitation is used during the etch process to prevent a preferential surface exposure. Once the damage is present, it will take a lot of effort to "undo" what has been done and quite often impossible without an additional lapping step.

Carriers used for the etch should be of a high quality to assure an even exposure to all the blanks and have a free flow of the etch solution throughout the carrier. During the etch process, all the blanks should move freely in the individual pockets and not locked-in to a tight spot. Typically, blanks used for the high frequency fundamentals application should be polished between 25 to 28 Mhz on a 4-way polisher and then exposed to the deep etch process.

Over the past several years, the technical publications were advertising various deep etching techniques and were offering a variety of equipment and means to accomplish the task. Some noted are plasma etching which have shown some promise but after tested by the author have exhibited surface deterioration due to the quite rapid quartz sputtering degrading "Q" of manufactured devices. NaOH (sodium hydroxide) used by other at very high temperatures will make the process very critical and does not produce consistantly good results with a superior "Q" crystal units. Employed for the past few years with some degree of success it has reproduced a high frequency fundamentals to satisfy most requirements for oscillator use but definitely not good enough for a quality filter application.

After trying most present methods, I have decided to go back to well known saturated ammonium bifluoride whose behavior and handling is relatively easy and handling is relatively easy and processes exist in almost every crystal operation making it readily available. demonstrated

There are some misconceptions about the etching of polished blanks which are drawbacks by not understanding a simple principle and the use of chemical analysis in this simple process.

First, most users will state that the ammonium bifluoride etch will degrade a polished surface making it at the beginning cloudy, foggy and as we follow etching the surface and "Q" of the blanks become progressively worse. This is true. Let us see what is happening but step by step. Most of us using etch do not try to thoroughly wet blank surfaces. And even if we did wet all blanks, the exposed temperature at 60 to 70 degree C will heat up in different times and depending on the degree of surface contamination they will develop an ideal condition for preferential etching. Also, teflon carriers warm up very slowly and from the outside in thereby creating temperature variation within a carrier. Let us assume the carriers have been preheated and we have used wetting agents to equally expose the crystal surfaces to the etch solution and we have used an agitation during the deep etch process. We are still facing another danger. Being familiar with ammonium bifluoride, we keep forgetting that part of its preparation is a saturation of the solution at its working temperature. Dropping that temperature by only a few degrees will crystallize instantly the ammonium bifluoride. Cloudiness that we observe on the surface of the blanks is nothing more than ammonium bifluoride crystals. The lesson is after etch the quartz wafers should be exposed to a rinse at higher than the etch solution temperature. If we leave undissolved ammonium bifluoride after the first etch the layer acts as an contamination barrier and combined with the moisture in the air it will make its own slow etch in a very unpredictable etch pattern. Every next exposure to the hot etch will make blanks progressively worse. Needless to mention, the cleaning of the blanks between each etching is as important as doing them at the beginning.

One of the inexpensive methods to avoid all of the described problems is to establish the etch rate for deep etch calculation. For that purpose, we can use any rejected blanks of the same surface finish as the ones we intended to etch. Reading of the blanks prior to etch and after substantial time lapse after etch will allow to establish material removal rate at the given time. For all polished material that removal rate will be totally linear. (See Fig. 2)

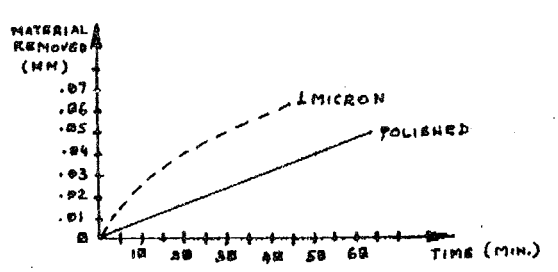


FIG. 2

Other surface finishes will exhibit initial "knee" and faster rates due to removal of all the surface damage first. Causing rapid frequency increases and as the surface improves you will notice a slow down in the etch rate. As I have mentioned earlier, repeatability of this process depends on the uniform surface finish, etch temperature and cleanliness of the blanks.

Etch process is known for decades to be a very useful tool to adjust final frequencies of blanks. There is no reason now knowing the behavior of ammonium fluoride not to take full advantage of it on polished surfaces especially on fundamental and third overtone modes.

Intentionally I have not pointed out the higher order of operating modes. These require almost perfect surface finishes and unless one masters the etch process, I would not recommend using it at this time. There are other ways to adjust final frequencies of blanks that are more conventional which will assure good performance of a crystal unit. The key to a good deep etch is to minimize the number of etches and keep surfaces as clean as possible at all times.

So far we have built 40 to 70 Mhz plano-plano geometry fundamental units. It will take just a base plating and mounting of a few samples to find out that the angle does not reflect the one we started with. Does the etch change the angle ? The answer is no. The higher we go in frequency the external mount, paste stresses will effect the more thinner blanks. Even a perfectly adjusted base does not prevent us from avoiding the results from a different thermal expansion coefficient of base and mount than that of quartz. The typical and not an unusual apparent angle shifts can be as high as +8 to +10 minutes. Building such units the designer has to accommodate those shifts rather than trying to eliminate them.

Inverted mesa geometry where deep etch process is identical to the described already shows no apparent angle changes, unless blanks are mounted under a definite stress and angular shift which are the result of imperfect lapping techniques. That is unquestionable proof that the deep etching process does not change the angle of the blanks regardless of their geometry with typical frequencies up to 200 Mhz fundamentals. It is also true that after the etched surface is identical to the one which we started with and even though a 1 micron finished blanks can become semi-transparent, they still exhibit certain roughness and much lower "Q" as compared with the polished samples.

The typical "Q" for open units with 1 micron finish and deep etched parts are usually not higher than 10 to 15k and about 25 to 30k for polished surfaces.

The real art of modern processing is the blend of a mechanical well known techniques along with chemical finishing thereby shifting the burden from lapping and polishing, where a frequency and geometry control has a "one by one" meaning to the etch area. Proper grouping of blanks to start with can produce large quantities on the batch basis in a very short period of time and a minimum of direct labor involvement.

There are a lot more factors in addition while building high frequency crystal units which are too numerous to mention in this paper. However, the described precautions and close microscopic examination of blanks at each step of a process, should lead to a successful conclusion.

I hope this paper will become a guideline for those who felt that the equipment and lapping techniques were the limiting factor in a pursuit of higher frequency fundamentals and others who have tried in part this concept and failed by not analyzing all of the components and necessary measure to safeguard the total process.

If there are any specific questions regarding the contents of this paper, or any other questions of the process engineering or designing high frequency units, the author will be glad to provide a guidance and any consultations.

Lech Morawiec of Quartz Works
5 Bacon Slip
Milford, MA 01757

Phn 508-473-5890
Fax 508-473-1921

Session 6A Number 56

DIELECTRIC RESONATOR WHISPERING GALLERY MODES APPLICATION TO MILLIMETER WAVELENGTH OSCILLATORS

D. CROS - P. GUILLON

I.R.C.O.M. - U.A. C.N.R.S. - Faculté des Sciences
123 Avenue Albert Thomas - F-87060 LIMOGES Cédex

ABSTRACT

Dielecric resonators excited on their Whispering Gallery Modes are used in millimeter wavelength devices because they present many advantages. In particular they offer oversized dimensions, good suppression of spurious modes and high Q factors.

This paper presents both electromagnetic and electrical parameters of Whispering Gallery dielectric resonators modes and their applications in millimeter wavelength filters, power combiners and oscillators.

INTRODUCTION

In millimeter wave military and communication systems there is a requirement for more compact packages and MIC compatibility. These results are of increasing interest in using low loss, temperature stable, high permittivity dielectric resonators and microstrip technologies in some millimeter wave components such as oscillators and filters. Unfortunately, at millimeter wave frequencies, the dimensions of the fundamental $TE_{0,1,\delta}$ mode dielectric resonators become quite small and are very difficult to control, which significantly limits their applications in the millimeter wave frequency band.

The Whispering Gallery Modes (W.G.M.) cylindrical dielectric resonators provide a promising solution to this problem.

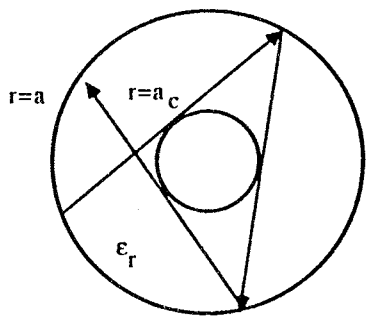
In this paper, after a description of the whispering gallery modes phenomena, we define both electromagnetic and circuits parameters of these resonators when they are coupled with transmission lines. Then the design and testing of W band whispering gallery mode DR oscillator is described.

WHISPERING GALLERY MODES DIELECTRIC RESONATOR CHARACTERISTICS

In a dielectric rod the WGM's are described as comprising waves running against the concave side of the cylindrical boundary of the rod [1] [2]. The waves moves essentially in the plane of the circular cross section. Most of the modal energy is confined between the cylindrical boundary a and an inner modal caustic a_c.

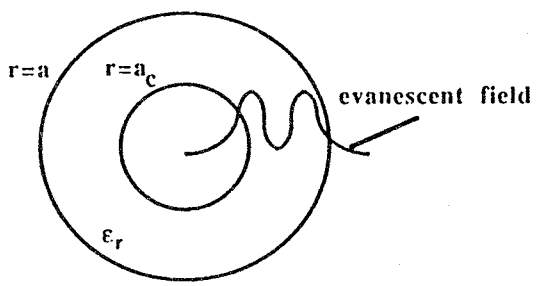
The energy confinement can be explained from a ray optics point of view. In this way, it is said that a ray is totally reflected at the dielectric air interface, it is then tangent to an inner circle called a caustic (figure 1).

Thus, the ray moves merely within a small region near the boundary. From an analytical point of view, it is well known that waves guided in a dielectric rod can be described by Bessel functions Jn(kr). The WG modes correspond to those for which the argument kr is of the order of n which is a large integer.



- Figure 1 -
Ray optics representation

In this way, it can be shown that the field is oscillating between the rod boundary and a slightly small radius called a caustic, while it exponentially decays elsewhere (figure 2).



- Figure 2 -
Analytical representation

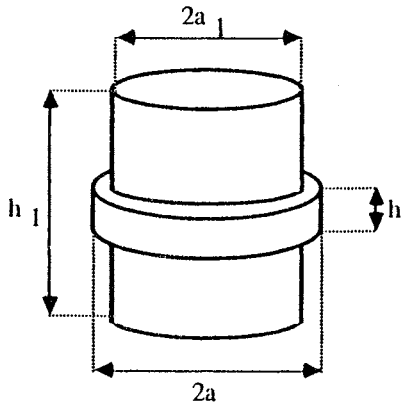
The WG modes can be excited in two different types of dielectric structures. In the first one the geometry of the DR considered is different from the conventional disk ones. In this structure (figure 3), the DR WGM's are characterized by a second energy confinement (axial) ensured by the enlargement of the resonator radius in the central zone. In this area the modes propagate axially with a propagation constant which is less than the radial ones and they decay exponentially away from the discontinuity of the resonator radius [2].

In the second type the electromagnetic characteristics of WG modes are essentially a function of the DR radius. So WGM's can be excited in their dielectric disks, the diameter of which 2a is greater than its thickness h (figure 4). These resonators are very interesting because they are compatible with the design of millimeter wavelength integrated devices.

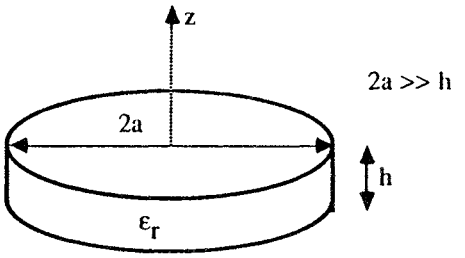
SESSION 6A

UHF AND SAW DEVICES

Chairman : P. DEFRANOULD



- Figure 3 -
Cylindrical dielectric resonator supporting WGM



- Figure 4 -
Planar dielectric resonator supporting WGM

WHISPERING GALLERY MODES ADVANTAGES

For both geometries described previously, DR's acting on WGM's are very interesting for a number of reasons :

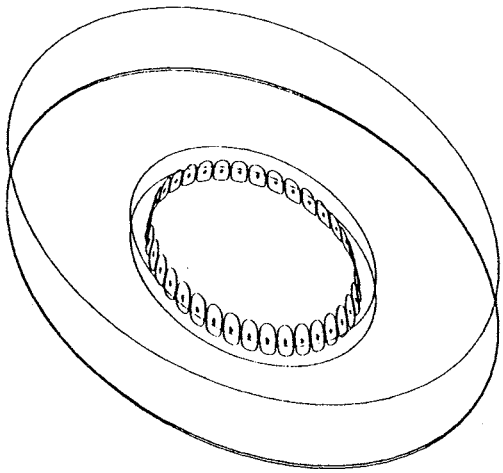
- their dimensions are large even in the millimeter wavelength band
- their quality factors are large : the Q_0 factor of these WGM is limited only by the value of the loss tangent of the material used to realize the dielectric resonator, since radiations losses are negligible
- they offer good suppression of spurious modes because the propagation constant along the z axis is very small and unwanted modes leak out axially and can be absorbed without perturbation
- since WGM's are running essentially in the azimuthal direction, they offer the possibility of new microwave devices, such as directional filters or power combiners, in which the reflection coefficient will be very low

WHISPERING GALLERY MODE NOMENCLATURE

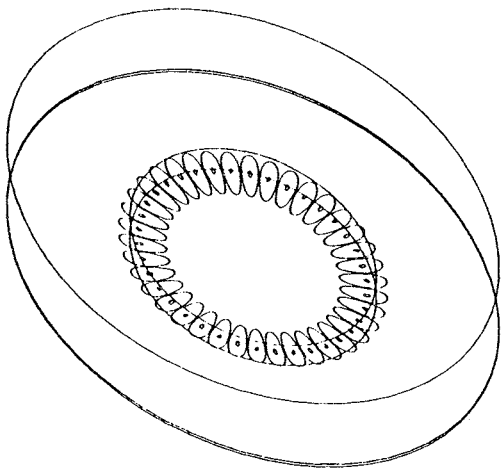
The DR WGM's are classified as either $WGE_{n,m,l,\pm1}$ in which the electric field is essentially transversal or $WGH_{n,m,l,\pm1}$ for which the electric field is essentially axial. The integer n denotes the azimuthal variations, m the radial variations and l the axial ones. Finally, two possible rotating senses are denoted by respectively ±1. However, in what follows, this designation will be omitted because in an isotropic dielectric medium the resonant frequencies of the mode are the same irrespective of the sense of rotation. WGM's are periodic according to the azimuthal number n, and the number of modes in a bandwidth increases with the diameter of the DR.

WHISPERING GALLERY MODES ELECTRO-MAGNETIC PARAMETERS

The sophisticated numerical finite element method is used to define resonant frequencies, electromagnetic fields, unloaded Q factors of DR WGM's. An example of the field cartography associated with WGE and WGH modes of a dielectric disk resonator is given in figure 5.



a) $WGE_{19\ 0\ 0}$ mode



b) $WGH_{19\ 0\ 0}$ mode

- Figure 5 -
Magnetic field lines of W.G. modes of a D.R.

Using this software we can also evaluate the WGM resonance frequencies of a DR of diameter 2a=5 mm, thickness 0.635 mm made with alumina. We can note for example that the frequency spacing between modes $WGE_{8,0,0}$ and $WGE_{9,0,0}$ is about 6 GHz (the resonance frequencies of which being respectively 92,4 GHz and 98,5 GHz).

COUPLED DIELECTRIC RESONATOR WGM's ANALYSIS

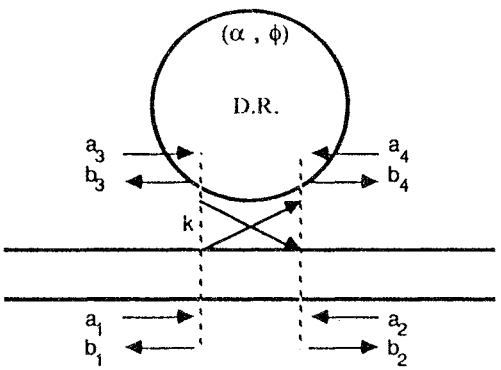
Dielectric resonators acting on their WGM's can be excited in different way so that we can obtained either a traveling WGM or a stationary one.

Due to their high Q and excellent integrability in circuits DR can be directly used as a frequency determining element for realizing a stable transistor or diode oscillator.

The DR in realizing a stable oscillator can be used either as a serie feedback element or a parallel feedback element. In this last case the DR is simultaneously coupled to two transmission lines, since in the first one it is only coupled with a line.

a) - Dielectric resonator WGM's coupled with a line

When a DR is coupled to transmission lines, at resonance all energy coupled remains stored in the resonator and no power is transmitted or reflected. A schematical representation of this system is given in figure 6.



- Figure 6 -
Schematic representation of a D.R. coupled to a transmission line

Let a_i and b_i be respectively, the incident and reflected normalized wave amplitudes in the coupling plane of the coupler [3] [4].The scattering matrix equation of the coupler can be expressed as follows.

$$S = \begin{bmatrix} 0 & \sqrt{1-k^2} & 0 & jk \\ \sqrt{1-k^2} & 0 & jk & 0 \\ 0 & jk & 0 & \sqrt{1-k^2} \\ jk & 0 & \sqrt{1-k^2} & 0 \end{bmatrix}$$

This system can be reduced by taking into account the relation between the wave amplitudes at ports 3 and ports 4.

$$a_3 = b_4\, e^{-j(\alpha+j\phi)}$$

$$a_4 = b_3\, e^{-j(\alpha+j\phi)}$$

where ϕ is the total phase shift around the ring resonator length. Resonance occurs when $\phi=2\pi n$. α is the total attenuation of the wave traveling in the resonator. If port 2 is terminated with a matched load and if the circuit is fed at port 1, the following relations for b_2 and b_4 are obtained.

$$b_2 = \frac{\sqrt{1-k^2} - e^{-\alpha}}{1 - \sqrt{1-k^2}\, e^{-\alpha}}\, a_1$$

$$b_4 = \frac{jk}{1 - \sqrt{1-k^2}\, e^{-\alpha}}\, a_1$$

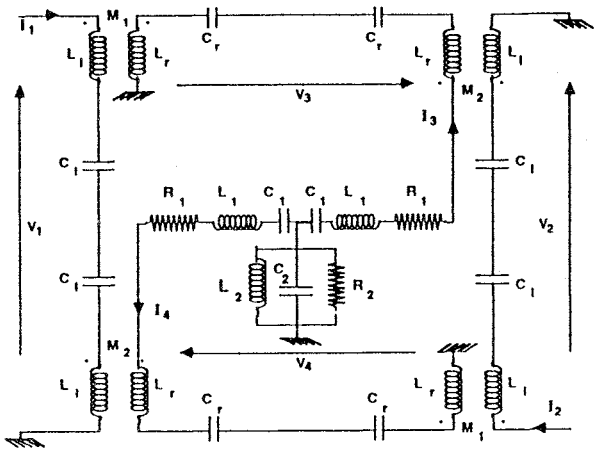
Here, it should be noticed that the backward transmission coefficient vanishes when $\sqrt{1-k^2}$ is equal to $e^{-\alpha}$; in this case waves at port 2 and 4 become :

$$|b_2| = 0$$

$$|b_4| = \frac{1}{\sqrt{1-e^{-2\alpha}}}\, a_1$$

This means that we can make the backward transmission vanish by adjusting the coupling coefficient k.

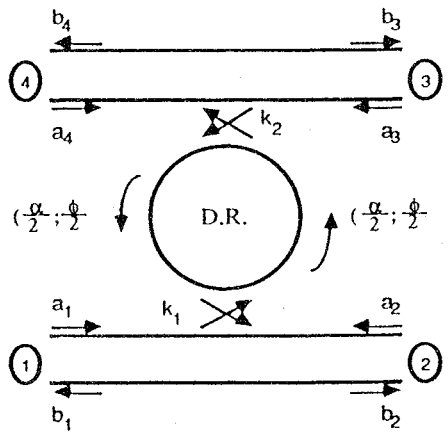
The circuit model of the dielectric resonator WG modes coupled with a transmission line is given in figure 7 where L_l, C_l are equivalent elements of the transmission lines, L_r, C_r are those of dielectric resonators and R_1, L_1, C_1, R_2, L_2, C_2 simulate the propagation phenomena in the dielectric resonator. M is the mutual inductance which represents coupling between DR and the transmission line.



- Figure 7 -
Equivalent network of a resonator coupled to one line

b) - Dielectric resonator WGM's coupled with two lines

One of the important advantages of the WGM of DR's is its analogy with a traveling wave ring resonator. So energy may be obtained from a resonator coupled to a line by coupling it to a second line (figure 8).



- Figure 8 -
Schematic representation of a D.R. coupled to two transmission lines

In this case, the response of the system becomes a directional filter [4]. The analysis given previously can be applied to determine the scattering parameters of this new device :

$$b_1 = 0$$

$$b_2 = \frac{\sqrt{1-k_1^2} + \sqrt{1-k_2^2}\, e^{-j(\alpha+j\phi)}}{1 - \sqrt{1-k_1^2}\,\sqrt{1-k_2^2}\, e^{-(\alpha+j\phi)}}\, a_1$$

$$b_3 = \frac{-k_1 k_2\, e^{-(\alpha+j\phi)}}{1 - \sqrt{1-k_1^2}\,\sqrt{1-k_2^2}\, e^{-(\alpha+j\phi)}}\, a_1$$

$$b_4 = 0$$

At the resonant frequency, the phase shift around the ring resonator length is given by $\phi=2\pi n$, so $e^{j\phi}=1$.

It should be noticed that the backward transmission coefficient at port 3 is maximum when the following relation is satisfied.

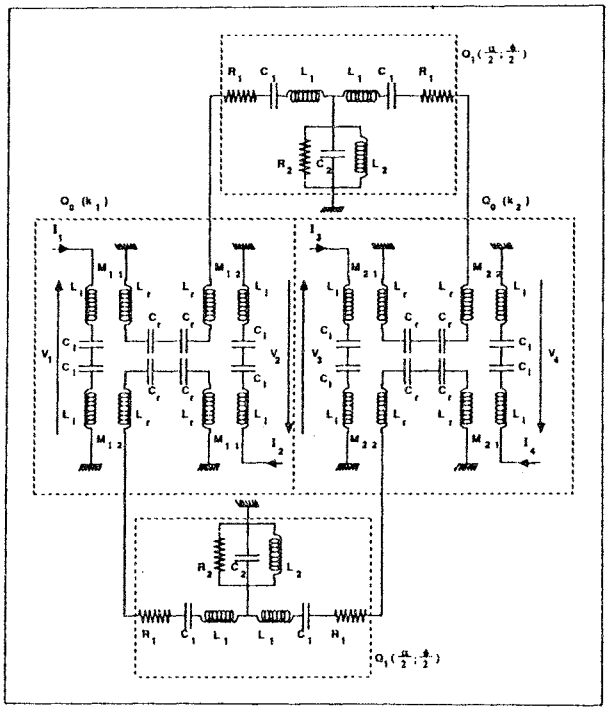
$$\sqrt{1-k_1^2} = \sqrt{1-k_2^2}\, e^{-\alpha}$$

Taking into account this relation, the scattering parameters become :

$$b_1 = b_2 = b_4 = 0$$

$$b_3 = \sqrt{1 - \frac{1 - e^{-\alpha}}{k_1^2}}\; a_1$$

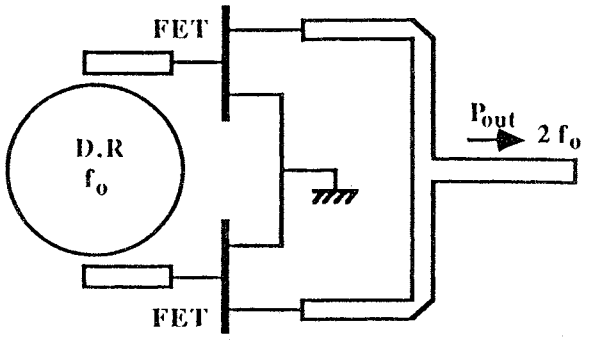
The equivalent model of the DR WGM's coupled with two transmissions lines as given in figure 9. The equivalent elements have been defined in figure 7.



- Figure 9 -
Equivalent network of a resonator coupled to two lines

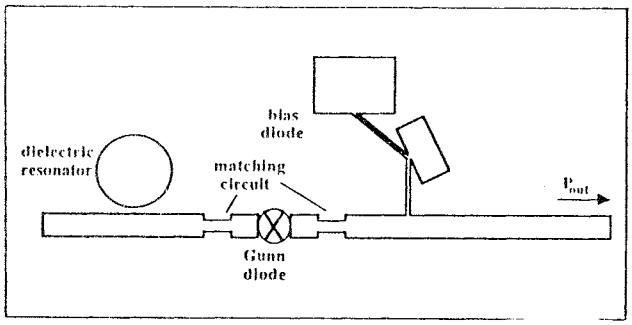
MILLIMETER WAVELENGTH DIELECTRIC RESONATOR WGM's OSCILLATOR

Two types of DR oscillators acting in the millimeter wavelength frequency band have been realized. One is an oscillator doubler using a FET acting at 47 GHz (figure 10) [5], the other is a Gunn oscillator acting directly at 94 GHz



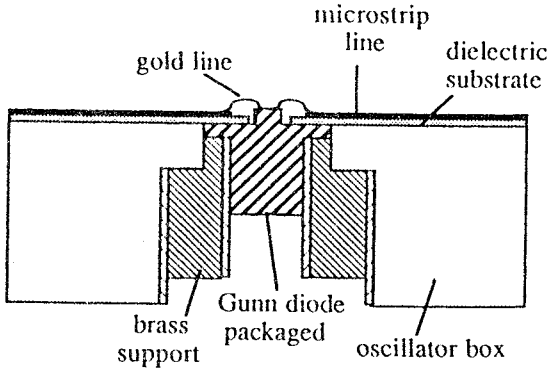
- Figure 10 -
Schematic representation of the oscillator doubler

The planar Gunn diode oscillator made with a DR excited on its WGM is described in figure 11.



- Figure 11 -
Planar Gunn diode oscillator

The circuit is formed on a 0.127 mm thick substrate of relative dielectric constant 2.2. The active element is a packaged Inp Gunn diode. The cross section of the Gunn diode in the circuit is given in figure 12.

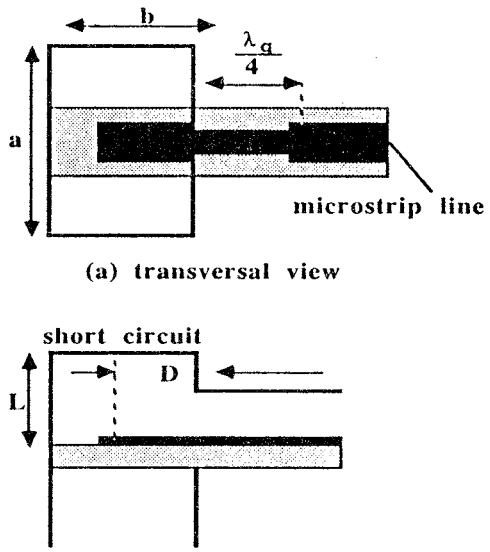


- Figure 12 -
Cross view of the Gunn diode report circuit

The dielectric resonator (ε_r 24, diameter 3.2 mm, height 0.4 mm) used in its $WGE_{10,0,0}$ is excited at 98.5 GHz by means of a microstrip line.

To respect the oscillations conditions, the load impedance was matched to the diode impedance with a quater wavelength line matching transformers. The Gunn bias circuit is realized by using the same technique.

Since most standard measurement equipments at millimeter wavelength frequency band uses waveguides. E plane probes transition from microstrip to waveguide has been computed and realized (figure 13) [6].

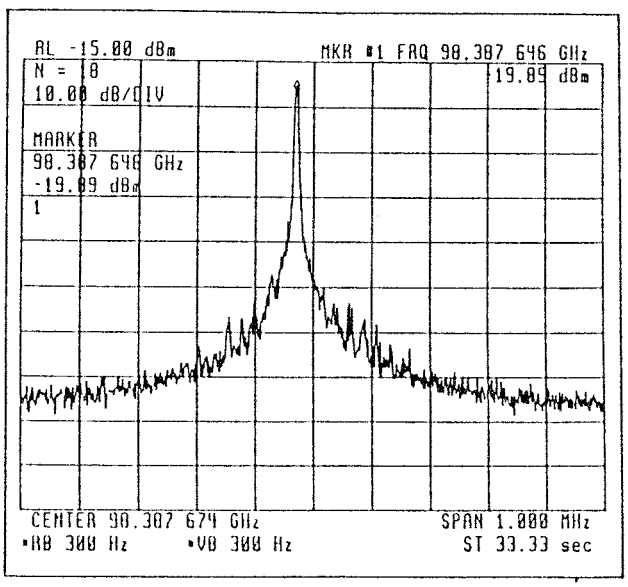


(a) transversal view

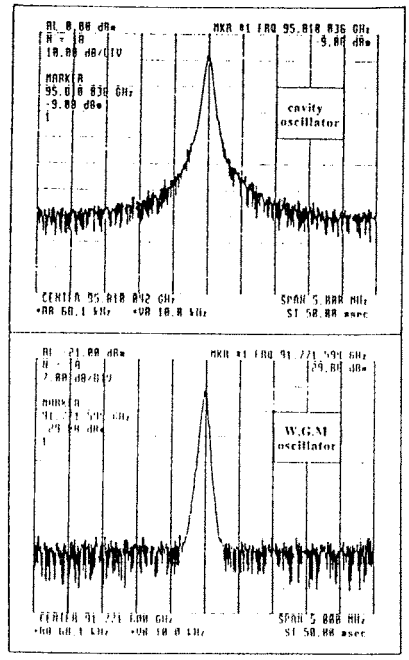
(b) lateral view

- Figure 13 -
Microstrip waveguide transition

The experimental response of the DR WGM's oscillator is shown in figure 14. The output power is 6 dBm for a bias voltage of 4.5 volts and current of 0.5 A. To evaluate the WG mode oscillator performances, its spectrum has been compared to that of a conventional metallic cavity oscillator (figure 15).



- Figure 14 -
Dielectric resonator oscillator response



- Figure 15 -
Comparison between WGM oscillator and cavity one's

CONCLUSION

This paper shows the feasibility of a planar Inp Gunn diode dielectric resonator oscillator operating in 90 - 100 GHz band and which has been realised using the WG dielectric resonator modes.

Next steps will be to use HEMT transistor to supply diode in W band oscillator devices to reduce power consumption and increase the relability.

This type of oscillator could be used for radiometric application type

REFERENCES

[1] C. VEDRENNE, J. ARNAUD
"Whispering Gallery Modes of Dielectric Resonators"
Proceedings IEE, part H, August 1982, pp.183-187

[2] J. ARNAUD
"Note on the use of Whispering Gallery Modes in a dielectric rod"
Bell Lab. momerandum,September 1971

[3] J.ALTMAN
"Microwave Circuits"
D. Van Nostrad, Co Inc. 1964, Chap V, pp.202-213

[4] X.H. JIAO, P. GUILLON and J. OBREGON
"Theoritical analysis of coupling between W.G. dielectric resonators modes and transmission lines"
Electronics Letters, Janvier 1985, pp.88-89

[5] L.A. BERMUDEZ, P. GUILLON, J. OBREGON, A. BERT
"A 94 GHz low noise Ga As FET oscillator using Whispering Gallery dielectric resonator modes and a new push-push configuration reducing 1/f converted noise"
IEEE International Microwave Symposium, New-York, June 1988

[6] YI-CHI SHIH
"Millimeter-wave device characterization"
Alta Frequenza, vol. LVIII, n°5-6, September-December 1989

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 6A Number 57

NO_2-SURFACE ACOUSTIC WAVE GAS SENSOR

M. Hoummady, D. Rebière*, D. Hauden, J. Pistré*, P. Cunin** and R. Planade**

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France
* IXL - Université de Bordeaux I (URA 846) - 351 cours de la Libérations - 33405 Talence Cedex - France
**Centre d'Etudes du Bouchet - Vert le Petit - France

Abstract

The operating of this gas sensors is based on the perturbation of the SAW velocity by the variation of the absorbed mass by an organic thin sensitive coating deposited on the propagation path.

In order to improve both gas response time and sensitivity, the operating temperature must be optimized. To reduce the temperature effect, SAW is propagating on quartz crystal substrates. The quartz cuts are selected with the right turnover point at working temperature. The organic cupper-phthalocyanine thin film coated on substrates is sensitive selectively to the NO_2 gas.

Gas adsorption has been modelized as a variation film mass. Sensors sensitivity has been calculated ; it depends on coated thickness, mass gas adsorbed and crystal cutting.

Experiments measurements has been effected both for various film thickness and different temperatures.

Introduction

Surface acoustic wave (SAW) devices are studied as gas sensors since the 1979 Wohltjen and Dessy's works [1].

A recent review of the corresponding applications made by Nieuwenhuizen and Venema [2] pointed out that several authors have up to now studied such NO_2 sensor devices [3]. The piezoelectric materials used for these studies were quartz [9], lithium niobate and zinc oxide. The sensitive layer was always a metalphthalocyanine (M-PC), M being copper, lead, ...

The work reported here concerns a Rayleigh wave NO_2 sensor using quartz as piezoelectric substrate and copper-phthalocyanine (Cu-PC) as chemical interface.

The first part is devoted to theoretical studies and modelings of the SAW device with and without his chemical coating. The approach for modelisation is somewhat different from the one usualy used in this domain.

A second part deals with the experimental techniques : measurement of the LAME constants of Cu-PC, design of SAW devices and of an automatic measurement bench for temperature, gas flow and frequency control as well as for data registration and processing.

The obtained results and discussion are developed in the third part.

This work is supported by DRET-CEB contract # 25068/89/ETCA/CEB/DED

1. SAW gas sensor principle on quartz crystal

The principle of a SAW gas sensor is given in Fig. 1 : two SAW delay lines are metallized on the same quartz plate and they are connected with two amplifiers to build the sensitive oscillator and the reference oscillator.

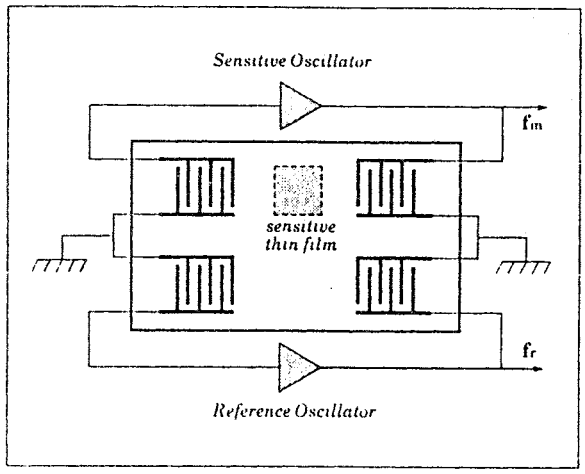


Fig. 1 : Principle of the NO2-gas sensor

When the NO_2 gas is adsorbed by the Cu-PC thin film deposited along the acoustic beam of the sensitive delay line, the mass of the film increases. This leads to a proportional wave velocity variation and because the corresponding phase shift, the frequency of this oscillator is also shifted.

Sensitivities versus temperature and mass variations are function of the quartz anisotropy. Then the choice of the quartz cut of the substrate must be first defined by the operating temperature of the sensor which corresponds to a turn over point of the frequency-temperature characteristic. Figure 2 shows the frequency-temperature characteristic of two singly rotated cuts of quartz that have been studied for gas sensing. The regular ST,X cut (Fig. 2a) is generally used at room temperature because its turn over point around 25°C. The second cut (Y+28°,X) has its turnover point around 120°C and at that operating temperature, water vapor adsorption is also limited (Fig. 2b).

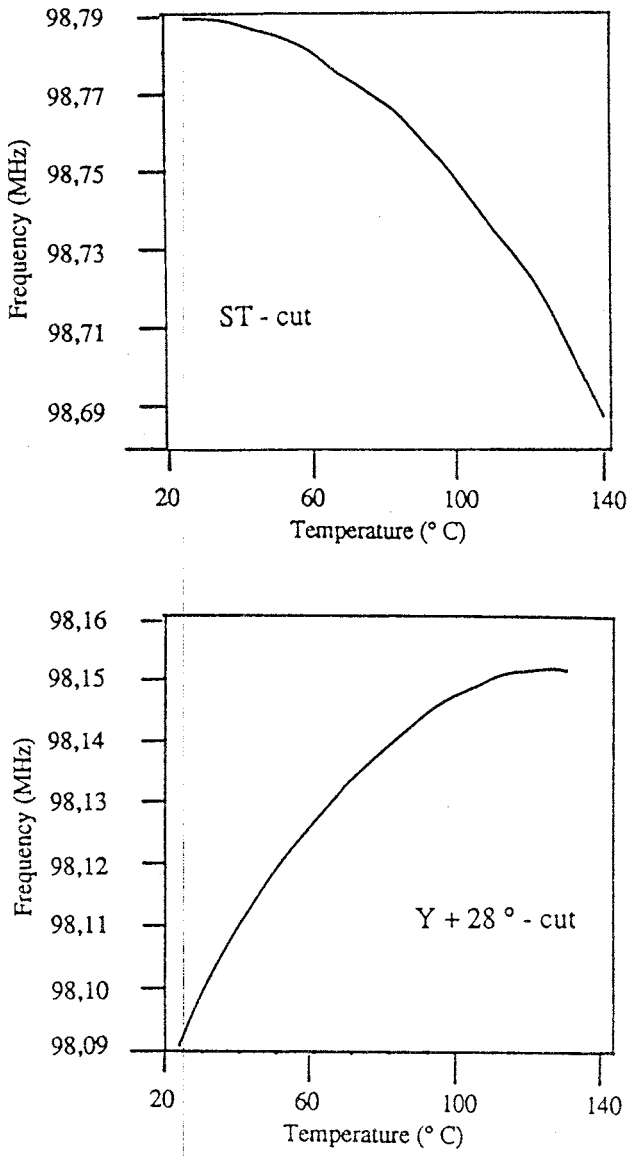


Fig. 2 : Static frequency-temperature characteristic
a) ST,X-cut b) Y + 28°,X cut

2. Mechanical properties of the thin film

The SAW velocity VR is function of the mechanical properties of the quartz substrate, then of the cut, and also of the mechanical properties of the thin film.

Theoretically, the velocity variations are given by the relation [4] :

$$\frac{\Delta V}{V_R} = \frac{V_R\, k}{4\, P_R} \left| \rho\,(V_{Rx})^2 + \left(\rho - \frac{\mu}{V_R^2}\right)(V_{Ry})^2 + \left(\rho - \frac{\mu}{V_R^2}\,\frac{\lambda+\mu}{\lambda+2\mu}\right)(V_{Rz})^2 \right|$$

calculated at the plate surface.

V_{Rx}, V_{Ry}, V_{Rz} are the components of the SAW velocity V_R determined for the free propagation surface. ρ, μ and λ are the specific mass and the Lamé constants of the deposited film and h is its thickness.

In this relation, SAW parameters (V_R, V_{Ri} and P_R) are easily calculated, but thin film mechanical properties must be measured because only hand ρ are known. Also ρ is equal to 1 630 kg/m³ for Cu-PC film.

To evaluate the Lamé constants μ and λ sensors made in the same technology with different quartz cuts are measured by an amplitude-phase method presented on Fig. 3.

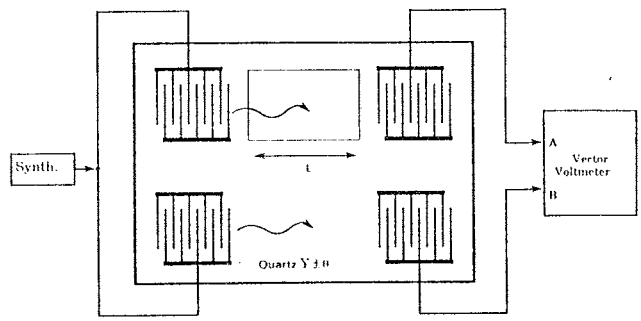


Fig. 3 : Phase difference measurement
with vector voltmeter (ΔΦ is proportional to ΔV)

The two delay lines are excited by the same signal, and the vector voltmeter measures simultaneously the amplitude ratio and the phase difference due to the thin film which is proportional to the SAW velocity difference.

This experimental data used in a software [5], [6] enables to simulate the behavior of the SAW in thin film on quartz cuts when Lamé constants μ and λ are taken as parameters. Figures 4a and 4b represent respectively the velocity variations for Y + 28°,X cut and ST,X cut versus λ when μ is taken as parameter.

Experimental points are ploted on the same graphes.

Only one couple of values are solution for the thin film Cu-PC : $\mu = 5.10^9$ N/m² , $\lambda = 2.10^9$ N/m²

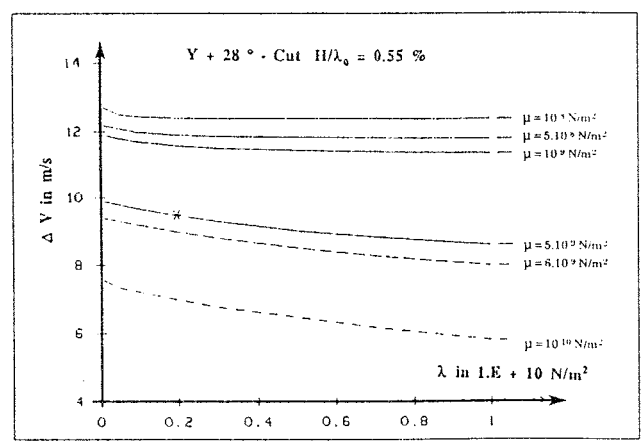


Fig. 4a : Velocity variation versus Lamé constants
for a Y + 28°,X cut
★ experimental point

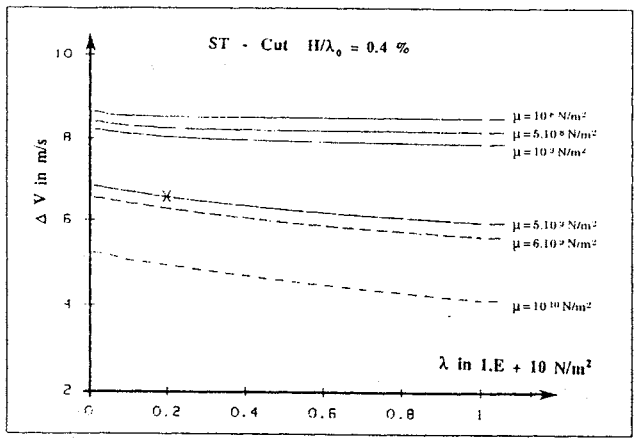


Fig. 4b : Velocity variation versus Lamé constants
for a ST,X cut
★ experimental point

3. Theoretical sensitivity of the NO2 gas sensor

To simulate the sensitivity of the gas sensor, it has been supposed that only the gas mass adsorbed acts on the specific mass. Figure 5 shows the calculated sensitivity versus the film thickness for ST-cut and Y + 28° cut.

Then, if the film thickness is 1 000 Å, the velocity variation will be equal to 20 kHz for a mass variation of 1%, this is corresponding to 1 ppm per nanogramm.

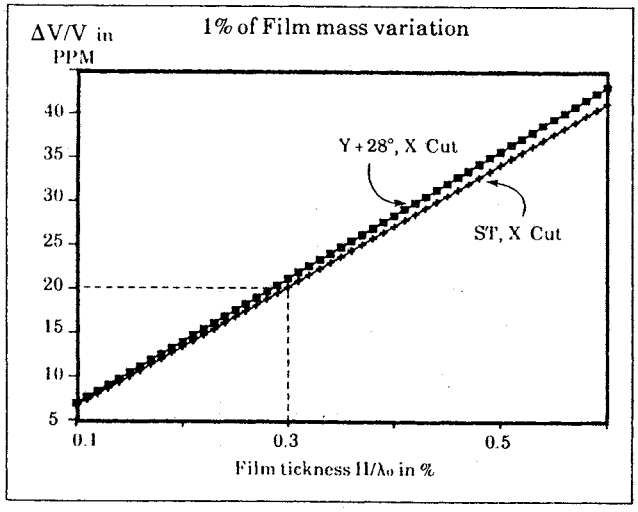


Fig. 5 : Sensitivities simulate for the two cuts versus film thickness for 1% of fibre mass variation due to gas adsorption

The (Y + 28°,X) quartz cut seems more interesting than ST,X cut in regards of the mass sensitivity lightly better and also its behavior at high temperature which limits water vapor adsorption and could improve the gas disorption.

4. Gas sensor cell

Two identical delay-lines are made on the same quartz substrate by chemical etching of a thin film of aluminum obtained by RF sputtering. The transducer geometry is as follows : 100 pairs of 4 µm fingers separated by 4 µm with an aperture of 1.5 mm ; the center to center distance between the transducers on a same line was 9.6 mm (free distance between the transducers was 6.4 mm). A 3×4 mm^2 layer of Cu-PC was deposited on one of the lines.

The quartz substrates were pasted on to alumina ($A\ell_2O_3$) plates with a thick film platinum resistor to heat the device. The temperature of the quartz is controlled by a thermocouple glued on to the surface. The entire device is then placed in an air-tight stainless steel box connected to the gaz line.

Photograph 1 shows the complete configuration of the SAW chemosensor cell.

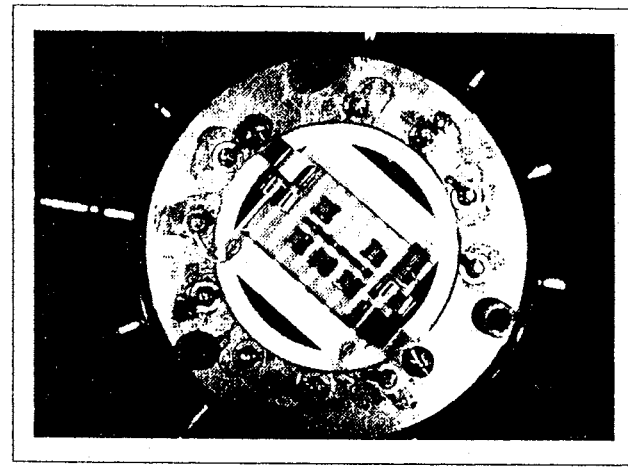


Photo 1 : Overall view of the open cell

5. Automated measuring bench

An automated measuring bench is built to follow the changes in the parameters of the sensor. This bench is controlled by an 80386 SX microcomputer, enabling the results to be stored in memory and processed. The basic characteristics are :

- separate measurements of the frequencies of each oscillator using an HP5334A (ultra-high stability counter). The measurements are made alternately every second.
- control and regulation at 0.05°C of the quartz temperature using an autoadaptive regulator controlled by RS232 liaison.
- recording and processing of results.

6. Results and comments

6.1 - Measurements without NO_2

The oscillators work at a frequency slightly higher than 98 MHz, depending on the SAW line. Their short term stability (relative frequency change) in the absence of gas flow is 2×10^{-8} s^{-1}. The frequency varies from 10 to 80.10^{-8} s^{-1}, according to the temperature, when the cell is submitted to a 0.1 litre per minute dry air flow.

The differentiel stability (evolution of the relative frequency difference of two oscillators made from SAW delay-lines belonging to the same quartz substrate) varies with the conditions (Cu-PC coating thickness, temperature), from 0.3 to 3.10^{-8} per minute. When the Cu-PC coating is relatively thick, 0.6 to 0.7 µm, and from 90°C and higher, a progressive increase in the coated SAW line controlled oscillator drift may be observed. This phenomenon was attributed to a Cu-PC decomposition which is followed, as shown here after, by a loss in sensor sensitivity.

6.2 - Measurements made with NO_2

Nitrogen dioxide is in fact a mixture of nitrogen oxides noted here as NO_x. The partial concentration or partial pressure of this gas is automatically controlled by the gas line. A vector gas (dry air) is used. Its flow is 0.1 litre per minute, remaining constant during the tests. NO_x partial pressures used vary from 1 to 100 ppm.

Figure 6 gives a typical response of the sensor.

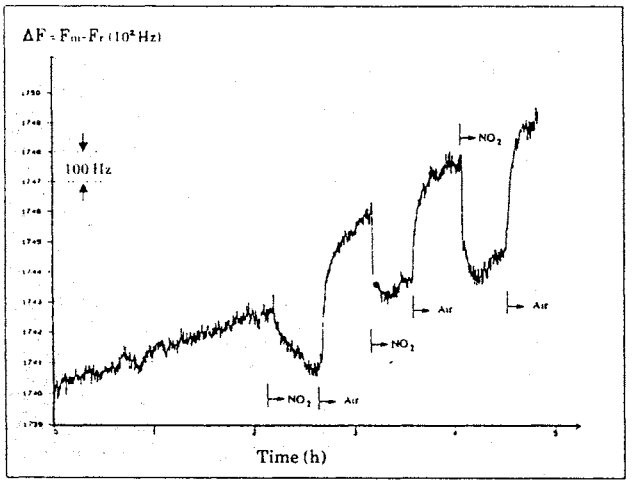
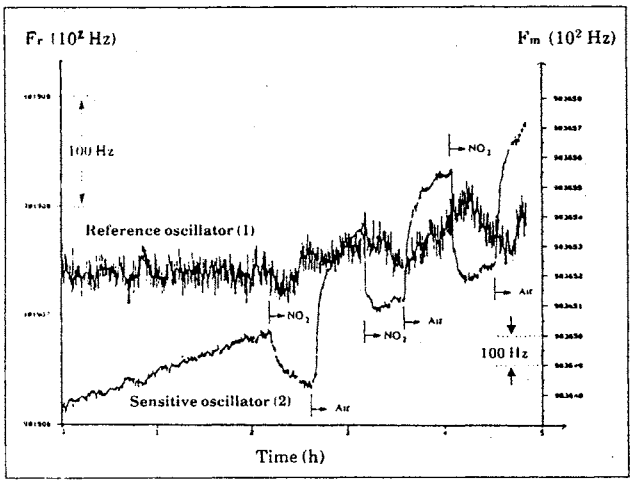


Fig. 6 : NO_x sensor response
quartz cut : Y + 28° ; quartz temperature : 90°C

As it can be seen, the response time of the sensor is important during the first test. This phenomenon is systematically observed each time the sensor was kept for a long time in free atmosphere. It is as if a certain "cleaning" of the chemical deposit occured in this first test.

Figure 7 represents the frequency evolution of both oscillators (reference oscillator and measurement oscillator) during the same test.



Fig; 7 : Frequency variation of each oscillator during the test
1 : reference oscillator ; 2 : measurement oscillator

Changes in the drift of the referenceoscillator are observed each time that NO_x is introduced, without however being able to correlate its response to the measurement oscillator drift. The reference and measurement oscillator long term drifts are different. The measurement oscillator drift behavior seems to indicate a slow mass loss of the deposit as previously observed.

Table 1 gives the geometricalc haracteristics of our sensor compared with TNO-TU Delft Sensor [7], [8].

	Our sensor	TNO-TU-Delft-Sensor
aperture (mm)	1.5	3
thin film surface (mm²)	3 × 4	4 × 6
active surface (mm²)	1.5 × 4	3 × 6
beam length (mm)	9.6	7
quartz surface (mm²)	15 × 20	14 × 6
quartz thickness (mm)	1	0.5

Table 2 summarizes the overall results obtained and compares them to those given in the literature. The following remarks may be made :

- the deposit thickness, more than 4 times lower for the sample with a ST quartz cut than the one corresponding to the Y + 28° cut, makes the sensor less sensitive. This result indicates that adsorption is not a phenomenon located at the surface, but rather a gas migration taking place in the porous bulk of the deposit. The sensitivity, normalised for 1 µm thickness, is approximately constant.
- response times to adsorption and desorption are temperature dependent. Bearing in mind the sensor degradation observed at 120°C (see Fig. 8), it appears desirable to limit its operation at 90°C.

Quartz-cut	Our sensor			TNO-TU-DELFT SENSOR ST,X
	ST,X	ST,X	Y + 28°,X	
Temperature (°C)	30	120	120	150
Film thickness (µm)	0.17	0.17	0.74	3
Sensitivity (HZ/ppm)	0.7	3	3	3
Frequency drift (HZ/min)	0.3	0.9	14	0.5
Frequency stability /10 s	10	80	20	(8)
Adsorption response time (min)	20	1	1	2
Desorption response time (min)	>30	5	>30	

Table 2 : Results and comparison

Whereas all sensitivity measurements of this study were carried out for NO_2 partial concentrations of 10 ppm, much higher values appear when measurements are made at lower concentrations.

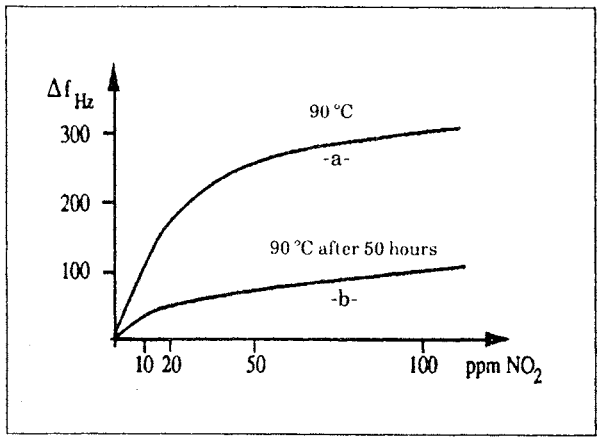


Fig. 8 : Sensor detection curves. Measurements are made at 90°C with a quartz substrate (Y + 28° cut), Cu-PC thickness : 0.72 mm
a : before aging ; b : after 50 hours at 120°C

Conclusion

This study allowed the design of a self-contained device, the electronics of which has been miniaturized and integrated in the measurement cell. The operating temperature is 90°C and the quartz substrate cut chose so as to present its compensation point at this temperature.

Experiments made with the measurement bench at 350 MHz showed that the device sensitivity could easily be increased when working at higher frequency.

Acknowledgments

It is a pleasure to acknowledge the Direction des Recherches et Etudes Techniques of the French Ministry of Defense (DRET) and especially the Centre d'Etudes du Bouchet (CEB) who sponsored this work.

Bibliography

[1] H. Wohltjen and R. Dessy, Analytical chemistry 51, 1458-1465 (1979)

[2] M.S. Nieuwenhuizen and A. Venema, "Surface Acoustic Wave chemical sensors", Sensor and Materiels 5, 261-300 (1989).

[3] B.A. Auld, Acoustic fields and waves in solids : John Wiley Publication, New York, vol. II, chapter 12, 1973.

[4] B.A. Auld, Acoustic fields and waves in solids, Wiley Interscience New York, Rayleigh wave properties, Appendix 4, vol. II, pp. 375-393n 1973.

[5] A.H. Fahmy and E.L. Adler, "Propagation of acoustic surface waves in multilayers : a matrix description", Appl. Phys. Lett. 22, 485-497 (1973).

[6] E.L. Adler, J.K. Slaboszewicz, G.W. Farnell and C.K. Jen, "PC Software for SAW propagation in anisotropic multilayers", IEEE Trans. on UFFC, 37, 215-223 (1990).

[7] A. Venema, E. Nieuwkoop and M.J. Vellekoop, "NO2 gas concentration measurement with a surface acoustic wave gas sensor for nitrogen dioxide", IEEE Trans. on UFFC 34 (2) (1987).

[8] M.S. Nieuwenhuizen, A.J. Nederlof and A.W. Barendz, "Metallophthalocyanine as chemical interfaces on a surface acoustic wave gas sensor for nitrogen dioxide", American chemical society 60, 230 (1988).

[9] M. Hoummady, D. Rebière, D. Hauden, J. Pistré, P. Cunin and R. Planade, "Surface acoustic wave (SAW) gas sensors : modelization and structure", 3rd Meeting on Chemical Sensors, Cleveland, Ohio (USA), sept. 90.

Session 6A Number 58

QUARTZ CRYSTAL ROTATION RATE SENSORS

J.J. Gagnepain*, G. Marianneau*, L. Gaudriot**

*Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France
** Société METRAVIB, 64 chemin des Mouilles - B.P. 182 - 69132 Ecully Cedex

Abstract

When a material particle is submitted to a rotation it experiences an acceleration, the Coriolis acceleration, proportional to the rotation rate and to its instantaneous velocity. In case of elastic waves in a solid the Coriolis acceleration is nonzero only under conditions on the configurations of the rotation and velocity vectors. The effects of the Coriolis acceleration on a wave is a change of the wave velocity and the driving of an orthogonal vibration. Both are discussed for sensing rotation rates. Comparison is also made between travelling waves and steady state waves, and between surface acoustic waves and bulk acoustic waves. A rotation rate sensor based on amplitude detection of the induced flexure vibration in a quartz crystal flexure bar resonator is described. The sensitivity of the sensor is given.

1. Introduction

The market of sensors is expanding today in new directions and in particular in the direction of the large consumer products. Automobile is one of the main applications, which requires sensors for controlling the efficiency of the engine and the dynamical stability of the car. Unexpansive navigation devices could also be developed in the next future. For such markets the first condition is the reduction in size and cost of the devices. Specifications in term of the various spurious sensitivities to the environment are not so easy. The levels of temperature variations, of accelerations and vibrations, etc. are generally high in a car. But new possibilities are offered by the technologies originated from microelectronics, which today can be successfully used for the micromachining of sensors made of silicon, quartz or other materials.

The principles which can be used for detecting angular rotations are of different types : velocity change of a travelling acoustic wave, induced orthogonal vibration in a steady state wave resonator, beam deflection of an ultrasonic wave in a gas, gyrolaser using light waves in an optic fiber or in a ring laser.

In the present work the effect of the Coriolis acceleration on an elastic wave is used. The first effect is the change of the velocity of the acoustic wave. But as it will be shown several conditions are required. Only travelling waves can be used since with steady state waves the Coriolis effect changes sign when the wave travelling direction is reversed ; therefore in steady state waves there is compensation. In a similar manner there is not effect when the motion of the material particle is rectilinear. The mechanical motion must have two components which are not in phase. Therefore only surface wave or bulk wave with elliptic or quasielliptic motions, Rayleigh waves and flexure waves, respectively are sensitive.

The second possibility is to use the vibration which is induced by the Coriolis acceleration. This can be easily achieved by means of piezoelectric effect as it will be shown, in particular in the case of a flexure wave. Therefore emphasis is put on flexure mode resonators in the present paper.

2. Coriolis force

Considering the motion represented by the displacement vector $\vec{u}$ of a material particle with specific mass ρ_0, submitted to an angular rotation of rotation rate vector $\vec{\Omega}$, the material particle experiences of an inertial force $\vec{F}_c$ due to the Coriolis acceleration [1] and given by

$$\vec{F}_c = 2\,\rho_o\,\vec{\Omega} \times \frac{\partial \vec{u}}{\partial t} \qquad (1)$$

where the symbol $\times$ denotes a vectorial product.

From Eq. (1) it is obvious that the Coriolis force is not zero only if the two vectors $\vec{\Omega}$ and $\vec{u}$ are not colinear. The Coriolis force $\vec{F}_c$ is orthogonal to the rotation vector $\vec{\Omega}$ and to the velocity vector $\partial\vec{u}/\partial t$ as shown in Fig. 1.

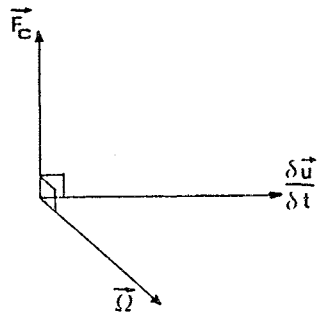


Fig. 1 : Orthogonality of the Coriolis force $\vec{F}c$ with the rotation vector $\vec{\Omega}$ and with the velocity vector $\partial\vec{u}/\partial t$.

In addition are two other forces which will not be considered here.

3. Velocity or frequency detection

The Coriolis force acts on each material particle of the solid in rotation and behaves as a body force. Following the mechanical equilibrium equation (linear momentum conservation)

$$\rho_o \frac{\partial^2 u_i}{\partial t^2} = C_{ijk\ell} \frac{\partial^2 u_k}{\partial x_j\, \partial x_\ell} + F_i \qquad (2)$$

where u_i is the mechanical displacement field, $C_{ijk\ell}$ the elastic constant tensor, ρ_0 the specific mass per unit volume, x_j the material coordinate reference system and F_i the body forces, the wave propagation equation under Coriolis acceleration becomes

$$\rho_o \frac{\partial^2 u_i}{\partial t^2} = C_{ijk\ell} \frac{\partial^2 u_k}{\partial x_j\, \partial x_\ell} - 2\,\rho_o\, \varepsilon_{ijk}\, \Omega_j \frac{\partial u_k}{\partial t} \qquad (3)$$

ε_{ijk} is the vectorial product operator.

$$\varepsilon_{ijk} = 0 \text{ for } i = j \text{ or } i = k \text{ or } j = k$$

and

$$\varepsilon_{ijk} = \pm 1 \text{ for } i \neq j \neq k$$

following a cyclic or an anticyclic permutation.

Applying a perturbation method [2, 3, 4] the corresponding fractional frequency change (or velocity) is

$$\frac{\Delta\omega}{\omega_o} = \frac{\int_V \varepsilon_{ijk}\, \Omega_j \frac{\partial u_k}{\partial t}\, u_i^* \, dV}{\omega_o^2 \int_V u_\ell \, u_\ell^* \, dV} \qquad (4)$$

where ω_0 is the angular frequency at rest. u_ℓ^* is the complex conjugate of u_ℓ. The integrations are performed in a given volume V. From Eq. (4) and from the properties of the operator ε_{ijk} it follows that

$\Delta\omega/\omega_0 = 0$

a) if Ω and u are colinear, i.e. if $j = i$ or $j = k$
b) if the vibration is rectilinear. This is equivalent to $i = k$
c) if the wave is a steady state wave.

Reversing the propagation direction of the wave changes the sign of the frequency shift. Since a steady state wave is composed ot two waves propagating in opposite directions the total effect is compensated. Therefore only travelling waves with elliptic mechanical polarization are sensitive to the Coriolis force.

Considering surface acoustic waves (Rayleigh waves) with general solution

$$u_i = \sum_r a_i^{(r)} e^{-j\beta_r x_2} e^{j\omega\left(t - \frac{x_1}{V}\right)} \qquad (5)$$

the relative frequency change takes the form

$$\frac{\Delta\omega}{\omega_o} = K \frac{\Omega}{\omega_o} \qquad (6)$$

where K is the Coriolis coefficient which was calculated for different materials and different crystal orientations and wave propagation direction [4]. In Table 1 and Table 2 are given the values of $|K|$ for various cuts of quartz and of lithium niobate.

Coupe	Ψ	θ	Φ	|K|
X	0	0	90	.138
Y	0	0	0	.103
Z	0	90	0	.054
ST	0	42.75	0	.136
AT	0	35.25	0	.138
FST	26	6.34	-41.5	.187
LST	-35	11.4	59.4	.148

Table 1 : Coriolis coefficient K for quartz (Rayleigh wave) Φ and θ are the cut angles. Ψ is the propagation direction [ref. 4]

Coupe	Ψ	θ	Φ	|K|
X	0	0	90	.070
X	90	0	90	.036
Y	0	0	0	.059
Y	90	0	0	.061
Z	0	90	0	.081
Z	90	90	0	.081
	0	-48.5	0	.095
	0	48.5	0	.384

Table 2 : Coriolis coefficient K for $LiNbO_3$ (Rayleigh wave) Φ and θ are the cut angles. Ψ is the propagation direction [ref. 4]

These tables show that the Coriolis coefficient is of the order of 0.1 (0.38 is the highest value). Considering surface acoustic wave oscillator with a frequency stability of the order of 10^{-8} the lowest Ω/ω_0 ratio which can be measured for $K = 0.1$ is $\Omega/\omega_0 = 10^{-7}$. For a typical SAW frequency of 100 MHz the lowest rotation rate which can be detected is 3 600°/sec. The only possibility to improve the resolution would be to reduce the frequency. This is not possible since for frequency much smaller than 100 MHz the size of SAW devices become too large. Only bulk waves with flexure modes present frequencies low enough to give a resolution better than 1°/sec, as it was proposed by Tiersten [5].

4. Amplitude detection

As shown in Fig. 1 the Coriolis force induces a mechanical displacement perpendicular to the plane of the rotation vector and the main mechanical displacement vector. Therefore if considering a flexure steady state wave and a rotation vector parallel to the propagation direction of the flexure waves, a second flexure vibration is induced, perpendicular to the previous one as shown in Fig. 2.

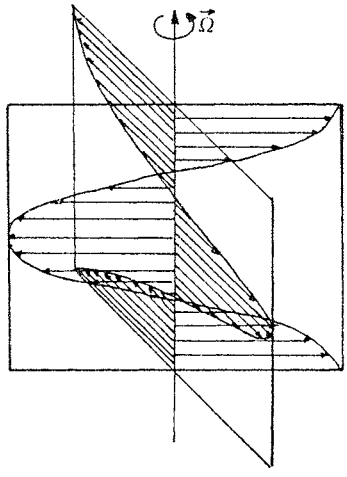


Fig. 2 : Main flexure and induced flexure for a rotation around the Ω axis

The main vibration mode is piezoelectrically driven at its resonance frequency in order to obtain the maximum amplitude, proportional to the resonator Q-factor and to the driving signal.

The induced vibration takes the form [6]

$$u_{induced} = u_{main} \frac{2\Omega}{\omega_o} \left| (1 - k^2)^2 + \frac{k^2}{Q_{ind}^2} \right|^{-1/2} \quad (7)$$

where k is the ratio between the driving frequency and the resonance frequency of the induced vibration

$$k = \frac{\omega_o}{\omega_{ind}} \quad (8)$$

A maximum of induced vibration amplitude will be obtained when both flexure vibrations are operated at their resonance frequency ($k = 1$) and

$$u_{induced} = u_{main} \frac{2\Omega Q}{\omega_o} \quad (9)$$

But a practical difficulty is to design the device in order to have identical resonance frequencies and to keep them identical under perturbations, temperature variations for instance. As a consequence it will be easier to operate the induced flexure vibration out of resonance. In the case $k < 1$

$$u_{induced} = u_{main} \frac{2\Omega}{\omega_o} \quad (10)$$

Different configurations are used for driving flexure mode vibrations. Since flexure modes are not pure modes, they can be obtained only by combining opposite length extensional modes or thickness shear modes.

a) Combining length extensional modes

The piezoelectric matrix, given in Table 3 shows that length or thickness extensional strains S_1 or S_2 can be electrically driven by means of the piezoelectric constants d_{11} or d_{12} by applying an electric field E_1.

d_{11}	d_{12}	0	d_{14}	0	0
0	0	0	0	d_{25}	d_{26}
0	0	0	0	0	0

Table 3 : Quartz crystal piezoelectric constants

- *flexure in the x_2x_3 plane*

Considering a bar with its length along the x_2 axis, the flexure is obtained by combining length extension and compression along the same axis (Fig. 3a).

The two opposite strains S_2 are obtained by applying in the crystal two opposite electric field E_1 ($S_2 = d_{12} E_1$) by means of electrode configurations as shown in Fig. 3b.

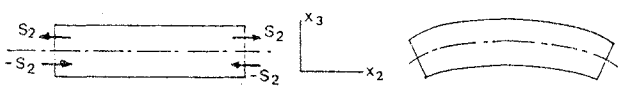


Fig. 3a : Flexure obtained by two opposite length extensions

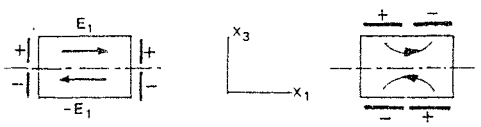


Fig. 3b : Electrode configurations for applying two opposite electric fields along the x_1 axis

- *flexure in the x_1x_2 plane*

The flexure is again obtained by applying two opposite E_1 fields. The two opposite fields are now distributed in the x_1x_3 cross section in the two zones separated by the x_3 axis (Fig. 4).

These fields are applied by means of four or six electrodes as shown in Fig. 4. These electrode configurations also give field components along the x_3 axis, which do not have any effect since quartz crystal is not piezoelectric along this axis (at least for crystal cuts which are not rotated).

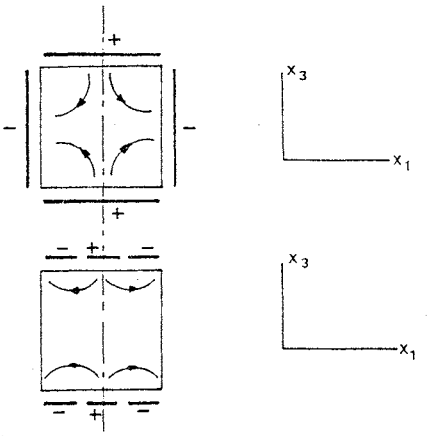


Fig. 4 : Electrode configuration for flexure in the x_1x_2 plane

b) Combining thickness shear modes

The second possibility for driving flexure modes is to apply to opposite thickness shear deformations. this can be done either by using the S_4, S_5 or S_6 strains by means respectively of the d_{14}, d_{25} or d_{26} piezoelectric constants and the E_1 or E_2 electric fields.

In Fig. 5 are shown two configurations using the S_4 and S_5 deformations. In Fig. 5a the opposite fields are along the x_1 axis by means of four electrodes. In Fig. 5b the strains S_5 are made by opposite electric fields along the x_2 axis.

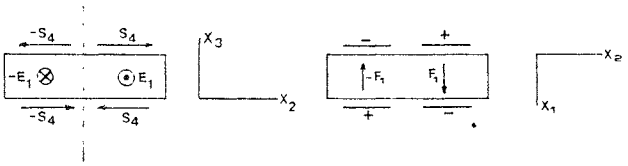


Fig. 5a : Flexure in the x_2x_3 plane obtained by two opposite thickness shear S_4 and corresponding electrode configuration for opposite E_1 fields

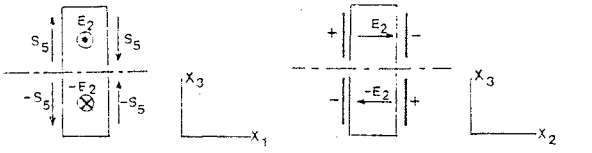


Fig. 5b : Flexure in the x_1x_3 plane obtained by two opposite thickness shear S_5 and corresponding electrode configurations for opposite E_2 fields

5. Gyrometer

Based on the previous principles a gyrometer was built composed of a quartz crystal bar with its length along the x_2 axis. The driving signal is applied by means of four electrodes to create opposite fields along the x_1 axis. These fields induce a flexure in the x_1x_2 plane as shown in Fig. 6. The rotation axis being oriented along x_2 the Coriolis force induces a flexure in the x_2x_3 plane.

One can consider that the induced x_2x_3 flexure is equivalent to two opposite S_4 thickness shear strains and to two S_2 opposite length extensions. The electrodes for detection were choosen to pick up the electric fields created by the S_2 strains (Fig. 6).

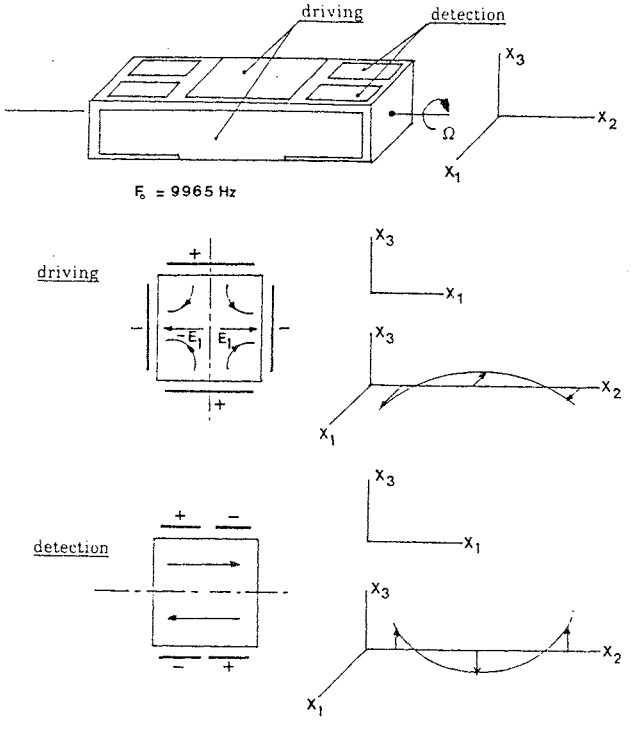


Fig. 6 : Flexure bar gyrometer configuration

The sensitivity of the gyrometer is given in Fig. 7. The dynamic corresponds to at least three orders of magnitude and the sensitivity is 20 mV/degree/sec. The resolution is limited by the background noise and corresponds to a rotation rate of 0.05 degree/sec.

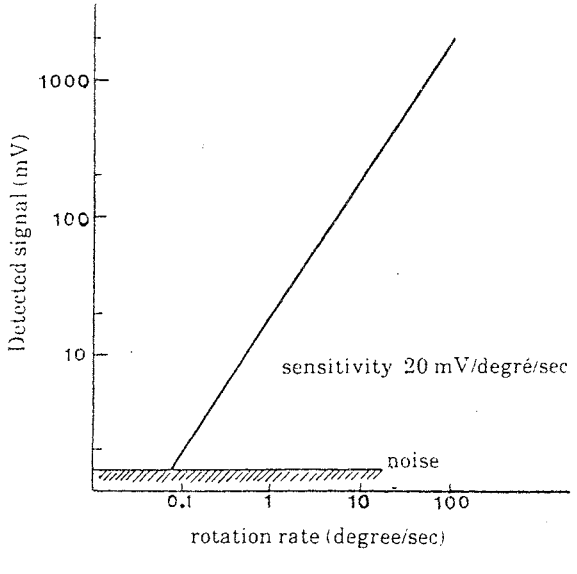
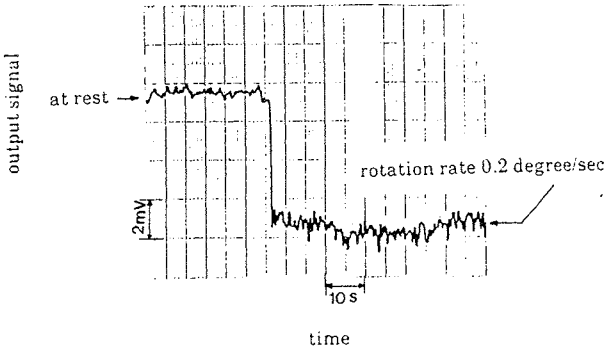


Fig. 7 : Sensitivity of the quartz crystal gyrometer

This means that it is possible to measure the rotation of a vehicle which makes a rotation of 360° in 2 hours (at least in laboratory environment).

The device was also tested in car in real conditions. As shown in Fig. 8 rotations of the order of 0.2 degres/sec were easily detected.



Background noise = 1 mV = 0.05 degré/sec

Fig. 8 : Response of the quartz crystal gyrometer in real conditions

6. Conclusion

The detection by piezoelectric effect of the amplitude of the vibration induced by the Coriolis force gives much better sensitivity that the detection of velocity or frequency variations. There are several reasons for that. The sensitivity of frequency detection is inversely proportional to the wave frequency. As a consequence only low frequency waves must be used, like flexure waves. However there is not frequency variation in steady state waves, i.e. in a resonator. Therefore the advantage of resonators, with high Q-factors is lost in frequency detection.

For these reasons amplitude detection was chosen. The highest sensitivity will be obtained when operating the main flexure vibration and the Coriolis induced flexure at the same resonance frequencies, but in that case an inherent difficulty will be to have and to keep the two resonances at the same frequencies, in particular as a function of temperature. The final choice was to operate only one flexure, the driving one at its resonance.

References

[1] G. Bruhat, "Mécanique", Masson 1961

[2] H.F. Tiersten, "Perturbation theory for linear electro-elastic equation for small fields superposed on a bias", J. Acoust. Soc. Am. 64 (3) sept. 1978.

[3] H.F. Tiersten, D.S. Stevens, P.K. Dias, "Acoustic surface wave accelerometer and rotation rate sensor", 1980 Ultrasonics Symposium Proc.

[4] Crochepierre, "Sensibilités des ondes élastiques de surface à l'accélération de Coriolis", Thèse de Doctorat, LPMO, Besançon, n° 477, 1985.

[5] H.F. Tiersten, D.S. Stevens, P.K. Dias, "Circulating flexural rotation rate sensor", 1981 Ultrasonics Symposium Proc.

[6] J. Söderkvist, "Piezoelectric beams and angular rate sensors", Proc. 44th Ann. Freq. Cont. Symp., 1990.

Session 6A Number 59

VISCO-ELASTICITY, PHASE TRANSITIONS AND ADHESIVITY MEASUREMENTS BY SHEAR HORIZONTAL ACOUSTIC WAVES

M. Hoummady and F. Bastien

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

1. Introduction

Usually interdigital transducers (I.D.T.) are used in order to produce surface acoustic waves (S.A.W.) but the same transducers can also generate bulk acoustic waves.

A device using shear bulk waves excited by interdigital transducers is presented here. Applications of the device for viscosity measurements, phase transition detection and adhesivity testing are demonstrated. Finally different operation possibilities of the present device will be discussed.

The device described here differs from previously described devices [1,2] in several points. The most important difference is the possibility to work with progressive waves (pulsed operation).

2. Device operation

The main part of the device is a quartz plate (Y cut) with interdigital transducers where the digits are parallel to the crystallographic X axis "x". In this orientation only shear horizontal waves can be generated [3,4,5]. The spatial transducer periodicity λ_0 is 34.4 µm. For this configuration surface skimming bulk waves are generated at 126 Mhz. Input and output transducers with 150 finger-pairs each (total length 5.16 mm, width 2.55 mm) are separated by 5.16 mm (see fig. 1), and quartz thickness is 1 mm. The previous description corresponds to the device effectively used to obtain experimental results presented in this paper. Obviously operating parameters can be modified, and several experiments with different devices of the same kind have been carried out.

A study of wave transmission in these devices has already been published [6,7]. Essential features will only be mentioned in order to understand device operation. Within a good approximation, one may consider the S.H. acoustic waves emitted from the transducers to be very close to plane waves. The received signal is a function of the overlap between the reflected acoustic beam and the output transducer. The reflection coefficient is practically equal to unity for the air-quartz interface, but with a liquid both effects of attenuation and phase shift appear. With a solid interface on the quartz the reflection coefficient depends on the transmission of waves into the solid deposited upon quartz surface. This transmission varies with properties of the solid and with adhesivity. Here the word adhesivity is used only to point out the macroscopic properties of the solid-quartz interface. Shear horizontal waves generated by the transducers can be produced continuously, in this case, the device operates in the plate mode regime. It is also possible to produce pulse emission with a long duration compared with signal period, but short compared with propagation time : for example 100 ns for a frequency of 190 Mhz. In this last case, it is possible to distinguish the contribution of electromagnetic signal from different acoustic signals corresponding to elastic waves travelling with different paths. On Fig 2 an example of output signals are shown. Different echos can be identified by comparison of theoretical and experimental propagation times.

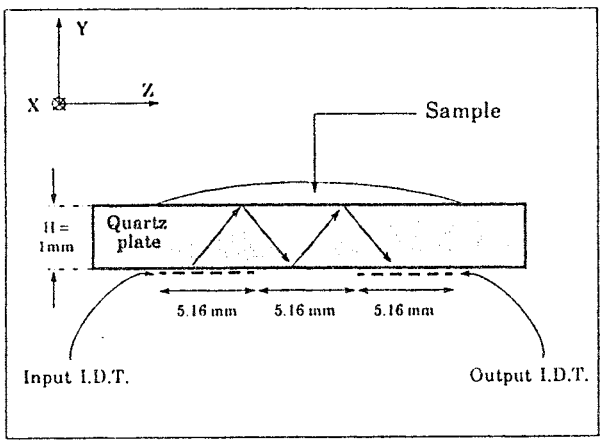


Fig. 1 : Experimental device

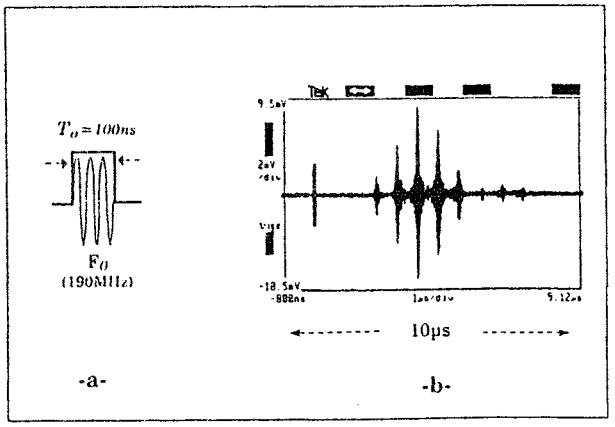


Fig. 2 : Typical input (a) and output signal (b)

On the output signal the electromagnetic signal and successive acoustic signals can be identified

3. Viscosity measurements

This first example will be presented in detail in other papers [8,9]. A summary of results obtained will only be given here. Measurements have been performed using pulsed operation. From the comparison of amplitude of echos with liquid on quartz or without liquid interface, calibration is obtained by reference to viscosity of different water-glycerol mixtures at a given temperature. The result is given as a curve $A = f(\eta)$ where $A = (1-R) \times 100$ and η is the viscosity coefficient obtained from available literature [10]. From the relation between pulsation ω and wave vector k in quartz and liquid, and from the continuity of displacements and stress at the interface, a theoretical value for the reflection coefficient R can be computed. Obviously the relation $\omega = f(k)$ in the liquid is obtained from the fluid propagation equation. The liquid is considered as visco-elastic therefore the equation is :

$$T + \tau \frac{\partial T}{\partial t} = \eta \frac{\partial S}{\partial t}$$

where T is the stress, S is the strain and τ the relaxation time. For sinusoidal waves the previous relation becomes $T = C_\ell . S$ with $C_\ell = j\omega\eta / (1 + j\omega\tau)$. Shear rigidity modulus μ is related to η and τ by $\tau = \eta / \mu$; μ can be considered almost as a constant. The value of μ has been adjusted to obtain the best fitting for large values of η. A relatively good fitting between theory and experiment is obtained for $\mu = 5.10^8$ N/m^2. Let us remark that for $\eta < 600$ centipoise the value of μ has no practical influence on the result, and therefore theoretical curves shown on Fig. 3 in the low value range of η can be obtained without adjustment. Futhermore values of $\mu = 5 \times 10^8$ N/m^2 and η = 60 centipoise = 0.06 N s m^{-2}, yield $\tau = 1.2 \times 10^{-10}$ s. With $f = 190$ Mhz, $\omega\tau = 0.143$ and therefore the simplified expression $C_\ell \sim j\omega\eta$ is valid.

This result explains the fact that calibration can be done using values of η given by literature at very low frequencies.

In conclusion it is possible to use S.H. acoustic waves to measure viscosity coefficients using very small quantity of liquid. The maximum value of η, previously discussed, depends on the range of frequencies used. The range of operating frequencies being itself determined by the I.D.T. dimensions.

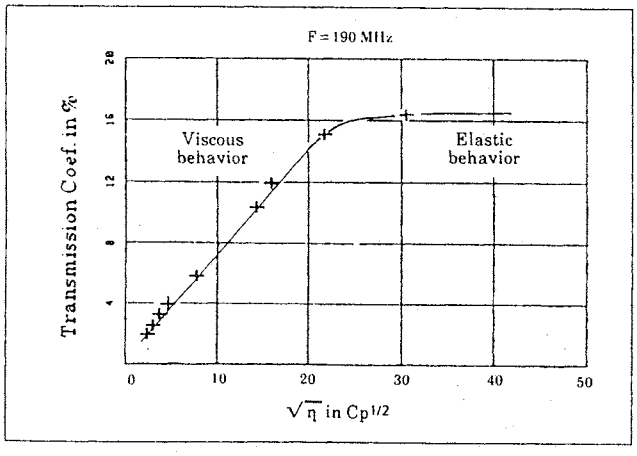


Fig. 3 : Transmission coefficient of the S.H. acoustic wave on the quartz-liquide interface.

+ : Experiment points
— : Theoretical points

4. Phase transitions study

In this application, the device previously described is used with harmonic excitation. S.H. waves propagate with several reflections on upper and lower sides of the plate without change of polarization. Interferences between incident and reflected waves produce S.H. plate mode. Such modes can be identified up to 2 or 3 times the cut off frequency (i.e. the frequency of surface skimming bulk waves). On Fig. 4, frequency response obtained with a frequency analyser lies between 180 and 200 MHz. Different modes can clearly be seen. Frequency interval between modes is about 1.4 Mhz.

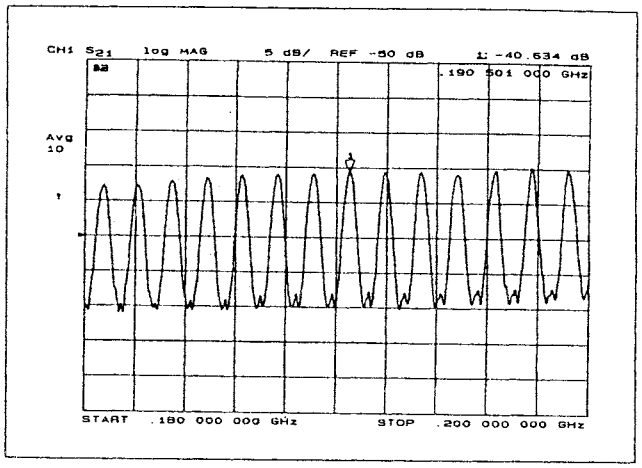


Fig. 4 : Device Transfert Function from 180 to 200 MHz.
Peaks correspond to different plate modes

The example given here corresponds to water-ice phase transition. Water is deposited on the uppon side of the quartz plate, covering about 1 cm^2. Temperature of the sample is measured by a thermoelectric couple (Chromel-alumel). A C.W. electrical excitation is generated by a frequency synthesizer at a frequency matched to the resonance of a plate mode ; here this frequency has been chosen at 190.5 MHz. The device is placed in a temperature-controlled chamber where output amplitude and water temperature are recorded at constant time intervals ($\Delta t = 50$ ms). Fig. 5a and 5b show the variations of insertion losses and temperature, recorded with demineralized water ; Fig. 6a and 6b give enlargements of previous figures. Different stages in the phase transition are demonstrated on both curves.

stage 1 Water temperature decreases down to -7.2°C. Water remains liquid and insertion losses stay at a constant value of - 41 db.

stage 2 Undercooled water crystallizes. Insertion losses rapidly increase for about 100 ms. These losses are due to the possibility of wave transmission into the ice. The wave being scattered on the upper surface of ice, a good approximation is to consider that almost all the energy transmitted into the ice will not be detected by the output transducer. On the enlargement of Fig. 6, this stage can be seen more precisely. From the slope of the curve crystallisation rate can be estimated at about 1 cm^2 for 0.12 s.

stage 3 Due to progressive freezing, temperature increases. Insertion losses decrease down to -44.7 dB as the variation of the ice-liquid water ratio at the 'quartz-liquid' interface. Heat produced during solidification is not immediatly evacuated and therefore produces a partial melting on the interface.

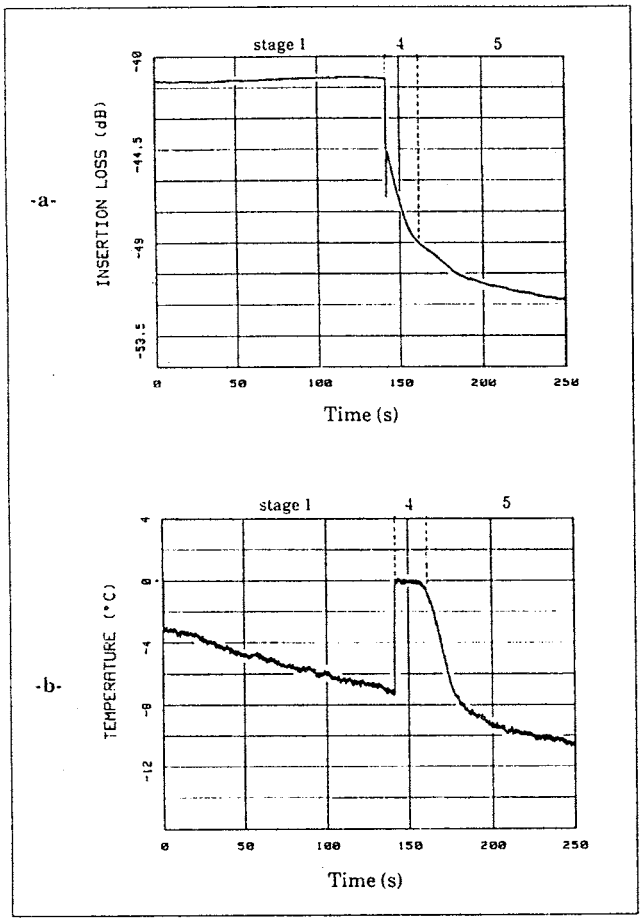


Fig. 5 : Demineralized water freezing

-a- acoustic plate mode attenuation (insertion loss)
-b- water temperature

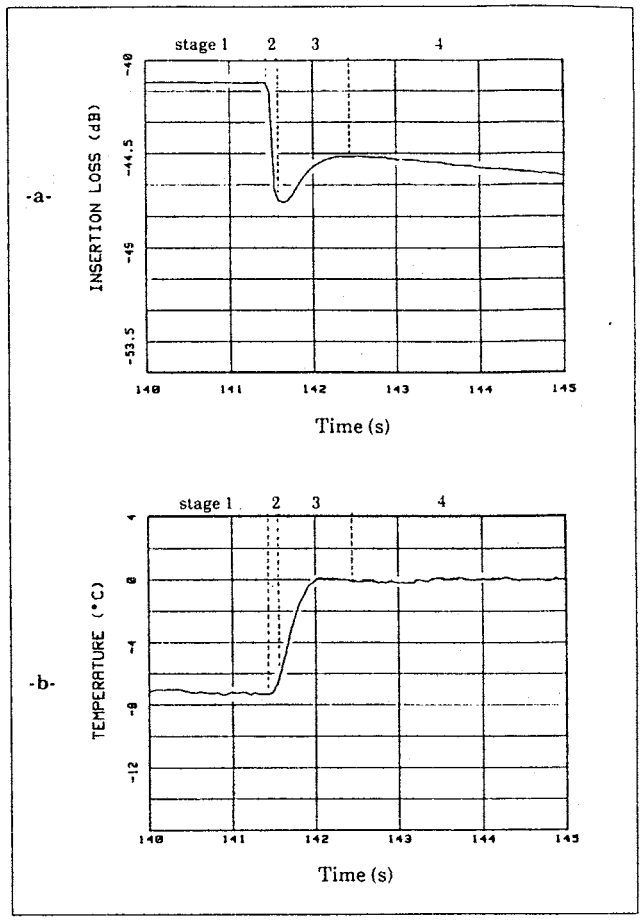


Fig. 6 : Enlargement of Fig. 5 around the transition

-a- acoustic plate mode attenuation (insertion loss)
-b- water temperature

stage 4 Insertion losses gradually increase. During this period liquid water freezes at 0°C. Temperature-controlled chamber is at about - 9 °C.

stage 5 Now only ice is present. The temperature of ice decreases from 0°C to the temperature of the controlled chamber. Insertion losses increase gradually toward -52 db.

Several other measurements show that the duration of the second stage is mainly a function of the difference T_0- T_s between the equilibrium phase transition temperature T_0 (0 °C for water) and the minimum undercooling temperature T_s. As it is well known, T_s is a function of water purity, quantity of water and external mechanical disturbances. [11,12].

S.H. acoustic plate modes excited by I.D.T. are a very useful tool to study phase transitions. The present device can be used for pratical applications like determination of icing conditions, or in order to understand the bond between solid and quartz. Futhermore it is possible to deposit a thin film of various materials on quartz in order to study the bond between film and solid.

5. Polymer adhesivity study

In this section the word 'adhesivity' is used to qualify the contact between polymer and quartz, but does not refer to a precise mechanism concerning the bond between materials. Let us describe the test in the case of polystyrene. A solution of polystyrene in toluen is deposited on a plane surface, and toluen evaporates. Obviously if the deposit is done directly on quartz the 'adhesivity' will be good but here the deposit has to be done on a surface of about 3 cm^2 on a 100 g mass. After toluen evaporation the polystyrene is pressed between the 100g mass and the upper face of the quartz device described in paragraph 3. The amplitude A of acoustic signal is measured as the temperature varies from 25°C to 105°C. The curves $A = f(t)$ present interesting features. Qualitatively these curves present large variations with different polystyrene preparations but the shape of the curve remains the same. In fact the properties of the polystyrene surface can be very different from one test to other. Figure 7 shows a typical curve. At the beginning of the cycle (point A) the adhesivity is very weak. The acoustic signal is almost the same with or without the presence of polystyrene and mass. From point A to B the heating induces an increase of adhesivity, therefore the amplitude of the signal decreases. From B to C, cooling does not change the adhesivity. Coming back to the temperature T_0 (point C) the polymer layer can be pulled out, then the amplitude of signal comes back to his initial value. Let us observe that with the mass but without polystyrene the signal amplitude variation between 25°C and 105°C is less than 6 %.

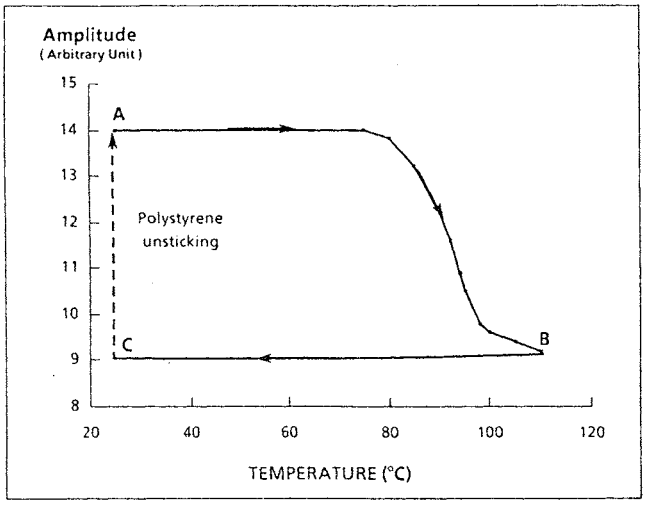


Fig. 7 : Amplitude of reflected S.H. wave on quartz-polystyrene interface during a temperature cycle

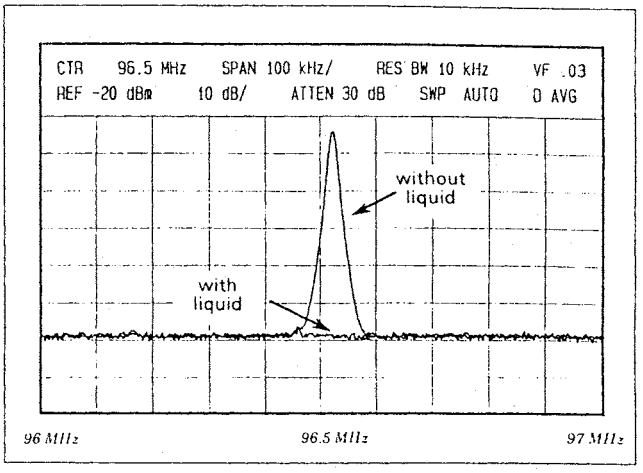


Fig. 9 : Spectral responses of the liquid detector (Fig. 8) with and without liquid

With atactic polypropylene, a signal attenuation at ambiant temperature is obtained. This attenuation depends on the pressure exerted on the polypropylene sample but can easily reach 40 %.

As a conclusion, the attenuation of the acoustical signal as temperature varies can produce a characteristic cycle. Additional tests with different materials are required in order to find a quantitative relation between this cycle and adhesivity.

6 Other possibilities and conclusion

Many other possibilities exist. Let us give only one other example. An oscillator has been built, using the previously described resonator (see Fig. 8). The system oscillates on a frequency corresponding to a plate mode chosen here at 96.52 Mhz (BT cut). Attenuation increases with a liquid deposited on quartz. If the loop gain is adjusted in order to maintain oscillations without liquid, but if excess losses stop the oscillator in the presence of liquid, a detector of liquid is designed. On figure 9, the frequency spectrum of the oscillator with or without liquid is presented.

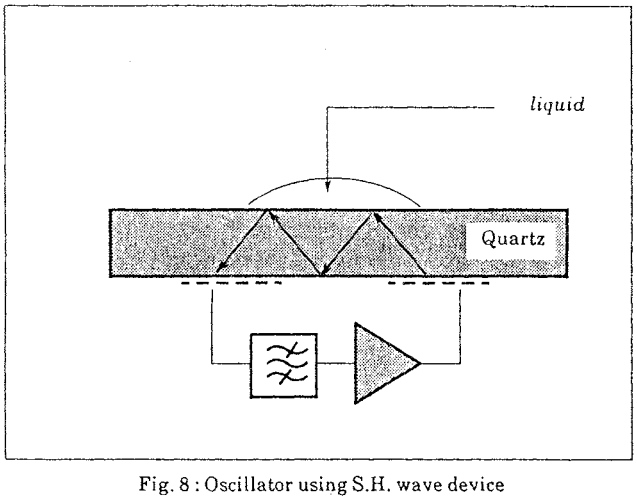


Fig. 8 : Oscillator using S.H. wave device

Comparison between different modes operation (i.e., C.W. operation and pulsed operation) can now be summarized.

pulsed operation
- advantages : It is possible to distinguish between acoustic and electromagnetic signals. The wave attenuation is only sensitive to the quartz-liquid interface and not to the bulk properties of quartz.
- disadvantages : The measurement corresponds to an attenuation, and therefore is relatively inaccurate. This difficulty can be partially overcome using a pair of matched devices for a differential measurement between a region with absorbing materials and without absorbing materials.

continuous operation (plate mode)
- advantages : The sytem is more simple. Furthermore it is possible to build an oscillator therefore the characteristic signal becomes a frequency which is more accurate and practical.
- disadvantages : the device is not only sensitive to the quartz-liquid interface properties but also to the bulk properties of quartz, for example the temperature variations.

In conclusion, devices using shear horizontal waves produced by I.D.T. can be used for various measurements and detection of different effects. Viscosity and phase transition applications as presented here can be developed immediately. Other applications are promising but will require further investigation.

Acknowledgements

The authors wish to thank E.Bigler from LPMO Besançon and B.Viard and G.Cudey from university of Besançon for helpful discussions, R. Coquerel, C. Bonjour and G.Martin from LPMO Besançon for acoustic devices preparation.

Bibliography

[1] A.J. Ricco and S.J. Martin, Appl. Phys. Lett. **50**, 1474-76 (1987).

[2] R.C. Hugues, S.J. Martin, G.C. Frye and A.J. Ricco, Sensors and Actuators, A21-A23, 693-699 (1990).

[3] M. Lewis, IEEE Ultrasonics Symp. Proc. pp. 744-752 (1977).

[4] C.N. Welmick, D.J. White and K.M. Lakin, IEEE Ultrasonics Symp. Proc., pp. 280-285 (1981).

[5] D.L. Lee, IEEE Trans. on Sonics and Ultrasonics SU-27 (1), 22-30 (1980).

[6] M. Hoummady, B.A. Auld, D. Hauden and E. Bigler, 1er Congrès Français d'Acoustique, Lyon, avril 1990, Proc. vol. 2, pp. 1245-1248 (1990).

[7] M. Hoummady thesis n° 202 (1991), Université de Franche-Comté-Besançon (France).

[8] M. Hoummady and F. Bastien, Acoustic wave viscometer, accepted at Review of Scientific Instruments.

[9] M. Hoummady and F. Bastien, Acoustic wave viscometer, 5th Int. Exhibition and Congress for Sensor & Systems Technology, Nuremberg (RFA), may 1991.

[10] J.B. Segur and H.E. Oberstar, Industrial and engineering chemistry **43**, 2117-2120 (1951).

[11] F. Franks, *Water a comprehensive treatise. Vol. 1*, The physics and physical chemistry of water, Plenum press, N.Y., (1972)

[12] L. Babin Thesis n° 96 (1966), Université de Bordeaux (France)

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 6A Number 60

The State-of-Art Electrostatic Field Calculation of Surface Acoustic Wave Devices

A.R. Baghai-Wadji, H.P. Zidek, H.P. Reichinger and B.H. Jakoby

Technische Universität Wien, Institut für Allgem. Elekt. und Elek.
Gußhausstr. 27-29, A-1040, Wien, Austria

Abstract: In previous work we have shown that Fourier integral transform, in conjunction with Green's function concept and the method of moments, can be succesfully applied to two- and three-dimensional representation of surface acoustic wave (SAW) devices in order to calculate efficiently and precisely the charge distribution on the electrodes and on the bus-bars.
The purpose of this paper is to discuss manifold extensions of our previous work. (i) Electrodes with arbitrary cross-section and geometries are included in the two- and three-dimensional analysis. (ii) The influence of the metallic SAW-package on the charge distribution is taken into account. (iii) Potential distribution patterns are presented in 3D and equipotential lines.

I. Introduction

I.a Peculiarities of SAW Devices: In the last two decades considerable attention has been directed towards the *fast* and *precise* calculation of charge distribution in surface acoustic wave (SAW) devices [1]-[20]. The calculation of the charge (source) distribution in SAW devices is a typical indirect problem, in contrast to the direct problems where the source functions are known and the resulting field distributions must be determined. Both cases have their relevance in SAW modeling (generation, detection). However, direct- and indirect- field problems are boundary value problems, for which powerful analytical, and in recent times, a variety of numerical and semi-numerical concepts have been developed. Thus, let us ask what, from a field theoretical point of view, are the peculiarities of SAW devices? And what are the resulting consequences for modeling?
First of all, the vast number of electrodes in these devices makes them different from other microwave components. In cases where the topology of the electrodes and the mutual spacing are regular, the large number of electrodes suggests methods developed for problems with periodic boundaries; i.e. methods which rely on Floquet's theorem. In cases where the regularity and the periodicity of the electrodes are slightly perturbed, iterative techniques can be employed to calculate the charge distribution. However, the efficiency of iterative methods strongly depends on the initial solutions. An excellent initial solution is, however, available from the corresponding (non-perturbed) solution. The reasons for the above-mentioned deviation from the periodicity are twofold. Firstly, electrode widths and positions can be seen as parameters to variate and to achieve a desired frequency characteristic of the device. Secondly, for the purpose of (partially) compensating undesired, so-called *second order effects*, the electrode geometry and their positions can be changed. When there is no regularity at all, as is the case in aperture weighted transducers, then the boundary conditions are generally extremely complicated, and a serious analysis demands an unacceptable amount of computer resources.

A second property of SAW devices is the plane surface and plane substrate backside. Some SAW devices employ different layers separated by plane interfaces. Even in a SAW device in its simplest realization with only a piezoelectric substrate plate, if we want to model the influence of the package on the charge distribution, and perhaps the influence of an adhesive layer, then we have to deal with a two- and three- layers system. However, the fact that the interfaces are parallel planes eminently facilitates the analysis. Integral techniques, preferably Fourier's integral transform, can be applied.

Thirdly, piezoelectric substrates are strongly anisotropic. However, by using integral transform techniques, the governing partial differential equations can be transformed routinely into algebraic equations which are naturally more complicated than in isotropic cases, but the complexity is of an analytical type and should cause no serious conceptual difficulties.

And finally, it should be noted that, in most applications, SAW devices can be modeled as linear devices with the great advantage of a linear calculation; i.e. superposition. Non-linear problems are preferably solved by iterative techniques; so in these cases a solution for the linearized problem is thought to offer an acceptable, perhaps even an excellent, initial solution.

I.b Motivation for Electrostatic Field Analysis: Let an interdigital transducer generate acoustic waves at a certain frequency ω_0. Depending on ω_0 the resulting piezoelectric fields are complex-valued. *Second order effects*, such as finite conductivity and the finite mass density of the metallic electrodes, make the physical picture even more complicated. At higher frequencies (in GHz range) electromagnetic radiation becomes relevant. The coupled piezoelectromagnetic-wave phenomena must be analyzed. Under these severe complications, a question arises: what has an electrostatically calculated charge distribution to do with an actual dynamic charge distribution? To realize the relevance of the electrostatic calculations in SAW devices, let us consider the most complicated piezoelectromagnetic problem. Here, we must deal with the system of Maxwell's equations coupled with Newton's equations of motion, subject to some imposed boundary conditions. Both in electromagnetic, as well as in piezoelectric problems and thus in coupled

piezoelectromagnetic cases, the electrostatic solution serves as asymptotic solution in the low frequency limit. In many applications, even up to several hundred MHz, comparison with experiments shows that the correct electrostatic solution leads to excellent simulation results. In cases where the electrostatic solution is not satisfactory, as discussed above, the electrostatic field can be regarded as an acceptable initial solution for an iterative solution scheme.

In the following, after a preparatory discussion, we present a representative selection of our recent calculations. We first discuss the two dimensional problems, and then the three dimensional ones. For a more detailed and comprehensive discussion the reader is refered to [14]-[17] (see also [18]-[20]).

II. Method of Analysis

The general features of the underlying analysis technique can be outlined in a few words. However, to make the discussion easier let us specify the problem more precisely.

II.a Statement of the Problem: Consider a metallic case which encloses a SAW device, consisting of a number of metallic electrodes on a substrate. In an electrostatic approximation, we assume the substrate to be a dielectric, with permittivity constants of the dielectric measured at a constant stress. This facilitates, in a phenomenological way, consideration of the piezoelectricity of the substrate, Fig.1.

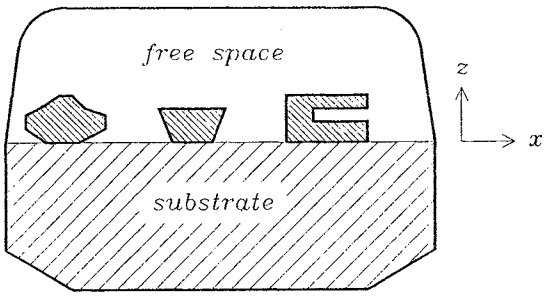


Fig.1 SAW device enclosed in a metallic package

Between the substrate back side and the package is assumed an adhesive material characterized by some dielectric constant. Above the substrate we assume free space. The electrodes are taken as metallic bodies on the substrate surface. Now assume that some of electrodes are driven by a sinusoidal voltage source. Some other electrodes are disconnected (floating electrodes). The package is assumed to be grounded. The problem is to find the field distribution within the package. In this specific case (electrostatic limit), the field variables can be derived from a scalar potential function. In the system of Maxwell's equations we consider the charges as the source for the electric field. According to the last two statements, another formulation of the problem will be as follows: determine the charge distribution on the electrodes so that the resulting potential function (a) takes the given constant values on the driven electrodes, (b) becomes zero on the package and on the grounded electrodes, and (c) assigns to the floating electrodes constant potentials so that the resulting charge integral on the individual floating blocks yields zero.

II.b Solution Procedures: (i) Let us try to solve this problem by a method of trial and error. Estimate an initial solution for the charge distribution and calculate the resulting potential function. Determine the difference between the calculated and the given potential values on the metallic regions. If this difference (residual) exceeds a prescribed tolerance, the initial guessed solution must be altered and the procedure repeated until, hopefully, after a few *iterations* the residual reduces to an acceptable level (the method converges). (ii) An alternative, and as we believe, a more natural way of solving field problems relies on energy minimizing variational schemes (finite element method). (iii) As a third possibility for attacking the problem at hand, we mention the powerful *method of weighted residuals*, also called *method of moments, MoM*, [21]-[23]. In contrast to the first method, this method is a direct non-iterative method. All the results presented in this paper are achieved by MoM application. MoM is a projective method and can be applied very successfully to an integral equation, to generate an approximate numerical solution. By assuming that the exact solution can be described in a complete infinite dimensional function space, we can work only in a finite dimensional function space. Thus we decide on a subspace and construct an approximation for the charge distrbution, by using a sum of the functions from the subspace, weighted by a priori unknown coefficients. Therefore, the problem now is the determination of the coefficients which enter into the sum. Analogous to the first method, we determine the resulting residual. Then we project the residual onto the coordinates of our subspace. We finally determine the unknown coefficients by requiring the projections to be set to zero. It is obvious that to perform the afore-mentioned steps, we have to define a scalar product and a norm.

II.c Green's Functions: The MoM, as applied to our problems, requires an integral equation which establishes a relationship between the source- (electric charge) and the response- (electric potential) function. At this point we remember the linearity of our problem. We consider the whole charges as a superposition of line- or point- charge sources, respectively, depending on the dimensionality of our problem. Similarly, we regard the potential distribution as the superposition of the potentials due to the individual line- or point-charges. In relating the charge- and the resulting potential- distribution in a physically realizable way and compatible with the imposed boundary conditions, a system-characterizing function exists (at least in a weak sense). This system's immanent, auxiliary function is called Green's function, related to the boundary value problem. Depending on the nature of the problem, Green's function can be scalar, vector, or dyadic. Green's function is the field response of the medium under consideration to a Dirac's delta function excitation (point-, or line- source).

In the next section we present some representative results concerning two and three dimensional analysis.

III. Results

Using basis- and weighting- functions as shown in Fig.2, we now present some representative results.

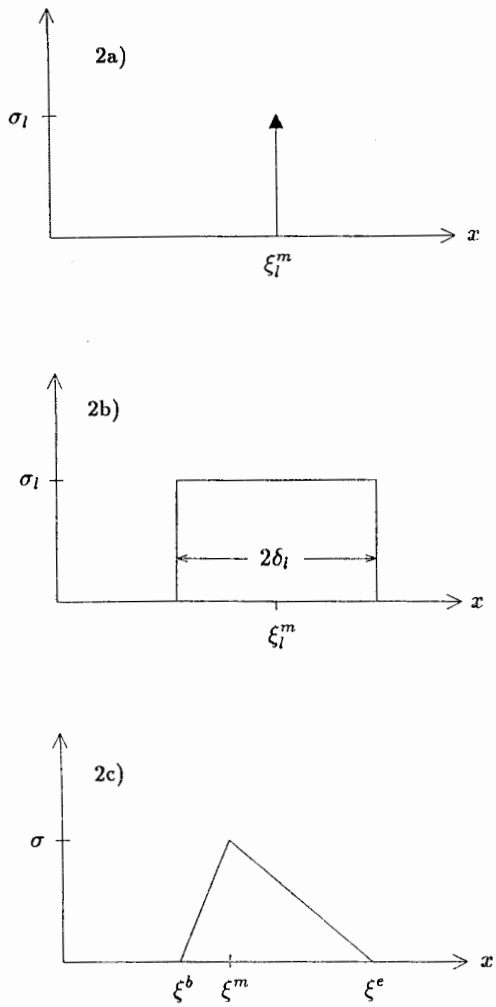
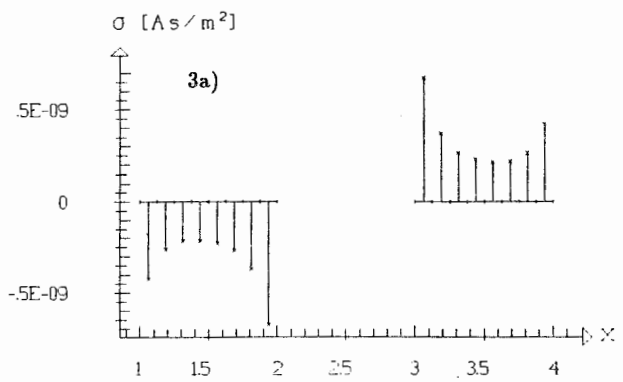


Fig.2 Basis- and weighting- functions. a) delta-function, b) pulse-function, c) rooftop-function

Fig.3 shows typical one-dimensional charge distributions, using afore-mentioned basis functions.



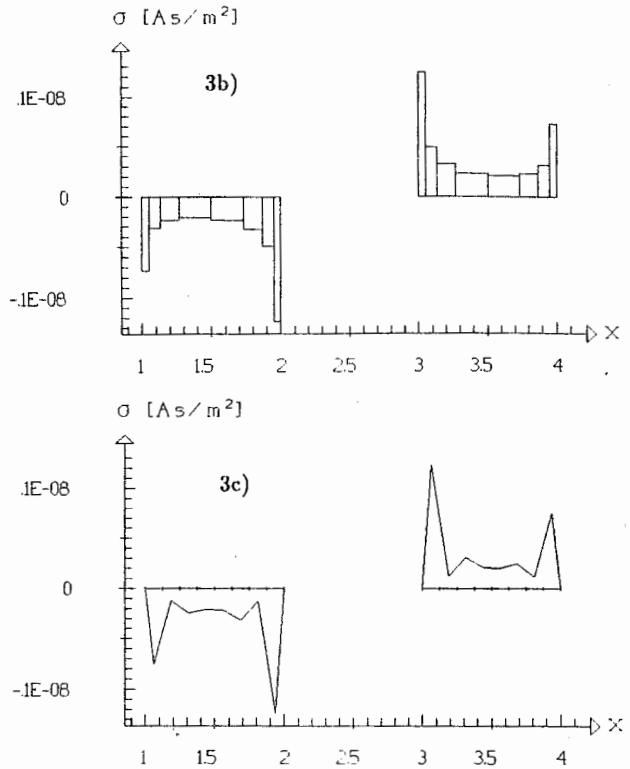
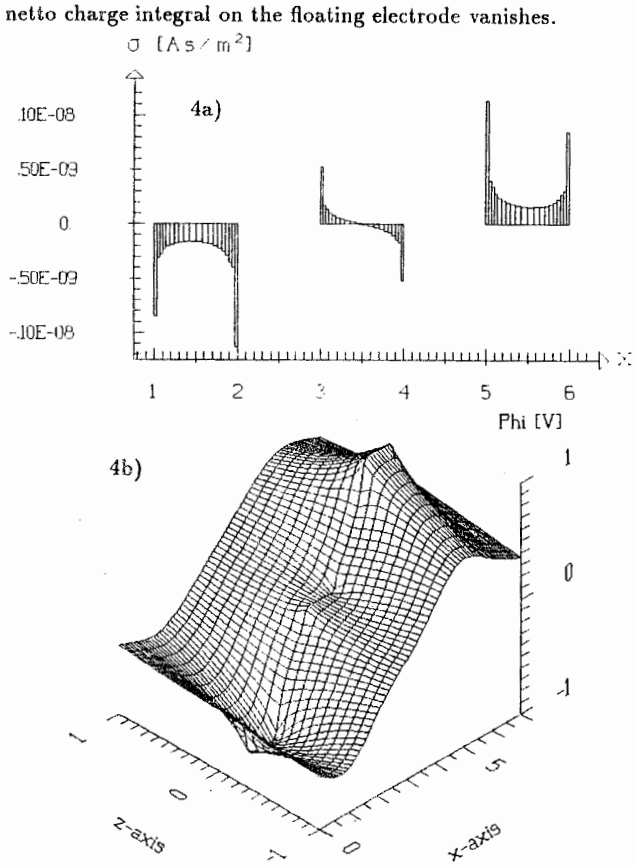


Fig.3 One-dimensional charge distribution using different basis- and weighting- functions. a) delta-function, b) pulse-function, c) rooftop-function

Employing a non-equidistant discretization, in Fig.4, the charge distribution, the equipotential lines and the potential distributions are presented. In this example the middle electrode is a disconnected, electrically floating electrode. Note that the netto charge integral on the floating electrode vanishes.



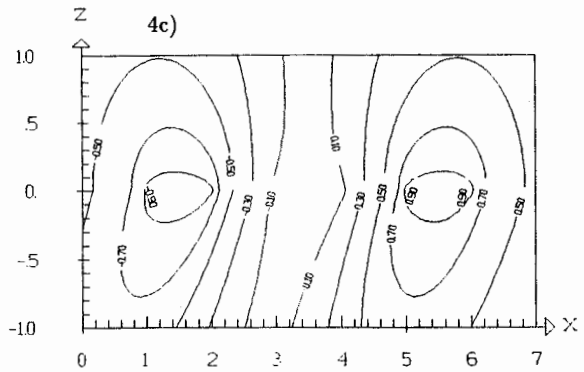


Fig.4 Non-equidistant discretization. a) charge distribution, b) potential distribution in 3D representation, c) potential distribution in equipotential lines

To demonstrate the generality of our method, in the following figures, the results concerning electrodes with rather arbitrary cross-section are showm.

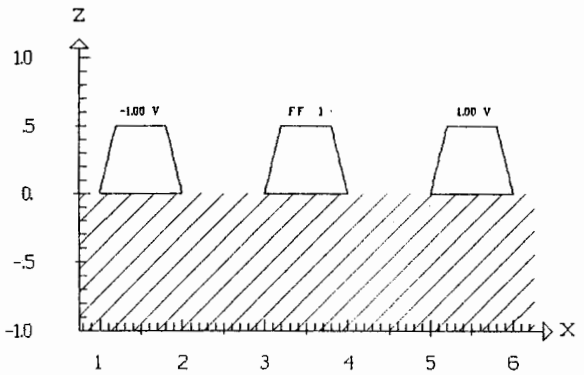


Fig.5 Electrodes with arbitrary cross-section

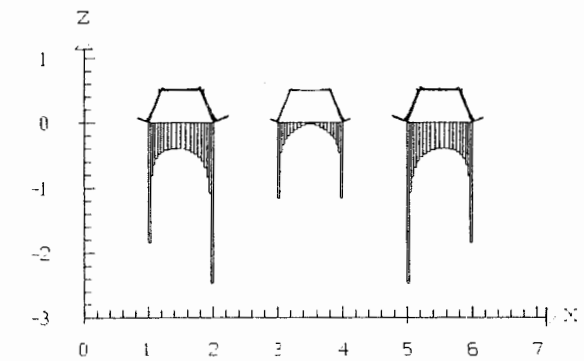


Fig.6 Magnitude of the charge density on the bounding surface of the electrodes

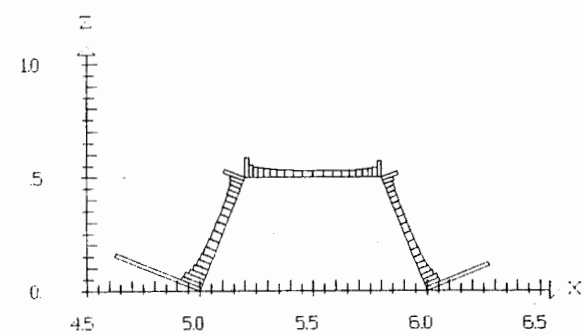


Fig.7 Magnitude of the charge density on the bounding surface of the most right electrode in Fig.6

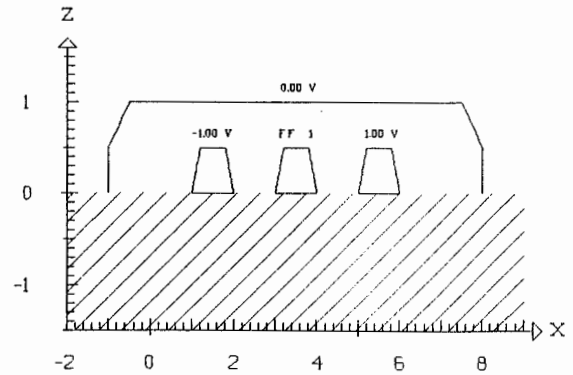


Fig.8 Grounded package over the electrodes

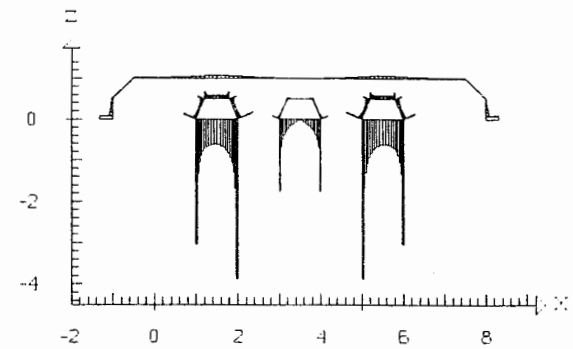


Fig.9 Charge distribution on the electrodes and on the package.

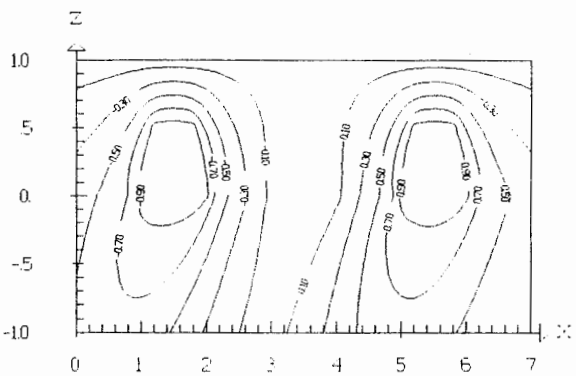


Fig.10 Equipotential lines

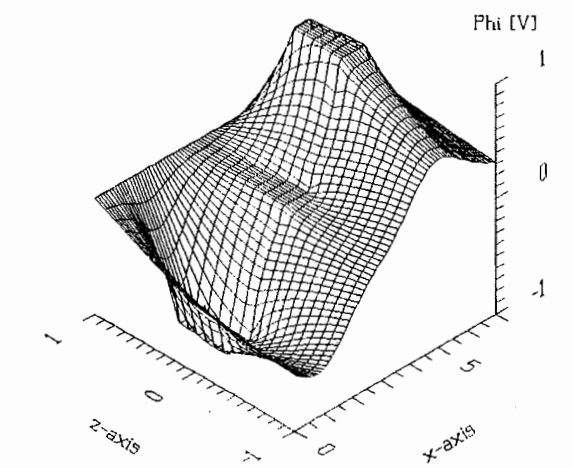


Fig.11 Potential distribution in 3D representation.

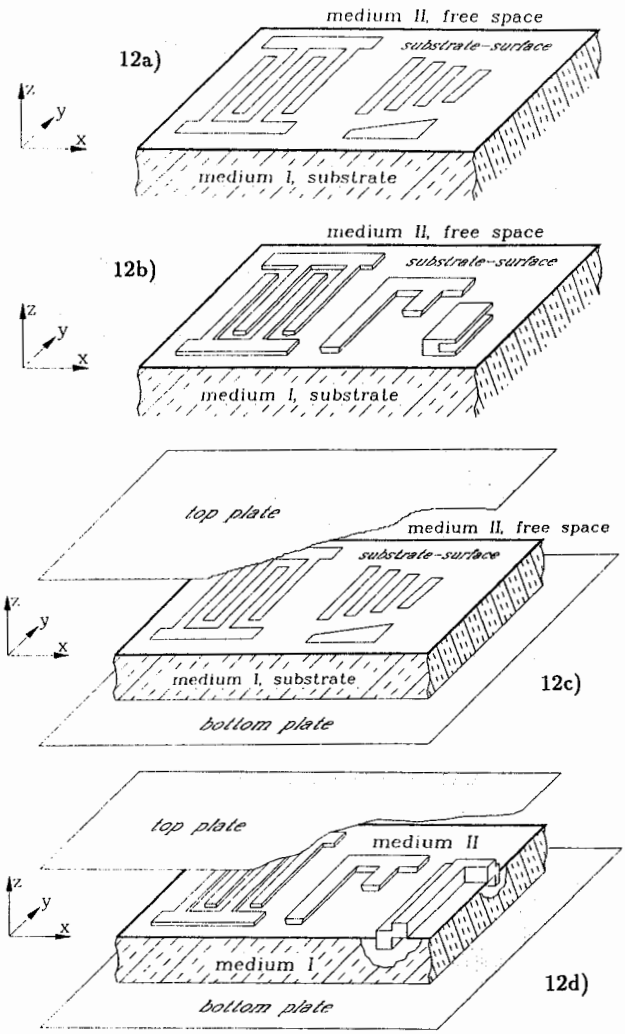
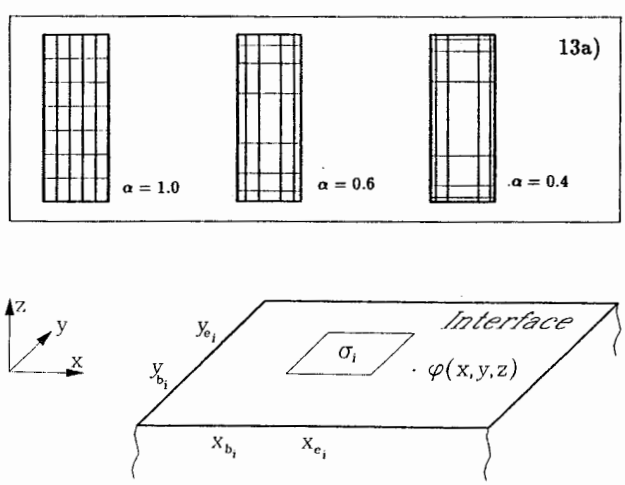


Fig.12 Different 3D models for SAW devices. a) half-space substrate with infinitely thin electrodes, b) half-space substrate with thick electrodes, c) substrate with finite thickness, infinitely thin electrodes and shielded by a lower and upper grounded plate, d) the same as in c) but with thick electrodes



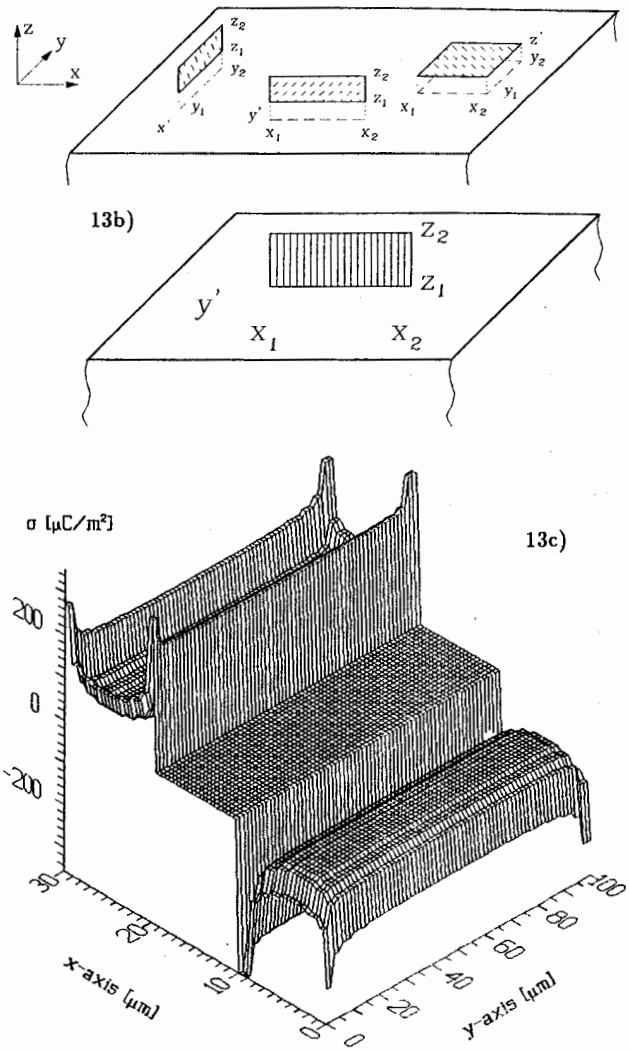
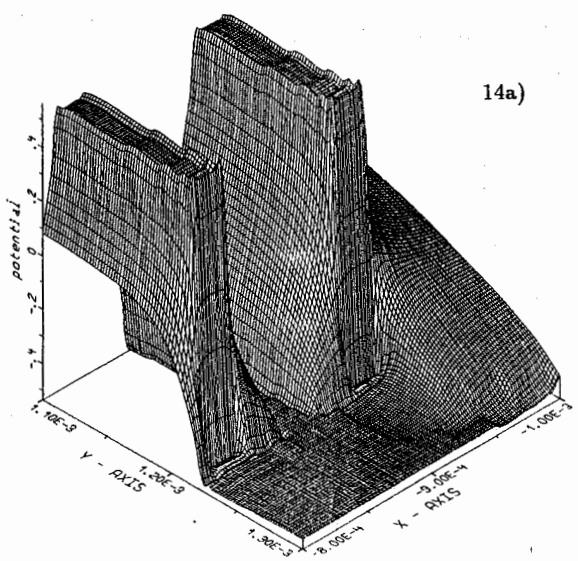


Fig.13 a) Equidistant and non-equidistant discretization, b) subelements for the calculation of the charge distribution, c) charge distribution on a simple structure



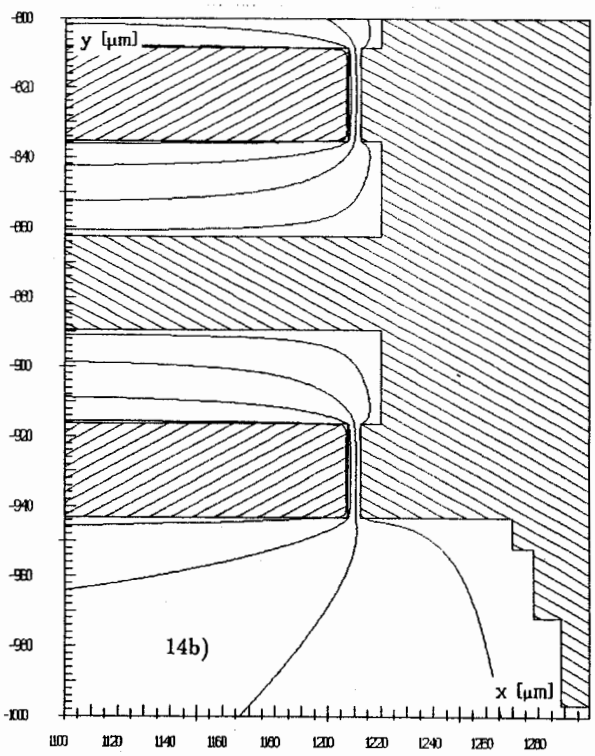


Fig.14 a) A typical potential distribution in 3D representation
b) in equipotential lines

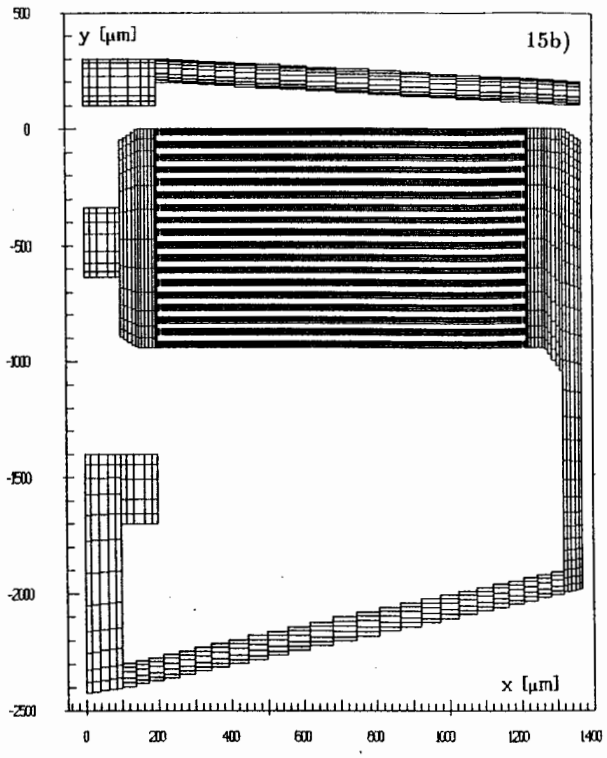
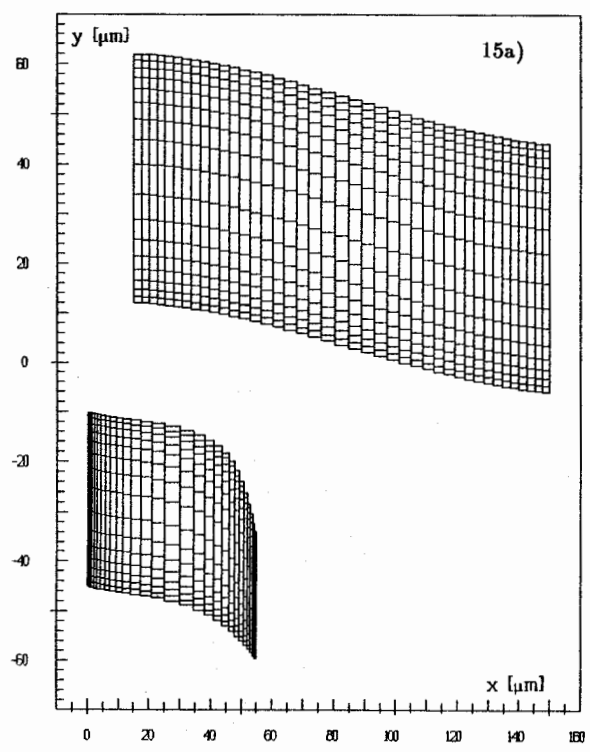
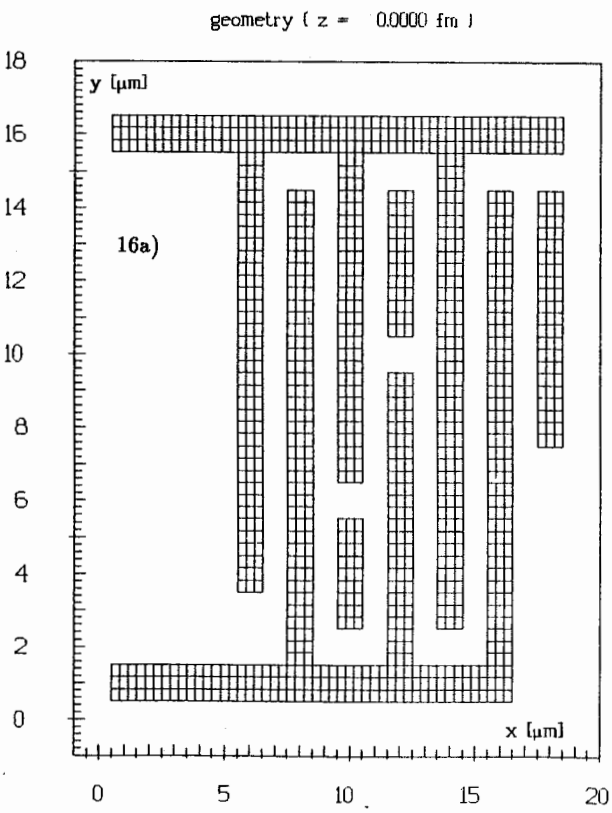
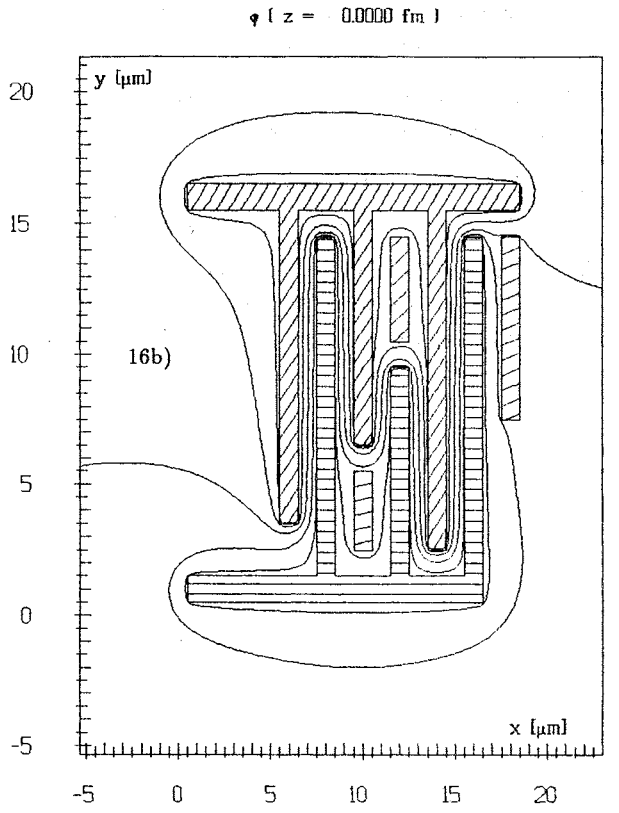


Fig.15 a,b) Examples for non-equidistant discretization







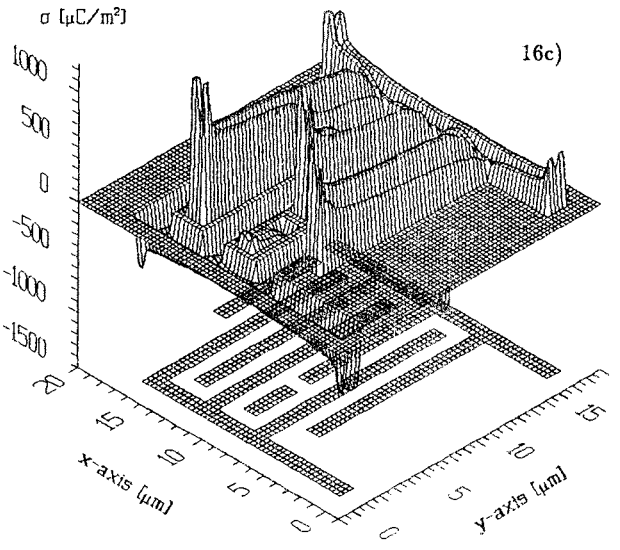


Fig.16 a) A transducer with finite aperture length and floating electrodes, b) corresponding potential distribution in equipotential lines, c) charge distribution

Conclusion

Representative calculations from our two- and three-dimensional field analysis are presented. The results show the generality of the method of moments, combined with Fourier transform and the concept of Green's function.

Most of the formulae are carried out in closed form. As a consequence of this, together with a non-equidistant discretization scheme of the electrodes, the required computation time and the computer memory, even for three dimensional problems are acceptable.

Finally, it should be mentioned that a further natural extension of our method makes possible to comprise the effects of the electromagnetic compatibility and the finite conductivity in SAW devices. It can be shown that the same solution scheme remains valid.

Acknowledgements

The work of A.R.B.W. was carried out partially under a Kurt Gödel grant, Ministry of Science and Research, Austria, whose support is gratefully acknowledged. The authors wish to thank Prof. F. Seifert and Dr. C. Ruppel for continuous encouragement and support.

References

[1] H. Engan, "Excitation of Elastic Surface Waves by Spatial Harmonics of Interdigital Transducers", IEEE Trans. Electron Devices, vol. ED-16, pp.1014-1017, 1969.

[2] C.S. Hartmann and B.G. Secrest, "End Effects in Interdigital Surface Wave Transducers", IEEE Ultrasonics Symp. Proc., pp.-413-416, 1972.

[3] R.C. Peach, "A General Approach to the Electrostatic Problem of the SAW Interdigital Transducer", IEEE Trans. Sonics and Unltrason., vol., SU-28, no.2, pp.96-105, 1981.

[4] V.M. Ristic and A.M. Hussein, "Surface Charge and Field Distribution ina Finite SAW Transducer", IEEE Trans. Microwave Theory Techniq., MTT-27, no.11, pp.897-901, 1979.

[5] S. Datta and B. Hunsinger, "Element Factor for Periodic Transducers", IEEE Trans. vol. SU-27, no. 1, pp.42-44, 1980.

[6] R.S. Wagers, "Transverse Electrostatic End Effects in Interdital Transducers", IEEE Ultrasonics Symp. Proc., pp.536-539, 1976.

[7] A.R. Baghai-Wadji, S. Selberherr and F. Seifert, "On the Calculation of Charge, Electrostatic Potential and Capacitance in Generalized Finite SAW Structures", Proc. IEEE Ultrason. Symp., pp.44-48, 1984.

[8] A.R. Baghai-Wadji, S. Selberherr and F. Seifert, "Closed-Form Electrostatic Field Analysis for Metallic Comb-Like Structures Contatining Single and Interconnected Floating Strips of Arbitrary Topological Complexity, part I: One-Dimensional Representation, part II: Two-Dimensional Representation", in Proc. IGTE Symp., Graz, Austria, pp.138-145 and pp.146-153, 1986.

[9] —, "A Green's Function Approach to the Electrostatic Problem of Single, Coupled and Comb-Like Metallic Structures in Anisotropic Multilayered Media", Proc. AMSE, vol.2.1, pp.109-120, Sorrento, Italy, 1986.

[10] —, "Two-Dimensional Green's Function of a Semi-Infinite Anisotropic Dielectric in the Wavenumber Domain", IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Controll, vol. UFFC-33, no.3., May 1986.

[11] A.R. Baghai-Wadji, S. Selberherr and F. Seifert, " Rigorous 3D Electrostatic Field Analysis of SAW Transducers with Closed-Form Formulae", Proc. Ultrasonics Symposium, vol.1, pp.23-28, Williamsburg, VA, 1987.

[12] H. Bachl and Baghai-Wadji, "3D Electrostatic Field Analysis of Periodic Two-Dimensional SAW Transducers with Closed-Form Formulae", in Proc. IEEE Ultrason. Symp., pp.359-362, 1989.

[13] A.R. Baghai-Wadji, O. Männer, and R. Ganß - Puchstein, "Analysis and Measurement of Transducer End Radiation in SAW in SAW Filters on Strongly Coupling Substrates", IEEE Trans. MTT, vol. 37, no. 1, pp.150-158, 1989.

[14] A.R. Baghai-Wadji, "PhD Dissertation", Technische Universität Wien, 1987.

[15] B.H. Jakoby, "Master Thesis", Technische Universität Wien, 1991.

[16] H.P. Reichinger, "Master Thesis", Technische Universität Wien, 1991.

[17] H.P. Zidek, "Master Thesis", Technische Universität Wien, 1991.

[18] B.H. Jakoby, A.R. Baghai-Wadji and C. Block, "Interaction of Electromagnetic Waves with Finitely Conductive Cylindrical Electrodes of Arbitrary Cross-Section", to be published in Proc. melecon'91, Ljubljana, Yugoslavia, 1991.

[19] H.P. Reichinger and A.R. Baghai-Wadji, "Rigorous Electrostatic Field Analysis of Surface Acoustic Wave Devices with Closed-Form Formulae, A Two-Dimensional Representation", to be published in Proc. melecon'91, Ljubljana, Yugoslavia, 1991.

[20] H.P. Zidek and A.R. Baghai-Wadji, "Electrostatic Field Analysis of Surface Acoustic Wave Devices, A Three-Dimensional Representation", to be published in Proc. melecon'91, Ljubljana, Yugoslavia, 1991.

[21] R.F. Harrigton, "Field Computation by Moment Method", New York, Mabmillan, 1968.

[23] R.F. Harrigton, "Matrix Methods for Field Problems", Proc. IEEE, vol. 55, pp. 136-149, Feb. 1967.

[24] M.M. Nye, "Method of Moments as Applied to Electromagnetic Problems", IEEE Trans., vol. MTT-33, pp.972-980, 1985.

SESSION 6B

OSCILLATORS I

Chairman : J-J GAGNEPAIN

Session 6B Number 61

Time and Frequency Devices in the Tomorrow Road Traffic

Sigfrido Leschiutta
Politecnico di Torino — Dipartimento di Elettronica
Torino, Italy

Summary

Electronic components in average car form today from 4 to 7% of the value, and according to some experts this incidence will reach 12–15% at end of the Century. Time and frequency devices — mostly crystal standards —, almost nil twenty years ago, presently number not less of 3–4 in a middle class car and can reach the figure of 15–20 for a luxury car equipped with CD reader, cellular, controlled drive, etc...

Around the car, new disciplines are emerging, called the Road Transport Informatics (RTI), or Advanced Road Transport Telematics (ARTT) to foster an Integrated Road Transport Environment (IRTE) with the final goal to implement a Integrated Road safety, Information and navigation System (IRIS).

All these disciplines, projects or programs, deals with the efficiency and safety of road traffic and relay heavily on telecommunication, informatics, computers, "smart" sensors, road navigation and ultimately on the use of frequency sources.

The aim of the paper is to present a panorama of the applications for oscillators in the road environment and to venture in offering an estimate of the trends and dimension of the relevant markets.

1 Introduction

The automotive share of the world market for passive components is tiny, about 4% of the total, as can be seen from Table 1, derived from [1] and other reliable sources, that presents the repartition of the market during 1990.

Also if the automotive share is small, this market is technically and industrially important, because:

- is growing,
- there is a large share of custom large scale integrated circuits, and

House hold electronics	590	M E.C.U.
Automotive	**590**	
Leisure	710	
Computers	1 780	
Industrial	1 990	
Telecommunications	2 590	
Entertainment	6 100	
Total	14 350	M E.C.U.
1 E.C.U = 1.39 US$		

Table 1: World market for passive components. 1990.

- all the components must face a very harsh environment.

With the introduction of new headlights, the bonnet will be lowered, the engine vane will become more and more cramped and it is estimated that the temperature under the bonnet, were most of the electronics must be located, will reach 125 centigrades, during this decennium.

An idea of the amount of electronic active components can be gained by inspection of Table 2, giving the combined number of integrated circuits plus large scale integrated circuits, on an existing japanese car.

Two additional figures can be quoted in order to reflect the american experience. In 1984, in the average Ford car, the share of electronic components was 585$ over the total cost and it is estimated that in 1992 this share will be around 1200$.

Anyway the electronic automotive market is growing and one of the fastest going segments is represented by the quartz oscillators.

The presence of time and frequency devices, almost nil twenty years ago, started with the introduction of a quartz clock and presently not less of 3–4 crystal are found in a middle-class average car and this figure can reach the value of 15–20 for a luxury car equipped with CD reader, cellular, etc.

Control system:	engine	20	
	transmission	9	
	anti-lock brakes	19	
	traction	24	
	air suspension	29	121
Entertainment:	crt	123	
	digital display	25	
	audio + radio	51	
	cellular	33	
	air conditioning	19	251
Others			37
	Total		**409**

Table 2: Number of a large scale integrated and integrated circuits on a top car (1990).

Moreover, around the car, new Disciplines are emerging, called the Road Transport Informatics (RTI) or Advanced Road Transport Telematics (ARTT), dealing with the efficiency and safety of road traffic. These disciplines relay heavily on telecommunication, informatics, computers, road navigation and ultimately on the use of frequency sources.

The next section covers the applications "inside" the car, while the third section deals with the infrastructure that is or will be needed "around" the car, in order to satisfy some requirements for safety, better use of natural resources and reductions of pollution.

Indeed, as it will be presented in section 4, large efforts are mounted in Europe, Japan and USA, at various levels and with different approaches, in order to reach some important goals. Some of these efforts will be referred mostly making mention to existing literature, and with reference to the largest initiatives that are under way in Europe, under support of the various Governments, program Prometheus, or the European Community Commission (ECC), project DRIVE.

In the final fifth section, an attempt will be made in forecasting the types of oscillators that will be required in the next decade for the automotive market, the relevant frequencies of operation, their requirements, such as frequency accuracy and stability, and, where possible, the number of units.

2 Oscillators Inside the Car

In this analisys only oscillators with accuracies ranging from 10^{-5} to 10^{-8} will be considered and consequently a number of simple RC oscillators present in a car will be disregarded.

The considered oscillators, all piezoelectric, can be located in all the various subsystems listed in Table 2; in particular resonators are used in the:

- clock,
- radio, fitted usually with an indirect frequency synthesizer,
- microprocessors, as a timing sources,
- bus interfaces used for the electric power distribution,
- cellular telephone, as oscillators, reference for synthesizers, filters,
- navigator, also as a sensor of rotation.

Additional quartz oscillators or dielectric oscillators will be used for the one- or two-way links under study for:

- "intelligent" cruise control, lane access, anticollision,
- automatic debiting, traffic information, etc.,

These last peculiar applications, the very first of which are now being available, will be considered in next section 3.

An estimate of the number of microprocessors and quartz crystals that could be found in a car produced in 1991, is presented in Table 3.

It must be stressed that while Table 2 represents the situation of one specific existing model, Table 3 was formed taking into account various devices fitted in a number of separate cars; in other terms, all the listed applications are now available or will be available during 1991, albeit in different models of different makers.

Some comments are of order, to explain a few entries.

As regards electrical networks, in some cars, the power distribution to the various utilities is performed via an unique heavy section feeder and the various devices are powered via individual local solid state switches that are digitally operated via a separate information serial bus. The interfaces along the information bus require a quartz crystal to insure correct decoding.

As regards front doors, these parts of the car are intensive for electrical appliances (external mirror regulation and defrosting, some courtesy and safety lights, glass panes movements, door locking, control of lock, anti-theft systems, loudspeakers) and it is becoming convenient to think to the door as an electrical subsystem , controlled and interconnected to the central computer. There is also a tendency to use the empty volume inside the front and rear doors, as a convenient seat — as regards vibration and temperature — for some of the electronics boxes. Flat and flexible electronic packages are being developed, following the "skin-electronics" approach used in aircrafts.

As regards, finally, the cathode ray tube, in some cases three microprocessors are used, one for character generation, one for the digital connections with other devices (electronic map reader and navigator,

		microproc.	quartzes
Control system:	engine	1	1
	emission control	1	2
	transmission	1	1
	anti-spin / anti-skid	1	1
	suspension	1	1
	steering wheels	1	1
	electrical network	1	11
	front doors	2	2
	air bags	1	1
Entertainment:	crt	3	3
	digital display	1	1
	audio-radio	2	4
	cellular	2	5
	air conditioning	1	1
	navigator	1	3
Totals		20	38

Table 3: Number of microprocessors and quartz crystals on a possible 1991 top car.

radio set, audio system, fault detection system, etc.), the last one for the CRT display itself, including the degaussing, luminosity control, touch-screen capability, etc.

3 Oscillators Outside the Car

In order to enhance the quality of transport, in its widest terms, a number of researches and experiments are performed in Japan, Canada, U.S.A. and Europe.

The objectives are to increase the safety, to reduce the pollution and to improve the general efficiency of the traffic.

Possible applications are countless, from automatic debiting to "intelligent" driving (safe lane changing, overtaking ...), to anticollision, to intersection collision warning, to dynamic traffic regulation, to parking management, etc..

An analysis of these applications is out the scopes of this papers and relevant information will be found in the books and papers quoted in the next section 4.

At any rate a couple of examples will be given in order to substantiate the thesis that will be presented in what follows, i.e. that near the totality of the considered applications, is based on telecommunications and consequently on oscillators.

Moreover all these oscillators will require quartz crystals, some ovenized — or temperature compensated —, or, finally, SAW or dielectric resonators.

Different examples, some existing, some planned, of the various possible applications are given in Table 4.

Some of these applications are well established, such as radiobroadcasting and the conviviality link, that uses the CB band, other as cellular telephone or automatic debiting are developing, other, such the

Road to vehicle links — one way
– Radiobroadcasting (also in tunnels)
– Mobile telephone and fax (also in tunnels)
– traffic information
– traffic guidance
– anticollision at intersections
– traffic light state repetition on board
– parking management
Road to vehicles links — two ways
– public transport operations and priority schemes
– automatic debiting
– fleet management
– traffic data collection using car transmissions
– aids to distressed vehicles
Vehicle to vehicle links
– emergency warning systems
– assistance to column automatic driving
– overtaking and changing of lane aids
– emergency vehicles signaling
– conviviality link

Table 4: Road-to-vehicle and vehicle-to-vehicle links.

Freight management:
1. On line cargo tracking and tracing
2. Automatic notification on arrival
3. Computer communication between shipper and carrier/vehicle
Fleet management:
1. Automatic fleet monitoring
2. On line route planning and scheduling updates
3. Route planning and scheduling based on actual road network data
Vehicle driver management:
1. Automatic vehicle positioning
2. Automatic vehicle navigation and route guidance
3. On board telefax for document exchange

Table 5: Future trends and attitudes towards road transport informatics for freight.

automatic driving will be a matter of the next century.

As as example, the two-way link applications characteristic of the freight industry are listed in Table 5.

Freight transport is considered only, because it is estimated that the first (in time) massive applications of transport Informatics will be in this field.

By inspection of Table 5, is evident that all the nine examples are based on a form of telecommunication.

The considered links range from leaky feeders to satellites, from 50 kHz to 80 GHz, but in any cases the number of oscillators involved will be very large and all with accuracies better than 10^{-4}, to be maintained in the field, under strong temperature and vibrations stresses.

4 Research and Development Programs

Leaders in Route Traffic Informatics studies, experiments and applications are Europe and Japan.

Two major initiatives are in progress in Europe. PROMETHEUS, promoted by the european Governments and Industries, and DRIVE organized directly by the Directorate Telecommunications, Information Industries and Innovation of the Commission of the European Communities.

PROMETHEUS stands for PROgraMme for European Traffic with Highest Efficiency and Unprecedented Safety, and is a nine year research program, now at its fourth year, in which near all the european car makers, and many Universities and Research Laboratories are involved.

DRIVE stands for Dedicated Road Infrastructure for Vehicle safety in Europe; the first group of activities begun in 1989 and now is at its end, with an ECC contribution of about 60 MECU (millions of European Currency Unit), the second group will be launched in 1992, with an ECC contribution of about 120 MECU.

Prometheus and Drive are very large research programs , in which in total about 200 topics are considered [2, 3, 4].

In Japan similar efforts, but with a different orientation are pursued by the Japanese Car Communication Research Committee, formed by the Ministry of Constructions, public users, Research Institutes and private sectors, represented by car makers and electronic components and systems factories.

In the U.S.A., in preparation of the Surface Transportation Renewal Act, to be debated in 1991, the RTI topics are discussed in an informal coalition of government, academic, and industry representatives under a program called "Mobility 2000", in preparation of a continent-wide Intelligent Vehicle-Highway System - IVHS.

5 Trends and Forecasting

5.1 Oscillators Inside the Car

The automotive market, being so relevant for the economy of a nation, is well known and studied; consequently also the forecast of the crystal oscillators that will be needed in future is somehow eased by the wealth of information available.

Table 6 gives the 1989 and 1990 car production for World, ECC, European Free Trade Association Nations and Europe.

Table VII reports the distributions in various classes of the cars sold in Europe during 1990, japanese cars included.

The figures given in Table VII are provisional, but it is estimated that the total error is bounded to 1%.

It would be really cumbersome to obtain an educated estimate of the number of quartzes mounted. A guess can be anyhow attempted: taking into account that in the class A very seldom the clock is mounted, while the clock starts to be a regular feature from the middle of class B onward. It can be therefore assumed that during 1990 at least 11 millions of electronic clocks, each with its crystal, were included in the cars sold in our Continent.

In class C, very few cars are encountered with a microprocessor, usually used for the injection control, and consequently it is safe to assume that also in class C the clock only can be found.

Classes A, B, and C accounts for about 60% of the sales, excluding classes H and I, it remains a population of about 38%, i.e. a number of about 5,800,000 cars, in which at least one-two microprocessors are included and consequently two-three crystals are used.

An accepted estimate of the total number of crystal units needed in the car industry in Europe for 1991 (frequency accuracy $\pm 10^{-5}$, temperature range

YEAR	WORLD	EUROPE		
		ECC	EFTA	TOTAL
1989	35.95	13.34	0.82	14.16
1990	36.10	13.40	0.82	14.22

Table 6: Car production (million of units).

Class	Percentage	Remarks	Examples
A	3.4	utility - small	126, Panda, R4
B	28.6	utility - large	Uno
C	27.0	average	Tipo, Ritmo, Alfa 33
D	24.5	medium	Audi 80, Prisma
E	9.8	medium-top	Alfa 164, Mercedes 190
F	3.2	top class	Bmw, Volvo, Citroen CX
G	0.7	luxury	
H	1.1	spider	
I	1.7	off-road	

Table 7: Distribution in classes of european 1990 sales (japanese cars included).

Uses	year 1991	year 1995	
clocks	11.0	12	
"smart radios"	3.5	5	
microprocessor	1.6	4	1 μP
		8	2 μP
		20	2-10 μP
		6	20 μP
totals	15.1	55	

Table 8: Quartz crystals for car industry in europe (million of units).

$-25°C$ to $+75°C$) is given in Table 8, along with a forecast for 1995, (same frequency accuracy, but extended temperature range ($-30°C$ to $+100°C$); the figures are in million of units for 1990.

Following some other experts, the need for 1991, will reach possibly 18 millions and for 1995 will be limited to 45-50 million; anyway we are confronted with very large number of crystal units.

As regards the ageing of frequency, the following figures are now considered adequate, both from the car-makers and the quartz-makers, $1 \cdot 10^{-6}$ /one year and $7\text{-}8 \cdot 10^{-6}$ / 10 years.

5.2 Oscillators Outside the Car

No forecast is possible as regards the number of "precise" oscillators that will be used in the applications "outside" the car; indeed the real introductions of the applications outlined in point 4, will depend on the World economy and, as regards, Europe, on the political agreements reached at level of the Community and on their subsequent implementations by the different Countries.

At any rate some forecast are possible as regards the types of "precise" oscillators (frequency, stability, aging, etc.) in order to orient the research and the technical developments.

As regards the frequency, in January 1991 the CEPT (Conference Europeenne des Postes et Telecommunications) has adopted tentatively an european plan of special frequency allocations for the automotive applications, given in Table 9.

Table 9 commands some remarks:

- for short distances (1-30 m) microwaves at 5.8 GHz (or infrared) are to be adopted. No technical problems are foreseen, the technology is available because a frequency stability of a few kHz/°C is offered by existing devices and consequently no frequency multiplications chain will be needed for most of the applications.
- for medium distances (30-500 m), millimetric waves will be used and DROs (Dielectric Resonator Oscillators) will be needed for frequency stability. This is indeed an open field for R&D, since at the moment rugged and inexpensive millimetric oscillators, suitable for mass production are not yet available.

For the microwave and the millimetric oscillators a frequency stability of 10^{-7}/°C can be tolerated, also if for some applications (Radar Doppler) a stability of $1 \cdot 10^{-8}$/°C would be improve the measurements.

It can be concluded that the use of precise frequency and time devices in the automotive Industry is now a reality.

Vehicle-infrastructure links, both one- and two-way		
5810-5815 MHz	power 3 dBm	
5795-5800 MHz	in future	
5800-5805 MHz	in future	
5805-5810 MHz	in future	
Vehicle-vehicle and vehicle-infrastructure links		
63-64 GHz	power 3-16 dBm	
	antenna gain 10-30 dBi	
Anticollision radars		
76-77 GHz	FM Radars	100 MHz bandwidth
		Pulsed Radars 500 MHz bandwidth
		power 16-20 dBm
		antenna gain 30-35 dBi

Table 9: Frequency allocations in Europe. 1991 CEPT recommendations for automotive applications.

The importance of these devices will increase with the time but some difficult problems are to be solved, such as the frequency stability versus temperature for the millimetric oscillators, that must be cheap and reliable.

Another problem requiring some efforts is the choice of more suitable mechanical suspension systems for the quartz oscillators to be used in board of a vehicle, for the very peculiar spectrum of mechanical vibrations encountered, with peaks at frequencies under 20 hertz.

References

[1] Siemens Components, various issues, and in particular Vol. XV, no. 6, p. 198, 1990.

[2] "R+D in advanced road transport telematics in Europe: Operation 1992," CEC, XIII / 494/89-EN, Brussels, 1989.

[3] "R+D in Advanced Road Transport Telematics in Europe, Workplan 91", CEC, XIII, Brussels,1990.

[4] "Advanced Telematics in Road Transport," 2 Vols, Elsevier, Amsterdam, 1991.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 6B Number 62

A NOVEL, LOW-POWER, FAST-WARMUP, MINIATURE OSCILLATOR

Wolfgang Pflaum, Paul Naepflin, John Fellner,
Wendy Chen , Jack Kusters , Pierre Stoermer
Ball Corporation, Efratom Division
Irvine, California U.S.A. 92718-1605

ABSTRACT

The goal of this operation was to design and produce an oscillator with the low-power features of the TMXO and frequency performance of the HP10811 oscillator, and do this in a package that is comparable in size to the TMXO. The design and manufacturing approach used was to develop a miniature ovenized oscillator using a 10 MHz, 3rd overtone SC-cut resonator and novel microelectronic module. The small thermal mass of the ovenized electronics warm-up to better than 1 x 10^{-8} of final operating frequency is achieved within 3 minutes at -55°C with an input power of less than 7 watts. Through the use of low thermal conductivity insulation operating at a partial pressure steady state power dissipation of the oscillator at 55°C is less than 900 milli-watts.

Other performance parameters:

Size, Weight	-	1.3" cube, 40 g.
Allan Variance, 0.1 to 10 sec.	-	< 5 x 10^{-12}
Phase noise floor	-	< -155 dBc
Aging	-	< ±5 x 10^{-10}/day
Operating range	-	-55° to +95° C.
Temperature stability	-	< ±5 x 10^{-9}
RF output	-	> 7dBm sinewave

Details are presented of the oscillator design, performance, production processes and the facility used to manufacture the device in volume.

INTRODUCTION

Current and future communication, navigation, and instrumentation programs require small, high-precision oscillators for timing and frequency reference. System specifications often require the oscillator to maintain high stability in various environments. Fast warm-up at any ambient temperature, stability over the operating temperature range, low phase noise under vibration, high reliability, and low cost are some of the typical requirements. In addition, as electronic equipment enclosures shrink, more demands are made on the oscillator in terms of its size and power consumption.

We describe the design and performance of a miniature, low power, high stability crystal oscillator intended for use in these types of systems. The oscillator utilizes a small precision SC-cut resonator and a unique microelectronic module housed within an insulated, hermetically sealed enclosure. Various tradeoffs were made during the design and development of the oscillator to achieve good performance in a small package without the associated high cost.

ELECTRICAL CONFIGURATION

The electrical configuration of the device consists of four basic blocks: regulator, oscillator, output buffer, and thermal control (*Figure 1*). A modified Pierce configuration oscillator with appropriate A and B-mode traps, automatic gain control (AGC) and electronic frequency control (EFC) drives a precision 10MHz, 3rd mode, SC-cut resonator. The drive current to the resonator is controlled to approximately 1.0 milliampere, dissipating about 50 microwatts of power. With a resonator motional capacitance of 0.3 femtofarads, the oscillator can be pulled in frequency over 1 ppm with a 5-volt DC control. Resonators are tuned to within ±0.3 ppm at their operating temperature thereby eliminating the need for mechanical frequency adjust and/or selected load capacitors.

The output of the Pierce oscillator drives a low-noise isolation amplifier which, in turn, drives the output buffer. A two-stage amplifier which provides current gain and impedance matching to 50 Ω is used for the output buffer. The sinewave output of the oscillator is at least 7dBm.

Temperature control of the resonator and oscillator components is provided by a thermistor bridge, error amplifier which proportionally controls drive current to a power transistor as the ambient temperature changes. A warm-up control circuit maintains nearly constant warm-up power over the oven input voltage range of 10 to 28 Vdc. Oven power requirements during the warm-up period is typically 7 Watts; although, special versions have been developed up to 10 Watts.

THERMAL DESIGN

As the ambient temperature increases, the drive to the power transistor proportionally decreases so that the oven operating temperature remains constant. Once the upper operating ambient temperature has been reached, the heater turns off and the electronics continue to dissipate idle power. Any further increase in ambient temperature causes loss of thermal control of the oven and subsequent large frequency shifts. The difference in upper operating ambient temperature and the oven set temperature is controlled by the idle power of the electronics and the thermal resistance between the oven and ambient. In order to maintain high reliability (MTBF) of the oscillator, the oven set temperature should be set as low as possible consistent with thermally controlled operation at the desired upper ambient operating temperature.

The following condition must be met to assure thermal control of the oven at the upper operating ambient temperature:

$$R_{th} = T_{oven} - T_{ambient} / P_{idle}$$

Where: R_{th} = Thermal resistance between oven and ambient
T_{oven} = Oven set temperature
$T_{ambient}$ = High ambient temperature
P_{idle} = Idle power of electronics

As R_{th} gets larger, when low power at -55°C is required, the need for minimizing the electronics idle power in order to keep the oven temperature relatively low becomes very important. Idle power of the oscillator and thermal control circuits has been kept to less than 30 milliwatts. If 95°C ambient operation is desired and the oven set temperature is held at 110°C, a thermal resistance of no greater than 500 °C/watt can be tolerated.

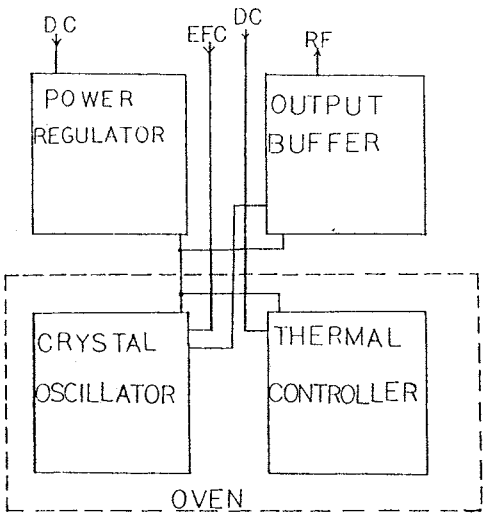


Figure 1. BLOCK DIAGRAM OF ELECTRONIC CIRCUITS

CRYSTAL RESONATOR DESIGN

The quartz crystal is manufactured in our facility as a 10.0 MHz, 3rd mode, SC-cut resonator. Previous papers on SC-cuts have mentioned the need for tight control on the orientation angle of each wafer in order to achieve the true SC-cut and to place the turnover point where it is desired. This is achieved here through individual wafer X-ray measurements and angle correction, prior to the more conventional wafer processing. *Figure 2* displays the mounting fixture for holding the wafers in the X-ray machine. This fixture is also the transfer device for holding the wafer during the lapping operation for correcting the angle. The three points at the edges of the fixture have diamond tips which define the reference plane, both on the X-ray machine and the lapping plate. *Figure 3* shows the X-ray machine which is used. The goniometer shown in the left of the photograph is a two-axis system which is used to measure individual X-ray reflections but can be adjusted to measure any set of crystal planes.

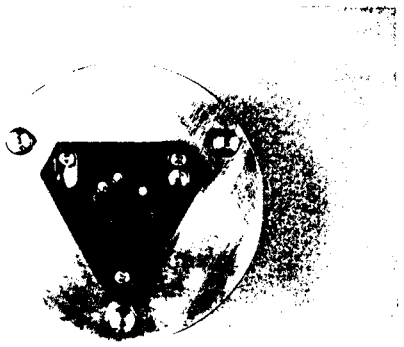


Figure 2. CRYSTAL WAFER FIXTURE



Figure 3. X-RAY MACHINES

It is usually used to measure three independent reflections. The other goniometer in the right background is a fixed system to measure all three X-ray reflections simultaneously. In this way, the true 3-D orientation can be determined precisely.

The crystal is processed in a clean room environment with cryo-pumped vacuum systems for the baseplate and final tune operations. *Figure 4* shows these two pieces of equipment with the baseplater on the left and the final tuner on the right. This system was described earlier[1]. All the crystals are mounted on a four-point configuration on a coldweld header as indicated in *Figure 5*. The side view in *Figure 6* indicates the very low profile that can be achieved with this type of mounting structure. All of the final crystal resonator processing, including the final tuning of frequency to a tolerance of 1 or 2 Hz, high temperature bake-out and coldweld sealing at the operating temperature of the oscillator, is done in the final tuner.

MECHANICAL CONFIGURATION

The thermal mass of the ovenized electronics, including the resonator, has been made as small as practical in order to achieve very fast warm-up with minimal energy input. The construction of the electronics module, shown in *Figure 7*, features an etched, copper-clad, polyimide, rigid/flex substrate which is used to provide electrical connection to the components. The module is separated into two parts which are connected by a multi-wired flex circuit. The ovenized oscillator with thermal control circuits and heat sink is

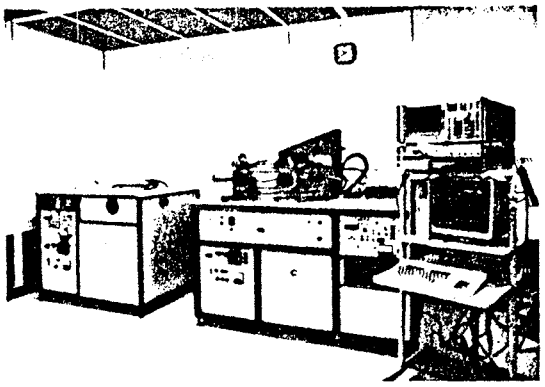


Figure 4. CLEAN ROOM WITH BASEPLATER AND FINAL TUNER

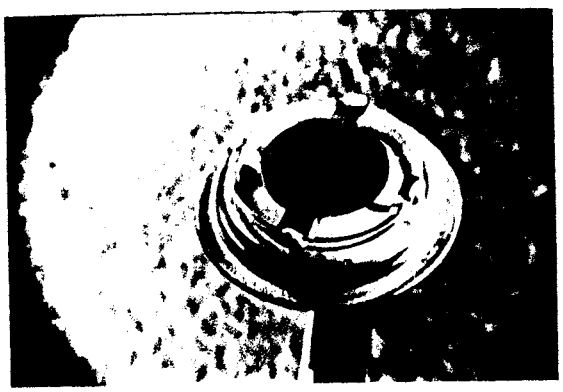


Figure 5. MOUNTED CRYSTAL RESONATOR



Figure 6. SIDE VIEW OF RESONATOR

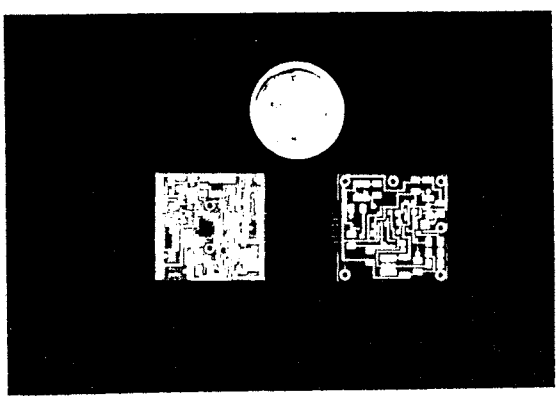


Figure 7. ELECTRONICS MODULE

located on one sub-module, while the voltage regulator and buffer electronics are assembled to the other. Semiconductors are epoxy die mounted to the substrate then wirebonded into the circuit. Passive devices are reflow soldered. An epoxy conformal coating is applied to all wirebonded devices for mechanical protection. Thermal connection between the crystal resonator and oscillator sub-module is made by contacting the resonator package to the heat sink. In this configuration, the oscillator/thermal control sub-module functions as a miniature hot plate for control of the resonator's operating temperature.

The buffer/regulator sub-module is soldered to hermetic feedthroughs in a copper header. Thermal insulation and mechanical support of the oven within the enclosure is provided by two molded pieces of a low thermal conductivity silica-based material. The oscillator cover is cold welded to the header under a partial vacuum to achieve hermeticity. Use of a partial vacuum decreases the thermal conductivity of the insulation so that the thermal resistance of the system may be increased significantly. The thermal resistance of the system is determined by two parallel thermal conduction paths, the insulation and the flexible jumper. Thermal losses down the flex jumper can be minimized by proper sizing of the etched copper traces. Losses down the flexible polyimide insulator are negligible by virtue of its low thermal conductivity and small cross sectional area. For reliable, low-power operation over the life of the oscillator, the partial pressure must be maintained. Factors contributing to the degradation of vacuum with time are loss of hermeticity of the outer enclosure and internal component outgassing. Cold welding of similar materials for achieving a hermetic seal has been used extensively by many resonator manufacturers with high yield. It is a relatively simple, inexpensive and reliable method for sealing packages of various form factors. Outgassing of internal components has been minimized by the use of vacuum compatible materials and thorough cleaning prior to assembly. *Figure 8* shows the glovebox and vacuum bake-out chamber that is used to outgas the insulation material. Dry nitrogen atmosphere in glovebox keeps the moisture level a low ppm levels. The coldweld sealer, shown in *Figure 9*, with a closeup in *Figure 10*, has a surrounding vacuum chamber which is used to evacuate the enclosure prior to sealing. The internal operating pressure of the oscillator enclosure is maintained relatively high so that residual outgassing does not significantly affect long-term operation. An additional benefit is that enclosure components can be leak tested in production to assure a level of hermeticity required for maintaining the operating pressure.

When the unit is sealed, a compressive force is applied to the oscillator/resonator assembly by the insulation thereby rigidizing the structure so that no internal mechanical resonances occur under vibration to 2000 Hz.

MANUFACTURABILITY AND TESTING

Once the resonator has been electrically and thermally characterized, it is attached to the heat sink on the oscillator sub-module and connected into the circuit. The buffer sub-module is soldered to the pin feed-throughs then cleaned thoroughly for removal of flux and any other possible contamination. After

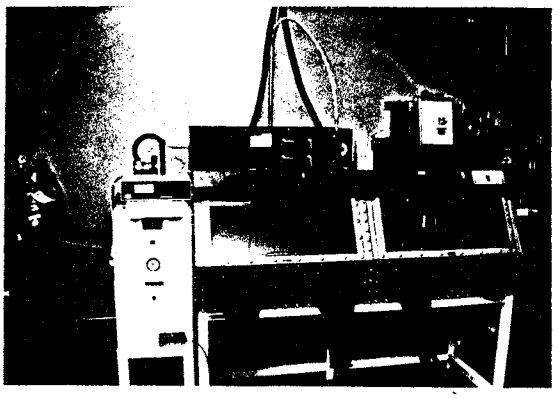


Figure 8. SEALING GLOVEBOX



Figure 9. COLDWELD SEALING OPERATION

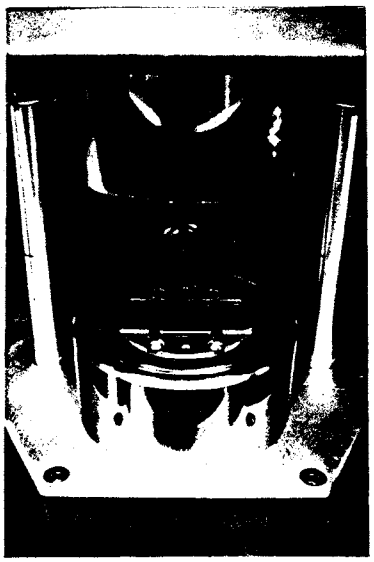


Figure 10. COLDWELD SEALER

chemical cleaning the electronics assembly, insulation and enclosure are inserted into a vacuum oven for extended bakeout at 150°C. The inside of the final enclosure, packed with insulation, before the electronics assembly is inserted, is shown in *Figure 11*. A computer controlled data acquisition system is used for data collection on sealed oscillators. Long-term aging and power aging data is collected, analyzed and used for monitoring the health of the entire process. The final product is shown in *Figure 12*.

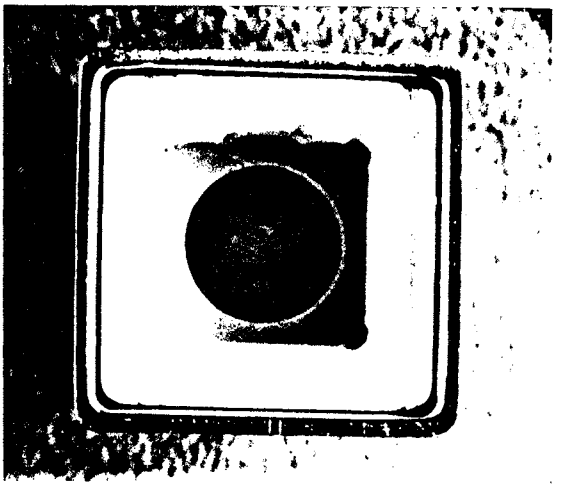


Figure 11. ENCLOSURE PACKED WITH INSULATION

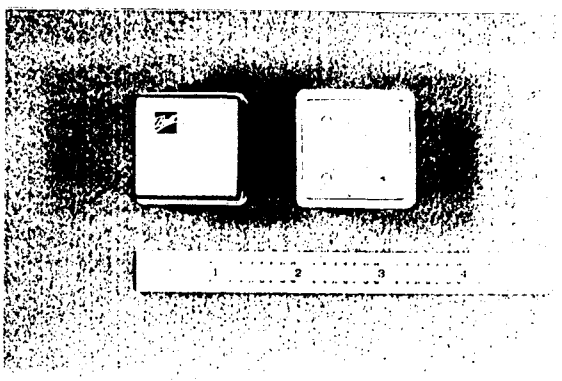


Figure 12. FINAL OSCILLATOR PRODUCT

OSCILLATOR PERFORMANCE

Typical frequency-temperature characteristic curves for SC-cut crystals are shown in *Figure 13*. The inflection temperature is near 92°C, and the upper temperature turnovers are well above 100°C. In our case, the oven temperature was set to 110°C so that the oven would still be stable at an ambient of 95°C. A detailed frequency-temperature curve for a crystal whose turnover point is at 110°C is indicated in *Figure 14*. The dashed line near the bottom of the figure represents the stability goal of the design at 5×10^{-9}. Actual performance results are indicated in *Figure 15*, which clearly met the design goal.

Naturally, the power dissipated by such a unit can be expected to vary somewhat over temperature. *Figure 16* indicates that even at -55°C the oscillator consumes less than 1 Watt for the full RF sine output variety. A variation which delivers a TTL compatible signal uses only 750 mWatts typically.

Another very useful feature of such a small oscillator is the ability to stabilize very rapidly even with the low power consumption. *Figure 17* indicates the typical warm-up performance of this oscillator. Most units that have been built have warm-up stabilization times to 1×10^{-8} of less than 4 minutes. Quite often, times of less than 2 minutes are observed.

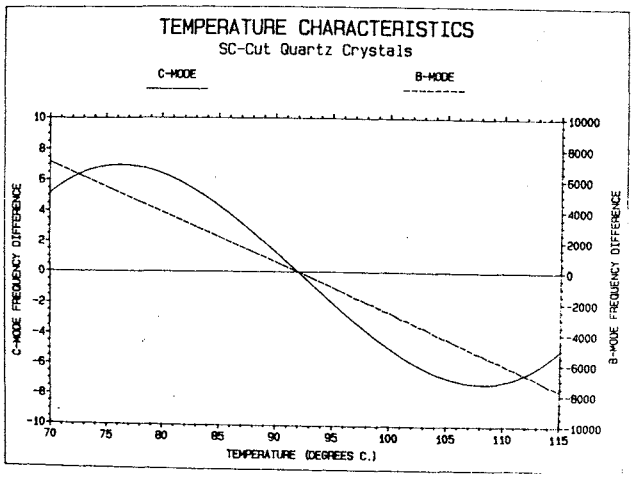


Figure 13. *TYPICAL SC-CUT CRYSTAL FREQUENCY-TEMPERATURE CURVES*

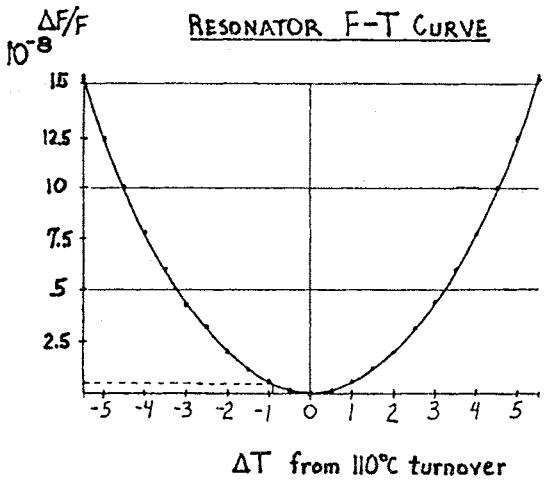


Figure 14. *DETAIL OF TURNOVERPOINT FREQUENCY-TEMPERATURE CURVE*

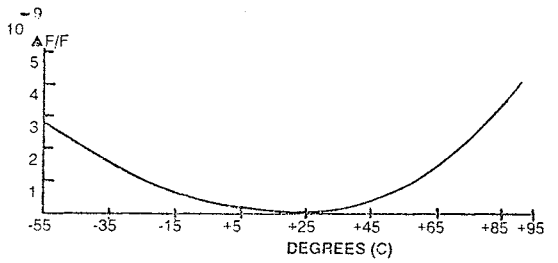


Figure 15. *RESULTANT OSCILLATOR TEMPERATURE STABILITY*

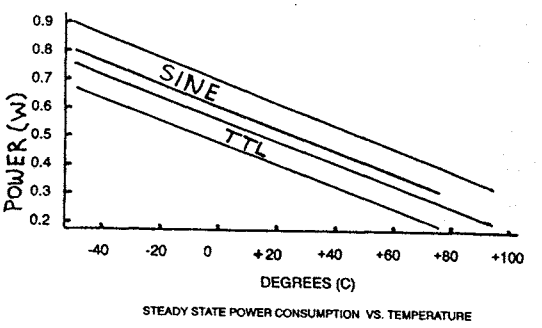


Figure 16. *POWER DISSIPATION*

The oscillators are stable within 15 minutes such that they can then be used as frequency references for very precise applications.

Frequency adjustment of oscillators is typically done with screw trimmers on varactors or capacitors. With this oscillator, since it is enclosed in a hermetic package, an electronic frequency control (EFC) function is available at one pin. *Figure 18* indicates the type of trim sensitivity and range that is available with the EFC function.

Figure 19 illustrates the phase noise performance of the oscillator, and typical short term stability calculated using Allan variance techniques results are $5x10^{-12}$ or better for 1 sec. time intervals.

A reliability analysis calculation was performed for this oscillator, per MIL-STD-217E, and the results for airborne environments are tabulated in *Figure 20*. Results are also shown for a variety of the oscillator with the oven set at 90°C. In both cases, an average ambient of 50°C was assumed.

The worst case operating scenario for the oscillator is repeated on-off cycling at -55°C. When the oscillator is turned off and allowed to stabilize at low temperatures, application of power causes rapid warm-up to the operating temperature of the oven. Approximate temperature slew rate for the oscillator in this case would be 2°C/second. In order to verify the reliability of the electronics assembly, units were on-off cycled at -55°C, 500 or more times with no failures. Retrace measurements have been made periodically while the power cycling was done, and the results indicate that oscillators retrace to several parts in 10^{-9} during the first several times and eventually retrace to parts in 10^{-10}.

Aging measurements are done in a fully-automated testing facility, and the results indicate that the crystal controlled oscillators typically start aging at the specified limit of $5x10^{-10}$ per day or better. However, a significant portion of the oscillators start aging in the low 10^{-11} per day range.

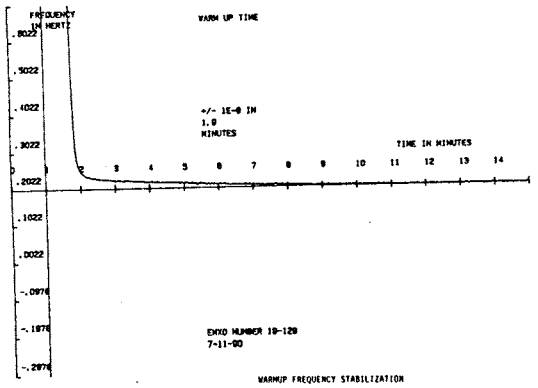


Figure 17a. FAST WARMUP TIME

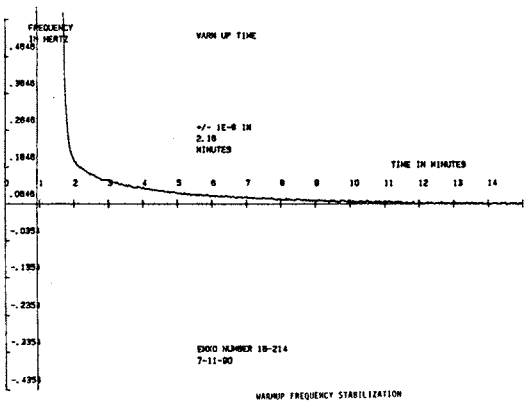


Figure 17b. TYPICAL WARMUP TIME

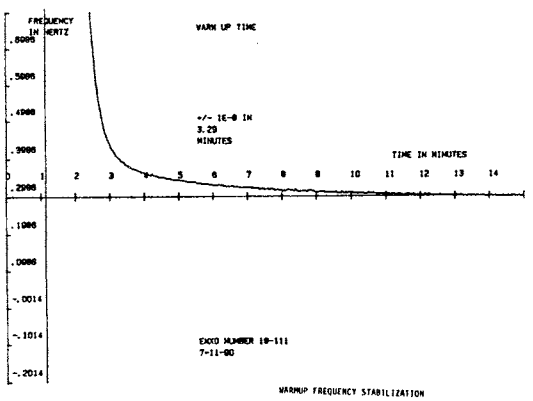


Figure 17c. SLOW WARMUP TIME

CONCLUSION

A new oscillator has been described which combines the very useful features of being small, light weight, consumes low power, is very stable over temperature, and ages at a low rate. It is our expectation that such a product can be practical in many applications and fulfill a void in the oscillator stability between the ultra-stable precision ovenized oscillators and the temperature compensated oscillators.

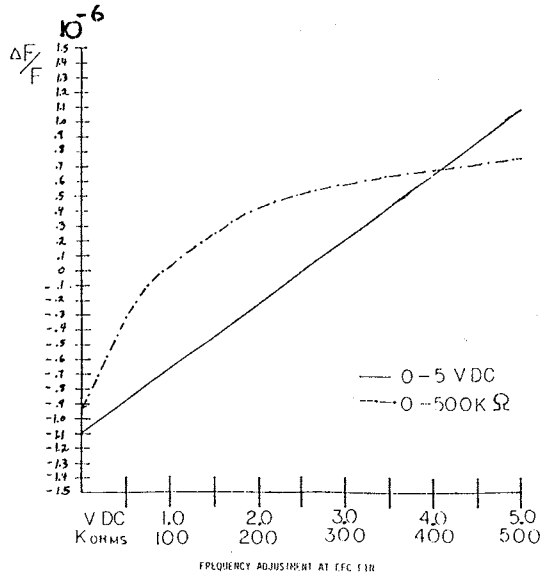


Figure 18. SENSITIVITY OF EFC

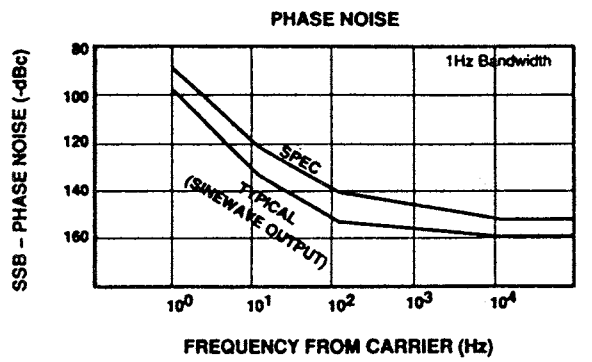


Figure 19. OSCILLATOR PHASE NOISE

RELIABILITY ANALYSIS

ENVIRONMENT @ 50° C	MTBF (HOURS) OVEN @ 90° C	MTBF (HOURS) OVEN @ 110° C
AIC	334,768	220,575
AUF	84,252	58,378
ARW	104,066	70,465

Figure 20. RELIABILITY ANALYSIS

REFERENCES

1. Pierre Stoermer, "Automated Measurement System for Final Frequency Plating and Characterization of High Precision SC-Cut Crystals," Proceedings of the 3rd European Time and Frequency Forum, Besancon, 1989, pp 143-146.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 6B Number 63

COMPACT MICROPROCESSOR CONTROLLED DTCXO

RALF OLBERT

TELE QUARZ GROUP GERMANY

SUMMARY

The paper describes a microprocessor controlled DTCXO as a miniature type.
Always new developments and techniques of electrical devices and systems make the electrical specifications harder and complicater. So the crystal oscillators have also been required to have a graeter frequency stability, be smaller in size and to have a lower power consumption.
In many systems have usually been applied Temperature Compensated Crystal Oscillators (TCXOs).
Now the TELE QUARZ GROUP has developped a Digitally Temperature Compensated Crystal Oscillator, called DCO-100, which is controlled by a micro-processor. The DTCXO consists of an analog and a digital part. The analog section contains the crystal oscillator, a reference voltage supply curcuit, buffer curcuits, a divider, a varicap and a curcuit, that produces a temperature dependent DC voltage. The digital section including the com-pensation, consists of an 8 bit A/D-converter, a microprocessor with a serial interface, a 256 word X 8 bit EEPROM, and an 8 bit D/A-converter. The digital and analog parts are placed separately on two pc-boards, which are lying one upon the other. The dimensions of the housing are 36.1mm x 27.2mm x 19.4mm.
Important features for the DTCXO are small frequency variations over a wide temperature range (±0.5 ppm from -40°C to +90°C) and low current consumption of 25 mA at 5 V supply voltage.

INTRODUCTION

A crystal oscillator is in general known as an oscillator with a high Q-factor and a high frequency accuracy verse time and temperature. For many applications is the accuracy of the crystal frequency over a temperature range not high enough and so it's necessary to reduce the temperature effects. There are two possibilities to compensate the frequency variation over temperature. The analog and the digital method. With analog compensation networks of a TCXO you can reach frequency stabilities from ±1 ppm over a temperature range from -40°C to +90°C. With other analog techniques is it only on the base of an Oven-Controlled Crystal Oscillator possible to get much better results (for example: ±0.05 ppm over -20°C to +70°C). But the disadvantages of such an oscillator are the long warm-up time, the higher power consumption (200mA, 12 V) and the higher production costs. The digital compensation is a considerable new method for compen-sating the frequency of a crystal oscillator. Because of its easy production engineering and its good frequency stability over a wide temperature range the DTCXO, called DCO-100 (s. Fig. 1), which is devel-opped by TELE QUARZ GROUP, is more than a alternative to other stable crystal oscillators. With a current consumption of 25 mA at 5 V supply voltage you can get a frequency stability of ±0.5 ppm over a wide temperature range from -40°C to 90°C.

PRINCIPLE

The DTCXO consists of three basic sections, as shown in Fig.2, a curcuit for temperature acquisition, a digital compensation curcuit and a crystal oscillator section. The digital compen-sation part includes an 8 bit A/D-converter, a microprocessor with a serial interface, a 256 word x 8 bit EEPROM and an 8 bit D/A-converter.
Guided by the three different operating modes of the DCO-100 the principle of digital compensation will be described (s. Fig. 3).
a) Pre-measurement-mode
The complet closed oscillator will be measured in a temperature test chamber by a number of temperatures as follows:
The digital temperature word, that's measured by the temperature sensor and converted by the A/D-converter will be sent via the serial interface of the microprocessor to an external computer. At the serial input pin of the microprocessor will be put on a digital pulling-voltage word. This digital voltage supplies the oscillator through a varicap over the D/A-converter and the lowpass filter. It will be changed so long with an iteration process till the load capacity of the crystal has reached the right load capacity term, that the oscillator is pulled on his nominal frequency. Then the temperature word and the corresponding pulling word will be stored on an external computer. These values have to be find out individually for each oscillator, because each crystal has got another temperature curve.

b) Compensation-mode
On an external computer the curvature of the digital compensation voltage will be calculated with an inter-polation procedure over the whole temperature range for all the inter-mediate temperature values that are not measured. The temperature data is digi- tized and characterized into 28 break points, where 8th is the number of bits used for the A/D-converter. The EEPROM is programmed over the serial interface of the microprocessor with a customized look-up table, which provides cor-rections directly to each digitized temperature word. The temperature words, from 0 to 256, is used to as the address for the 256word x 8bit EEPROM.

In the pre-measurement- and the compensation-mode we need for the data transmission beetween the external computer and the microprocessor a serial/parallel interface, because the microprocessor uses the oscillator

frequency as its on clock-frequency and so the microprocessor can't reach the defined baudrates of the serial interface of the external computer like 4800, 9600, etc.. But the advantage is, that we needn't a second crystal for the microprocessor in the oscillator. So we can save place on the pc-board and we have only on frequency in the oscillator.

c) Operating-mode
With a repetition rate, which is deri-vated from the oscillator frequency, the microprocessor continually reads in the current temperature voltage through the A/D-converter. The digital tempera-ture word points at its corresponding compensation voltage in the EEPROM-table and puts out the digital correction voltage to the 8 bit D/A-converter. The D/A-converter converts this value to an analog pulling voltage, which supplies over a lowpass-filter the varactor of the crystal oscillator. So the load capacity of the crystal will be always changed, so that the oscillator swings on his nominal frequency.
With an external potentiometer you have got the possibility to modulate the frequency. The external modulation voltage, which is produced by the potentiometer, is put on a second A/D-converter. The microprocessor works up the digital value and changes the digital pulling word and consequently the oscillator frequency.

Experimental Results

Compensation experiments, that are presented, were made by using an AT cut 10.0 MHz crystal unit. On figure 4 are first shown the temperature characte-ristics of the uncompensated crystal. The figure 5 indicates the compensated result. The less than ±0.5 ppm frequency variation is achieved over the temperature range from -40°C to 90°C. Over the temperature range from -55°C to +95°C you've got a frequency variation of ±0.6 ppm.
On figure 6 is only shown a part of the compensated frequency diagramm. Here is it better possible to show the 'frequency-step' per bit and the 'temperature-step' per bit.

'temperature-step'/bit = T / bt

'frequency-step'/bit = F / bv

Where:
T = total temperature range
bt = number of bits in the temperature word (ADC)
F = maximum peak to peak frequency excursion of the crystal in ppm
bv = number of bits in the correction voltage word (DAC)

The figure 7a shows the digital pulling voltage and 7b shows the digital temperature voltage, both over temperature. On figure 7b you can see the monotonic curvature of the digital temperature voltage, that is required to get different digital temperature words at each temperature value.
On figure 8 is shown the size and the pinning of the DCO-100.

Characteristics of the DTCXO

The DTCXO, called DCO-100 has the following characteristics:

- Only one thermistor is needed
- Wide temperature range compensation is possible
- Easy pre-measurement with the serial interface

Directions of future work

In these days the TELE QUARZ GROUP is developping a new DTCXO on the base of an ASIC. In the ASIC will be integrated the digital and analog sections and a EEPROM array; only two block capacities and the crystal will be necessary as external devices. The different compensation method of the 'ASIC-DTCXO' is this that only the coeefficients of the digital compensation voltage will be stored in the EEPROM array and not the whole look-up table as in the DCO-100. The ASIC will be designed in CMOS 1.5 um technic and is stored into a SO-20 housing. It is also possible to get the ASIC chip as a single die with no housing to make the whole oscillator still smaller. Table 1 shows some characteristics of the 'ASIC-DTCXO'.

References

1) Proceedings of the 44th annual Symposium on Frequency Control 1990
2) Proceedings of the 41st annual Frequency Symposium 1987
3) Symposium Schwingquarze '89 ZVEI Fachverband Bauelemnte der Elektronik

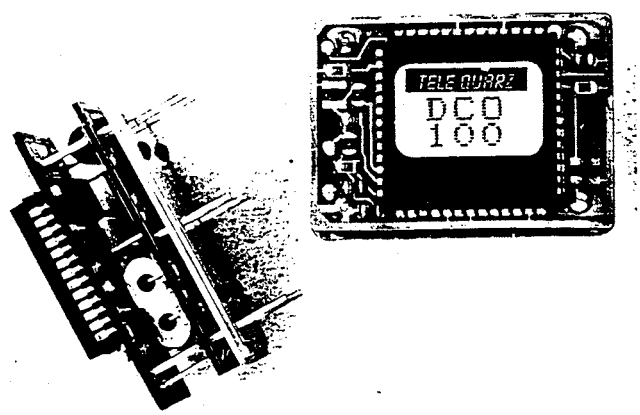


Fig.1 The DTCXO,called DCO 100

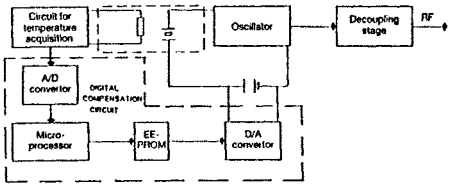


Fig.2 Block diagramm of the DTCXO

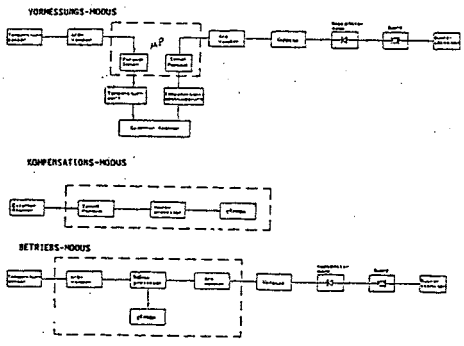


Fig.3 Block diagrams of three different operating modes

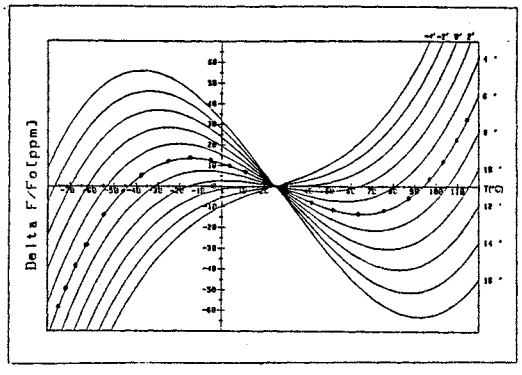


Fig.4 Temperature Characteristics of the uncompensated crystal

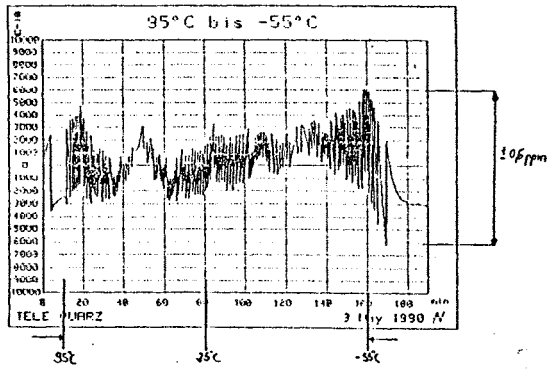


Fig.5 Compensated frequency vs. temperature characteristics

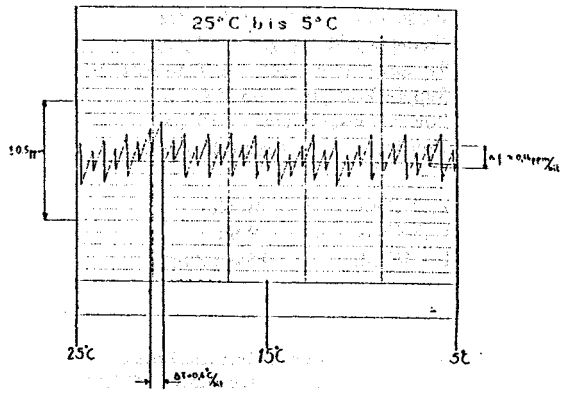


Fig.6 Part of the Compensated frequency

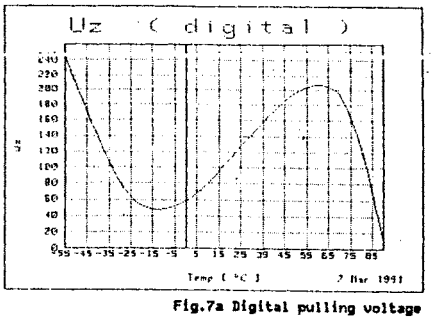


Fig.7a Digital pulling voltage over temperature

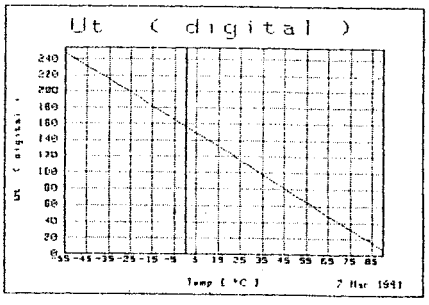


Fig.7b Digital temperature voltage over temperature

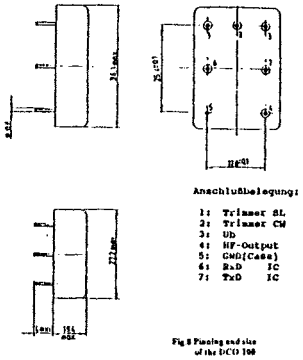


Fig.8 Pinning and size of the DCO 100

Frequency Range: 3-30 MHz
Frequency Stability: < ±0.3 ppm
Temperature Range: -40°C to 90°C
Supply Voltage: +5 V ±5 %
Power Consumption: < 15 mW
Housing: 20x20x10 mm3

A/D-Converter: 10 bit
D/A-Converter: 9 bit
EEPROM: 16w x 8 bit

Table 1 Characteristics of the "ASIC-DTCXO"

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 6B Number 64

DESIGN AND QUALIFICATION OF AN ULTRA STABLE OSCILLATOR FOR SPACE APPLICATIONS

by Evelyne GERARD and Roger MOLLE
C.E.P.E. 44 avenue de la Glacière
95100 ARGENTEUIL (FRANCE)

Abstract

This paper presents the state of the art at CEPE concerning the Ultra Stable Oscillator (USO) designed and qualified for space applications, such as Telecommunications or Global Positioning satellites. A synthesis of all effective results obtained on the nine models, is given and shows that the technical objectives are achieved with the described design and the use of a QAS resonator, in particular the short term and long term stabilities, and thermal frequency stability requirements.

I.- INTRODUCTION

The new generation of telecommunication satellites such as TELECOM 2, DRS needs on board, a frequency reference of very high daily stability - in range 10-11/10-10 including all variables.

Similarly, for Global Positioning satellites, an on-board frequency is required with Ultra Short Term stability 10-13/10-12 for an integration time of 1 second.

In both cases, the frequency reference is located inside the satellite and must to have a life time covering the total mission time, which is now up to ten years under vacuum and of course to pass the launching without damage.

To meet all these specific needs, CEPE with funding from a CNES R/D contract, has developed an Ultra Stable Oscillator at 10 MHz, suitable to meet all these required characteristics.

II.-TECHNICAL SPECIFICATIONS

The main characteristics required to the 10 MHz U.S.O. are defined in the following paragraphs

2.1. Frequency accuracy

Setting accuracy after 8 hours :
$\pm$ 5.10-9

. Two options are available :

- the first one includes only an internal adjustment, easily accessible if it is necessary to readjust the frequency after a long period of storage.

- the second one contains an adjustment driven by a numeric word sent by a telecommand line as so to assure the frequency correction from ground when the USO is on board of the satellite.

2.2. Frequency stability

The frequency stability versus external parameters is defined in table 1.

2.3 Output characteristics

One sine output is available with the performances of table 2.

table 1

PARAMETERS	SPECIFICATIONS	REMARKS
a) Temperature over - 10 to + 50 °C	+/-2.5 10-11/10 °C +/- 2.10-10	with a slope of 10 °C/h under vacuum and at atm.pressure
b) Pressure	1.10-8/bar	
c) g-sensitivity	1.10-9/g	
d) Time Short term stability	10 ms 5.10-11 100ms 5.10-12 1 s 1.10-12 10 s 5.10-13 100 s 7 10-13 1000 s 1.10-12	Allan Variance
Mid term stability Mean square slope	4.10-13/mn	
Long term stability	3.10-10/day 1.10-9/month 1.10-8/year 5.10-8 over 10 years	
e) Retracability	5.10-9	cut off time of 24 hours after 24 hours of turn-on
f) DC supply variation	1.10-10	for a total variation of 5%
g) Load variation	1.10-10	for a RSWR < 1.5 on 50 ohms load
h) Magnetic field	5.10-11/Gauss	
i) Irradiations	1.10-12/day	

table 2

OUTPUT CHARACTERISTICS	SPECIFICATIONS	REMARKS
a) output level	0 to 12 dBm	
b) output lev. stability	1 dB	in temperature range
c) harmonics	40 dBc	
d) Non harmonics	100 dBc	
e) Phase Noise	10 Hz -120 dBc/Hz 100 Hz -130 dBc/Hz 1 kHz -140 dBc/Hz >10 kHz -150 dBc/Hz	
f) Source RSWR	>1.25 on 50 ohms load	

2.4 DC Supply interface

The DC supply characteristics are as follow :

DC voltage : 24 V or 20 V ± 5 %

Ripples on the DC line : 50 mV or 480 mV peak to peak between 10 Hz to 50 MHz depending on the application.

With a 50 mV ripples, all specifications regarding short term stability, thermal stability and phase noise have to be met.

With a 480 mV ripple, only the non harmonic specification is required.

Input powers are defined in the table hereafter :

INPUT POWER	AMBIANT pressure	under VACUUM
Max	8 W	8.0 W
at 0 °C	7 W	3.5 W
at 20 °C	5 W	2.5 W

2.5 Environmental conditions

The qualification levels for mechanical and thermal environments are :

a) Random vibrations

10 Hz to 100 Hz + 4dB/oct
100 Hz to 1000 Hz 0,2 g2/Hz
1000 Hz to 2000 Hz - 3dB/oct

composite value : 18.2 g2/Hz
on the 3 axes,
with a duration of 2 mn/axis.

b) Constant acceleration :

20 g / 60 s
On each of the two directions of the 3 axes.

c) Sine vibrations :

10 - 100 Hz / 20 g
2 oct / mn 3 axes

d) Thermal environment

Operating : - 10 to + 50°C
Non operating : - 40 to + 70°C
cold start-up : - 30°C

e) Irradiations : TOPEX dose curve

f) Life time : Minimum 10 years

III.- USO DEVELOPMENT PHASES

Only two years were needed to design and to qualify the Ultra stable Oscillator with all its characteristics complying the technical specifications described in paragraph II, and with respect of the product assurance requirements specific to space projects.(1)

3.1 Phases A and B : Feasability and Design phase

The first stage has consisted to design the USO taking account all of all the technical specifications but also the specific space requirements as selection of parts, materials, processes in Preferential or Qualified lists(2), checking of stresses applied on parts according to the Space rules,(3) and so.

The optimization of the design has been concluded by both theoretical analyses - mechanical and thermal - and experimental results.

During this phase, were also conducted qualification works on specific technologies used for the first time by CEPE in a space application, according to programs approved by CNES. The qualification

reports have been issued and presented during the first Design review.

After the Preliminary Design Review (PDR) conducted by the CNES, with intent to analyse and to make recommendations, if needed, to improve the presented design, one Elegant Breadboard model, representative of the Flight Model (FM) design, has been built and tested according to a test sequence with qualification level.

The test results obtained on this first model has demonstrated its ability to meet the foreseen technical and environmental requirements.

3.2 Phase C : Engineering and Qualification Phase

The success of the first stage has called for the manufacture and the qualification of six engineering models (EQM). The configuration of these models is the same as the FM units' , except for parts procurement level, Military instead of space.

These EQM units were built according to the same manufacture flow chart, with the same procedures and same inspections as Flight Models.

In addition of the EQM manufacture, all the documentation, technical and product assurance as well, edited in preliminary form for the PDR, has been up-dated according to CNES comments.

This documentation includes :

- all technical analyses
electrical, thermal and mechanical

- the reliabilty file with

. the failure modes, effect and criticity analysis,
. the parts application review
. the worst case analysis
. the radiation sensitivity analysis
. the reliability prediction

- the design and manufacture files

- the product assurance file :
. Quality plan describing the quality organization and rules carried out to meet the quality assurance requirements
. The declared parts, materials and processes lists
. The qualification plans and reports of specific processes implemented in the design of the USO.
.The configuration management plan showing the management and the control of the documentation issued in the project.

This phase was closed by the Critical Design Review (CDR) that has successfully concluded in favour of the qualification of the presented USO design for space applications.

IV.- TECHNICAL DESCRIPTION

4.1 Thermal and mechanical design

Due to the required thermal performances and the need to pass with success the vibrations spectrum, the design of the mechanical structure has been defined in taking account simultaneously of these both requirements .

The USO design is based on a stracking boxes concept made by three boxes, defined in metal nature, dimensions and shapes, as so to minimize the convection effects and to reach the same thermal behaviour at atmospheric pressure or under vacuum.

1) The inner box, a copper metal quasi-cylinder, is ovenized by power transistors, welded on 2 faces. The crystal resonator in HC40U holder and the oscillator network are fitted in this oven.

To avoid thermal effects, a particular care is taken in the choice of wires, parts and mounting aspects.

2) The intermediate box, in aluminium alloy, not ovenized but with internal homogeneous temperature environment contains the oven, fixed by screws, through thermal insulators.

The thermal control board is also fixed inside this box, in order to benefit by the thermal environment and also to make the connections with the power transistors easier.

15 wires are needed to assure these connections.

Here again, their diameter, length and location are defined to minimize the thermal effects, to avoid mechanical stresses during vibrations, and to ensure a mounting/demounting facility.

Outside, on 2 faces, are fixed two PC boards :

- the first one contains the DC power distribution with filtering function, the HF amplifiers and the telemetries,

- the second one, on the top, easily accessible is the frequency adjustment function, and the telecommand interface when required.

3) The external box, holder of the USO, improves the thermal and mechanical behaviour.

The thickness of its walls is defined to reduce the radiation effects on the sensitive components.

To assure the electrical interface, two connectors D*M type and one RIM/SMA coaxial connector are located on one face.

The mechanical interface is realised by fixation with four screws M4.

All mechanical parts are manufactured from solid metal and assembled by screw to ensure the mechanical behaviour.

4.2 Electrical Design

To ensure the ultra short term stability and the low temperature sensitivity the main features of the electrical design are :

- the use of a QAS resonator, 10 MHz, SC cut with very high performances of short term and long term stabilities,

- the design of the thermal control using original PID function in the loop,

- the design of each function after analyse of the external effects, specially temperature, DC supply and time drifts, so as to optimize the schematic, as well in the parts type selection as in the values really used.

The electronic diagram contains four functions :

1) the metrologic function consists of the oscillator built as using a QAS 10 MHz resonator in PQ swept material, and the two amplifiers stages.

2) The thermal control, including a PID in the loop allows to maintain the oven at the given temperature, adjusted with an accuracy better than 0.1 °C for each unit at the turn-on point of the thermal characteristic, with a precision better than 10-3°C.

The temperature information is given by a thermistor, with a very high long term stability, and the heating is assured by power transistors.

3) The DC power distribution with the specific filtering function at the head, to avoid the ripples effects on the output signal is built to obtain for each function a separate and filtered DC voltage with a good stability.

4) Two telemetries have also designed to detect the possible disfunction and delivere a DC analogic voltage representing the information to be checked :

- current of the heating,
- RF output level.

When a telecommand interface is required, this one allows to control the adjustement by a numeric word applied on a DA converter and to keep this numeric state between two frequency controls.

V.- USO PERFORMANCES

Today, CEPE experience in the domain of USO for Space applications, is demonstrated by the results obtained on the first models.

Since 1988, have been built twelve units including six Engineering Models and five Flight Models, according to the same configuration, with nevertheless, some minor variants well-identified.

The tables 3 to 7. summarizes for nine models the effective results and gives comparison to the required specifications defined in the paragraph II.

5.1 Frequency stability

The table 3 details the frequency sensitivity versus external parameters obtained for each unit.

On this table, the stability in temperature range is given peak to peak. Typical curves are given in figures 1 and 2.

We see, from the set of values listed for all units, that the proposed design allows to achieve the same behaviour at atmospheric pressure and under vacuum.

Morever the DC supply and load sensitivities are very low and show the efficiency of the electronic design and particularly, of the DC supply distribution.

About the magnetic field sensitivity, the given results in this table are the values obtained in the worst axis.

It appears two groups of results, the first four models have a mean sensitivity of about 1.5.10-10/Gauss when the others have a mean sensitivity divided per 20. This improvement is realized with the change of the technology of oscillator components.(5)

The pressure and g sensitivities, short term and long term stabilities are primarily depending on the performances of the resonator used.

So, this one is characterised before its integration inside the USO. The table 4 shows the good correlation between the results obtained on the

resonator alone and when in the USO. The given value of long term stability is the slope calculated per day after one month of continuous operation.

Typical results are shown in figure 3 , ageing on one month at atmospheric pressure, and in figure 4 , daily stability under vacuum.

About irradiation behaviour of the USO, theoritical analyses have been performed to calculate for each sensitive component the equivalent cumulated dose, taking account the USO structure and the satellite environment and to evaluate the impact of component sensitivity on the frequency.

So, in the case of TOPEX satellite, the total dose seen by the resonator a QAS in PQ SWEPT crystal, during the mission time, is estimated to 6 Krad and implies a frequency drift of about 1.10-8 .(6)

Tests performed by CNES directly on USO units, with a slope of 1 rad/h have shown a frequency drift of about 1.10-12/rad,(7) in good agreement with the calculated drift.

table 3 :Frequency stability results

Para	spec	unit	EM01	EM02	EM03	EM04	EM05	EM06	FM08	FM10	FM09
0 to 40°C. atm.pressure	2	10-10	1.0	2.0	1.2	1.2	0.5	1.9	NA	NA	NA
0 to 40 °C under vacuum	2	10-10	1.0	1.3	0.5	1.4	0.1	1.7	0.5	0.7	0.3
pressure	1	10-8	0.6	0.7	0.6	1.0	1.0	0.9	0.4	1.0	0.2
g-sensitivity	1	10-9	0.2	0.6	0.1	0.1	0.2	0.2	0.1	0.3	0.2
time short term											
1 s	10	10-13	4	4	6	4	4	4	5	5	8
10 s	5	10-13	3.5	4.0	4.5	4.8	3.6	3.8	3.8	3.9	4.8
100s	7	10-13	4.5	4.5	7.0	NA	NA	NA	NA	NA	5
1000s	10	10-13	16	13	14	NA	NA	NA	NA	NA	14
mid term stability	4	10-13 /mn	NA	NA	NA	3	1.5	2.8	3.0	3.0	NA
long term stability per day	10 3	10-11 10-11	3	3	2	2	2	3	5	10.	3
retracability	5	10-9	0.4	0.4	0.2	0.1	0.01	0.15	3.50	0.1	1.00
DC supply	1	10-11/ %	0.4	0.8	0.6	0.4	0.4	0.6	0.6	0.6	0.3
load	1	10-10	0.4	0.4	0.2	0.15	0.3	0.3	0.3	0.3	0.3
magnetic field	5	10-11 /gauss	13	20	10	20	1.5	0.4	0.3	0.5	NA

NA : not applicable

table 4 : Comparison between resonator and USO characteristics

	EM 01	EM 02	EM 03	EM 04	EM 05	EM 06	FM 08	FM 10	FM 09
PRESSURE SENSITIVITY	0.7	0.4	0.2	1.0	1.0	0.8	0.3	1.0	0.2
10-8/bar	0.8	0.6	0.3	1.0	1.0	0.9	0.4	1.0	0.2
g-SENSITIVITY	2.8	6.2	1.1	1.3	2.0	2.4	1.4	4.6	2.2
10-10/g	2.3	6.5	1.0	1.2	2.0	2.3	1.0	3.0	2.0
SHORT TERM STABILITY	3.2	4.2	4.7	4.2	2.2	4.2	3.2	4.5	3.6
10-13	3.5	4.0	4.5	3.8	3.6	3.8	3.8	4.2	4.8
LONG TERM STABILITY	3	2	1	2	3	3	1	4	1
10-11/day	3	3	2	2	2	3	5	10	3

5.2. Phase noise performances

The figure 6 shows the phase noise performances obtained on these USO.

From this curve ,we have reported in table 5 the Phase noise table characteristics for particular Fourrier Frequencies in order to compare with DRSS USO phase noise specifications.(8)

table 5

FOURRIER FREQUENCY	L(f) dBc/Hz	DRSS USO spec.
0.5 Hz	-112	-105dBc/Hz
5.0	-136	-135
100	-153	-149
1000	-158	-149

The table 6 shows all results obtained on the nine models.
All values are grouped in a window of +/- 3 dB and are depending on the effective output level, two different requirements, low level 0 to 3 dBm for models SN 04, 05, 06, 08 and 10 and high level 8 to 12 dBm for the others.

table 6 : Phase Noise performances in dBc/Hz

	Pout	**PHASE NOISE L(F) dBc/Hz**				
SN	**dBm**	**1 Hz**	**10 Hz**	**100 Hz**	**1 KHz**	**10k Hz**
01	10.	-122	-143	-153	-158	-158
02	9.4	-120	-145	-153	-157	-157
03	9.8	-120	-142	-152	-158	-158
04	1.9	-119	-142	-152	-158	-158
05	1.9	-120	-140	-150	-155	-157
06	1.6	-118	-140	-148	-153	-155
08	0.4	-116	-137	-146	-153	-157
10	0.5.	-118	-138	-146	-155	-156
09	9.8	-120	-140	-148	-155	-157

5.3. DC Power characterictics

The table 7 gives the DC power measured at ambiant pressure and under vacuum versus temperature for each model

table 7 : DC power in Watts

	Atmospheric Pressure.			Under vacuum		
SN	0 °C	20 °C	40 °C	0 °C	20 °C	40 °C
01	7.2	5.3	3.4	2.4	1.9	1.5
02	7.0	5.2	3.4	2.4	1.9	1.5
03	7.0	5.2	3.3	2.4	1.9	1.5
04	5.9	4.3	2.5	1.9	1.6	1.1
05	5.6	4.0	2.2	1.8	1.4	1.0
06	6.4	4.7	2.9	2.0	1.6	1.2
08	6.5	4.7	2.9	2.0	1.6	1.2
10	6.0	4.2	2.3	2.0	1.5	1.1
09	6.8	5.0	3.2	2.4	1.9	1.5

A good correlation can be made between the oven temperature and the measured DC power values. The reproductibility of the values listed in each row of this table ,versus a same thermal state , shows that the design , without a DEWAR vase , allows to ensure the same thermal behaviour independent of the model unit assembly.

5.4. Mechanical characteristics

Each model was successfully tested under vibrations . Frequency measurements performed before and after each axis test, shows a frequency drift less then 5.10-10.

The weight of the USO relatively high due to an oven in copper metal is between 1080 g and 1150 g, depending on the required options.

Details of outline dimensions (105 * 73 * 94.5 mm) are represented in the figure 7.

VI. CONCLUSION

The new generation of telecommunication or earth observation satellites need on-board a frequency reference with highly daily stability, in 10-11/10-10 range, Ultra short term stability, in 10-12/10-13 range and very low sensitivity to environmental parameters, as temperature, pressure, acceleration magnetic field and total dose radiation.

For mass, volume, power consumption and life time considerations, it is preferable to use an Ultra Stable Crystal Oscillator for the Frequency Reference.

To be present on the European and U.S. market of USO, CEPE, with funding of a CNES R/D contract, has specifically developed and qualified for space applications, an USO at 10 MHz, based on the stracking boxes concept and the use of a QAS resonator.

Today, after completion of the first Flight Models and delivery of some of them, - two units are actually under integration phase in the TOPEX satellite - CEPE USO design has demonstrated its ability to meet and sometimes exceed the technical and environmental specifications.

The specific CEPE Space organisation, stand up since two years, allows to meet at the best, the Project and Product Assurance requirements.

CEPE experience in USO oscillators and in Space Projects as well, permit today, to answer any requirement about USO and others products , and to ensure the best prestations in terms of service, prices and delivery times.

BIBLIOGRAPHY

(1) Product Assurance Management and Audit System for ESA Spacecraft and Associated Equipments
ESA Document ESA PSS 01-10

(2) CNES Preferential Part List
CNES Document QFT/IN 500

(3) Derating Requirements and Application Rules for Electronic Components
ESA Document ESA PSS 01-301

(4) Further Results on 5 and 10 MHz Resonators of BVA and QAS design.
J.P.AUBRY - A.DEBAISIEUX ,CEPE
38th Annual Symposium of Frequency Control

(5) Magnetic Sensitivity of Quartz Oscillator
F.DEYZAC ONERA
EFTF 90 Proceeding

(6) Influence de la Qualite du Matériau sur les Performances des Résonateurs à Quartz
J.P.AUBRY X.BUISSON M.FRAGNEAU
J.GUILLEREY D.SCHOTTAND
Revue Technique THOMSON CSF Vol 19 N°2Juin 1987

(7) Etude Expérimentale sur la dérive de Fréquence des OUS Spatiaux soumis à Irradiations.
M.BRUNET CNES
Rapport CNES Ref RA/TE/AE/IT/TF N°276

(8) Frequency Sources for Space Applications
S.J.FELTHAM ESA
EFTF90 Proceedings

figure 1 : thermal stability at atm.pressure

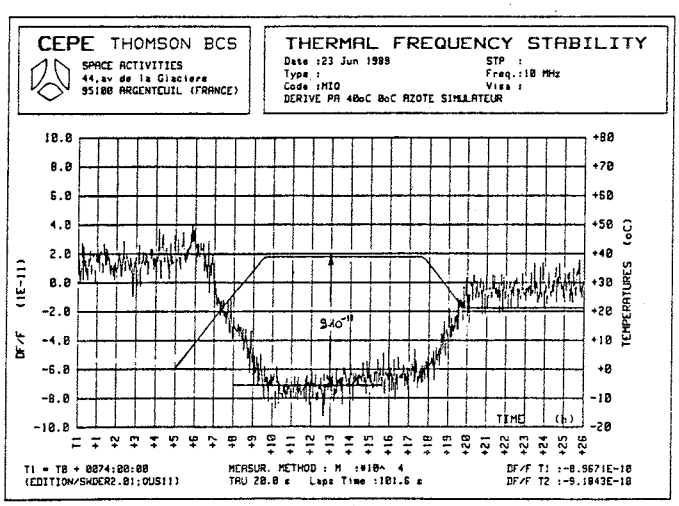


figure 2 : thermal stability under vacuum

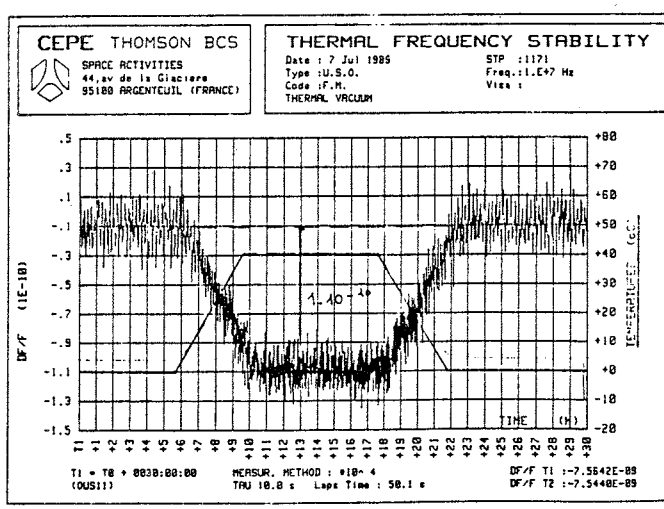


figure 3 : ageing of a QAS resonator 10 MHz

slope/day : 3.1 10 -12;

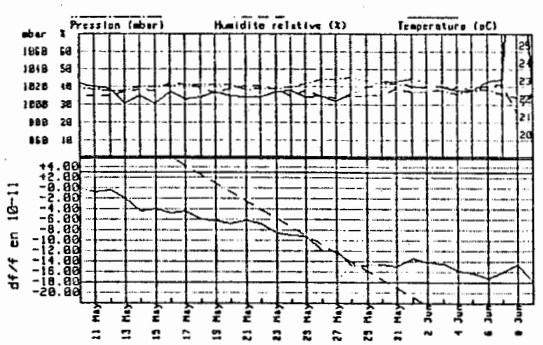


figure 4 : Ageing of a USO under vacuum

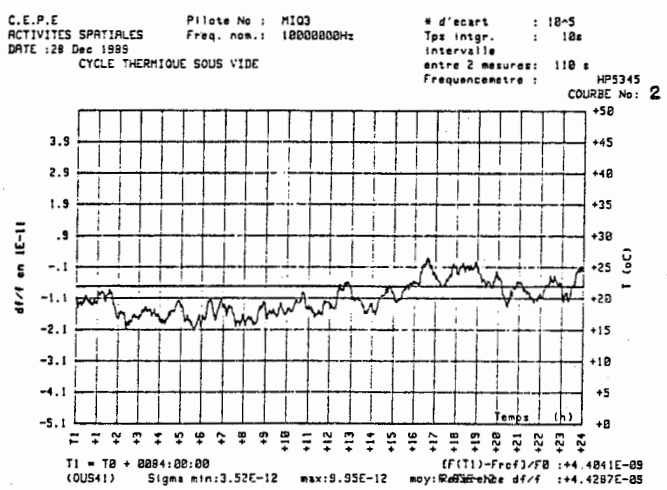


figure 5 : short term characteristic

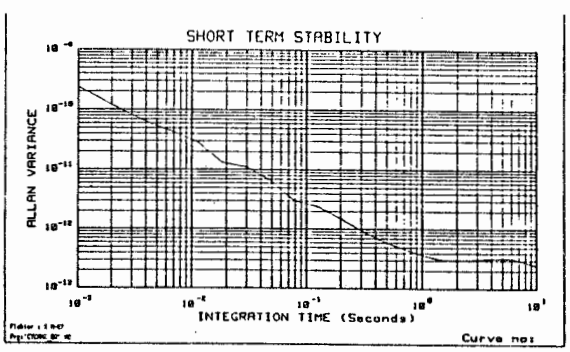


figure 6 : phase noise characteristic

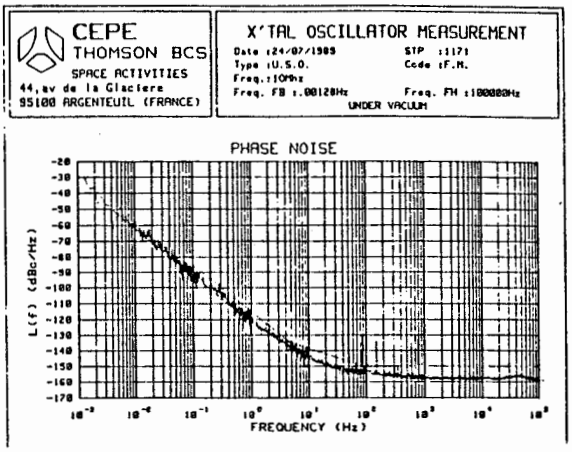
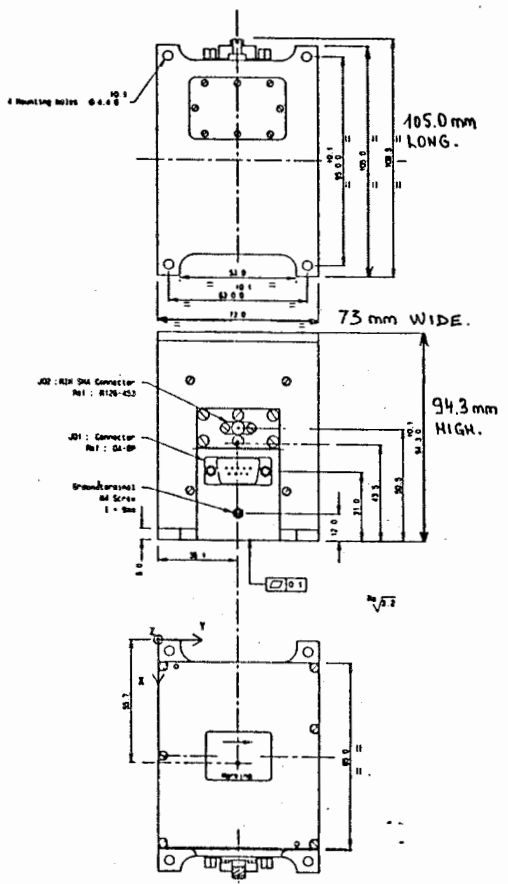


figure 7 : USO Interface Drawing



SESSION 7A

RESONATORS III

Chairman : W. PFLAUM

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 7A Number 65

NEW RESULTS CONCERNING QUARTZ AND BERLINITE RESONATORS OBTAINED BY SYNCHROTRON RADIATION X-RAY TOPOGRAPHY

A. ZARKA*, B. CAPELLE*, Y. ZHENG*, J. DETAINT**, J. SCHWARTZEL** and E. PHILIPPOT ***

* Laboratoire de Minéralogie-Cristallographie - Université P. et M. CURIE
4, Place Jussieu 75252 PARIS Cedex 05 France

**C.N.E.T. dept P.A.B./B.A.G./M.C.T. 196,Av Ravera 92220 BAGNEUX France

***L.P.C.M. Université de Montpellier Place E.Bataillon 34000 MONTPELLIER France

SUMMARY Further observations of the influence of the defects on the vibration modes were performed mainly on very high Q factor plano-convex resonators vibrating on the fifth overtone. These resonators were obtained from quartz with dislocations or growth bands. It was observed that the normal extra Pendellösung fringes due to the vibrations are perturbed and can eventually disappear when the density of dislocations is high. These fringes are also very deformed by the presence of growth bands in natural quartz resonators. From these observations it can be induced that these defects have detrimental effects on the spatial coherence of the vibrational state. Similar observations have been done on berlinite resonators.

I - INTRODUCTION

Many works are presently done to improve the stability of the more stable quartz oscillators and also to reduce the stabilities degradation due to the environmental variables such as the forces, the acceleration and so on (1). One of the reasons of the sensibility to these factors is the modification of the symmetries of the vibration modes due to the presence of the defects in the materials (2). Another reason, probably important, is the presence of coupled components, eventually non stationary of the vibration mode which modify the symmetry and can also spread over the whole surface of the slice (3). In the aim of understanding the influence of the defects on the vibration modes, investigations were performed on various resonators containing different amounts of defects. In AT-cut quartz samples without defect the observations obtained by stroboscopic topography with the use of the synchrotron radiation, concerning the contrast of the vibration modes were correlated with the results obtained by different theories. In the commercial samples the dislocations density is generally about 10^2 to 10^3/cm2 , this leads to a considerable deformation of the mode shape. The natural crystals usually present a reduced amount of dislocations but very often show growth bands which have also a negative effect on the geometry of the vibration mode. The dislocation density observed in berlinite samples is similar to that observed in synthetic commercial quartz and interactions between the defects and the vibration modes are also observed. The purpose of this study was to determine the role of the defects on the acoustic vibrations and also to examine some properties of the vibration modes in quartz and berlinite.

II - STROBOSCOPIC X-RAY TOPOGRAPHY

The stroboscopic technique used, was that employed in our previous experiments, it is described in details for example in (4). The stroboscopic X-ray topographs are obtained using a double axis spectrometer at the L.U.R.E. (5). This spectrometer permits to obtain a Laue pattern of topographs when the sample is placed in the incident white beam (6). An interesting possibility of this apparatus consists in the monochromatic setting. In this case, on the first axis is adjusted a 110 sample of germanium which selects from the white beam only one wavelength. Then, monochromatized images from the resonator adjusted on the second axis, can be obtained. Furthermore in the two previous cases, using a very fine slit section topographs can also be obtained.

III - STROBOSCOPIC X-RAY TOPOGRAPHY OF EXCITED QUARTZ RESONATORS

The theory for slightly distorted crystals (7) was applied to the case of AT-cut quartz resonators to interpret stroboscopic topographs. The theoretical understanding of topographs of strongly deformed crystals is more difficult than the case of slightly distorted crystals for which several theories are available (Kato etc..). So that most of our experiments concern resonator vibrating with at a low amplitude. The features in the diffraction process, such as ray-trajectories, phase changes and intensities of diffracted waves were drown from this theory. It was shown that extra Pendellösung fringes appearing on section topographs could be related to the vibration amplitude within a precision of about 0,3 Å. The experimental images were compared with Kato's theoretical profile and simulated images obtained by solving Takagi's equations. In the following, examples of section and traverse topographs of weakly excited quartz resonator are presented.

III.1 - X-ray section topographs of weakly excited quartz resonators

The resonators were made in such a way that the resonance frequency of the (5.0.0.) mode is exactly the pulsed synchrotron-radiation frequency or one multiple of this frequency ($\Omega_{5.0.0}/2\pi$= 2 x 3.169280 MHz). The electrical excitation of the resonator is derived from a signal taken from the radiation pulses, so that the crystal is forced to vibrate at its (5.0.0.) resonance frequency and in synchronism with the radiation pulse. A phase shifter permits a variation of the relative time phase between the radiation pulse and the vibration, i.e. one may observe the vibration at different stages of the period of vibration, with a time resolution of 1 ns.

The crystals were set with a Bragg angle Θ of 8.068°, so that the selected incident X-ray wavelength λ for the $\bar{2}1.0$ reflection was about 0.689 Å. An experimental image obtained with a high vibration amplitude is shown in Figure 1b. Only one half of the section is presented. The excitation voltage V was 6.046 mV and the section topograph was taken near the middle of the electrode on the resonator with its centre at x_{1_s}= 1.11 mm, x_{3_s} = 0 . The value of the shear displacement estimated by the piezoelectric theory, was very nearly $u_0\{x_{1_s}$=1.11 mm, x_{3_s}=0.00$\}$=4.30 Å (7). The determination of this value of the shear displacement using the intensity profile derived from Kato's theory, leads to $u_0\{x_{1_s}$=1.11 mm, x_{3_s}=0.00$\}$=4.80 Å (Fig 1a). Another determination was performed by solving

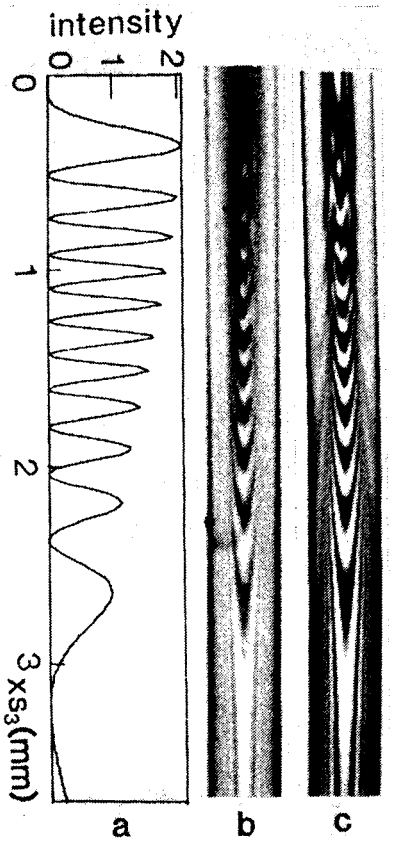


Figure 1: Extra Pendellosung fringes in low amplitude vibrating plano-convex resonators

1a: Intensity profile derived from the theory of Kato.

1b: Experimental section topograph.

1c: Numerical simulation made by resolution of the Tagaki's equations.

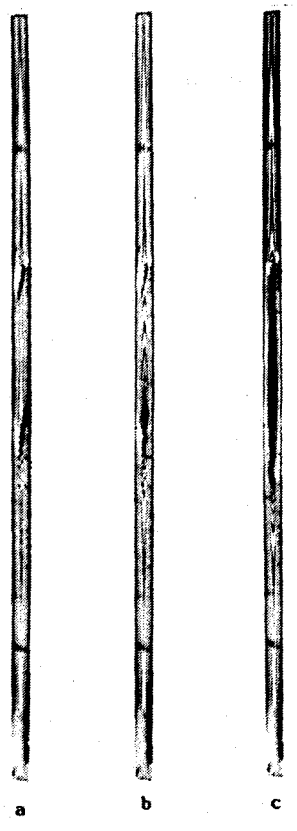


Figure 2: Perturbation of the extra-fringes due to the presence of a dislocation.

2a: section without vibration.

2b: low amplitude of vibration.

2c: large amplitude of vibration.

numerically the Takagi's equations and it was found that the image computed with a shear displacement $u0\{x_{1_s}=1.11, x_{3_s}=0.00\}=5.00$Å reproduced the experimental fringes position best (Fig 1c). This example shows that extra Pendellösung fringes can be understood and analysed by Kato's theory. This means also, that for resonators operating at higher excitation levels, one can always apply Kato's theory for zones of low amplitudes and get vibration parameters. The perturbation of the fringes due to the presence of dislocations is given also in figures 2a, 2b and 2c where, their deformations due to these defects may be observed . Respectively in figure 2a, 2b and 2c are shown a section of a crystal without vibration, with a small amplitude of vibration, and with a higher vibration level.

III.2 - Stroboscopic X-ray traverse topography of quartz resonators

Stroboscopic X-ray traverse topographs were performed for a vibration amplitude range similar to previously. A typical image is shown in Figure 3a where the estimated value of u_0 is 6.2 Å. In this figure concerning an AT plano-convex resonator made with an experimental synthetic crystal (SICN) having an exceptionally reduced dislocation density, "Circular contrasts" can be seen. These contrasts are perturbed by two dislocations. In figure 3b is shown a corresponding section topograph taken in the middle of the sample. The traverse topographs can be understood as an integration of section topographs over the whole crystal, so that, the "circular contrasts" result of the extra fringes observed in section topographs. As it was shown that the extra fringes positions are determined by the local vibration amplitudes, "the circular contrasts" represent most probably the equal-amplitude curves of the vibration. This point is supported by the two following remarks. First, for a same excitation voltage, the number of "circular contrasts" on traverse topographs is found to be the same as the number of extra fringes on the corresponding section topographs. Secondly, these

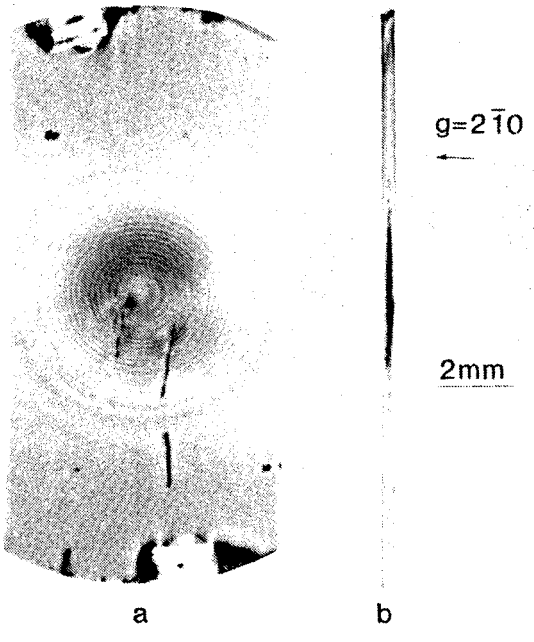


Figure 3: Relation between traverse topograph and section topograph.

(Experimental synthetic quartz with an exceptionally low dislocation density)

3a: traverse topograph.

3b: section topograph.

contrasts have a slight ellipticity which corresponds to the elliptic equal-amplitude contour from the piezoelectric theory. It is recalled that the equal-amplitude curves are defined by :

$u_0 \exp \{- a_5 x_1^2 - b_5 x_3^2\} = \text{constant}$

or

$a_5 x_1^2 + b_5 x_3^2 = \text{constant}$

and the ratio between the two axial lengths of the ellipses is then equal to:

$a_5 / b_5 = 0.935$.

This value was actually measured on the experimental images.

The fact that the "circular contrasts" of stroboscopic topographs can be related to the vibration amplitude provides another method to study the vibration in the resonator since the examination of the equal amplitude curves can give directly a determination of the spatial dependence of the components of the displacement. This is very useful for comparison with theoretical models and to study the vibration modes of resonators for which no theory is already available. The perturbation of such contrasts by dislocations is also a method to study the influence of dislocations on acoustic vibrations. This can be done, in considering use the bidimensional mapping constituted by the deformation of these "circular contrasts" that gives informations on the departure of the lateral variations of the vibration amplitude from that of the theoretical case which is observed in perfect quartz crystals. This permits thus to estimate the global effect of the dislocations. This will be carried out in several of the following examples. Another method that can give more quantitative informations and makes use of section topographs was presented in reference (7).

IV - INFLUENCE OF DISLOCATIONS ON THE VIBRATION MODE.

IV.1. - Quartz crystals

When the density of dislocation in the resonator is increased by choice of the material, traverse topographs, different of the previous one, are obtained with more and more complicated contrasts.

As an example in Figure 4, is presented a topograph of an AT-cut quartz resonator made from an experimental crystal of extreme purity grown by the RADC Laboratories. The number of dislocations in this sample is much lower than in commercial samples but higher than that observed in figure 3. In this case, we observe the same elliptical fringes contrast as in the previous example (Fig 3a).

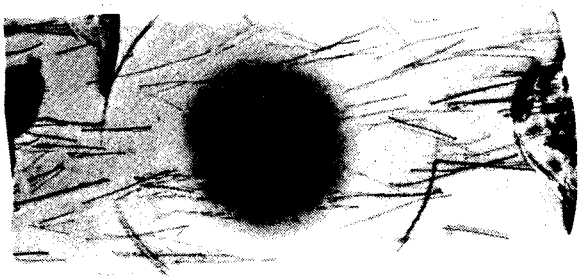


Figure 4: Traverse topograph of a resonator made with a very high quality quartz (low density of dislocation).

Other examples are given in Figures 5 and 6 which correspond to resonators made with commercial crystals with a large density of dislocations. It may be observed in Figure 5a that this resonator (not in vibration) was made using a crystal with a large density of dislocations. When this crystal is excited at its 5.0.0.resonance, we can observe (Figure 5b) that this large density of dislocation induce a great perturbation of the mode shape in the vibrating zone. The vibration, instead of having a quite circular form is completely deformed up to the edge of the sample.

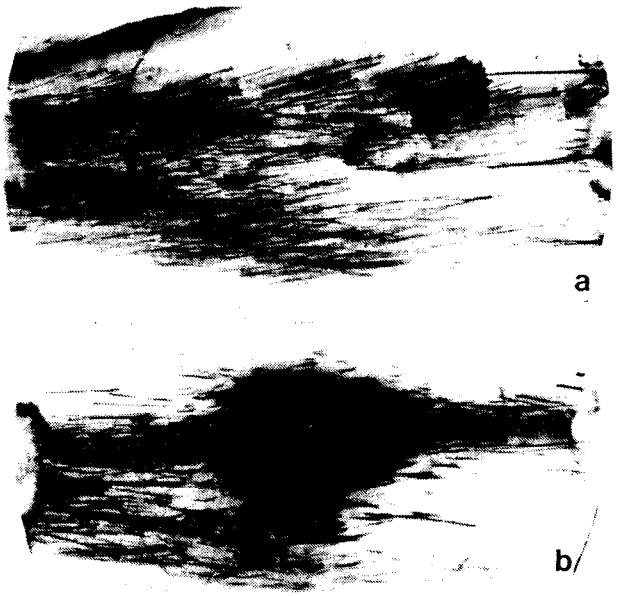


Figure 5: Traverse topographs of a resonator presenting a large density of dislocations
5a: without excitation
5b: with a large amplitude of vibration.

The example given in Figure 6 represents a topograph of another AT-cut quartz resonator. In this sample, the amount of dislocations is important and it can be seen that the classic central black contrast is perturbed by the dislocations. The deformation fringes corresponding to the region where the thickness shear mode has a small amplitude can also be observed, in this figure, in the external region of the vibration mode.

So, these observations show that the dislocations contribute to unconfine the theoretically gaussian mode and can be , by this way, at the origin of acoustical energy losses in the resonators.

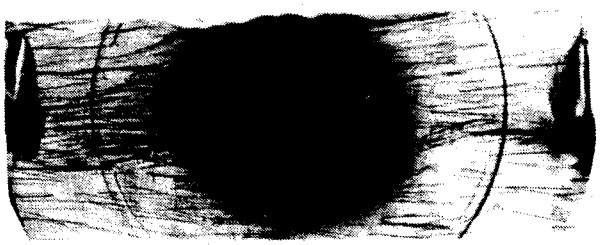


Figure 6: Traverse topographs of a plano convex resonator made with Premium Q commercial quartz.

A last example is given in Figures 7a, 7b, 7c, 7d and 7e. This crystal is a Q.A.S. Which was made from premium Q commercial synthetic quartz (S.I.C.N. Company). In Figure 7a is presented one topograph of the crystal out of vibration. In figures 7b, 7c, 7d are respectively shown stroboscopic topographs of same the resonator under vibration(large excitation), taken at different times of the period of vibration. These times of observation (apparent duration of observation 1 ns) in the period are measured by the relative phase between the resonator current and the X-ray pulses (4) ; the later ones were respectively $\Phi = 0°$, $\Phi = 45°$ and $\Phi = 90°$. In figure 7e the resonator is excited with a still larger large signal and is observed with a relative phase $\Phi = 90°$. In Figure 7b and in Figure 7e, a double central contrast due to the high level of the vibration may be observed. This coupling was observed to be related, in

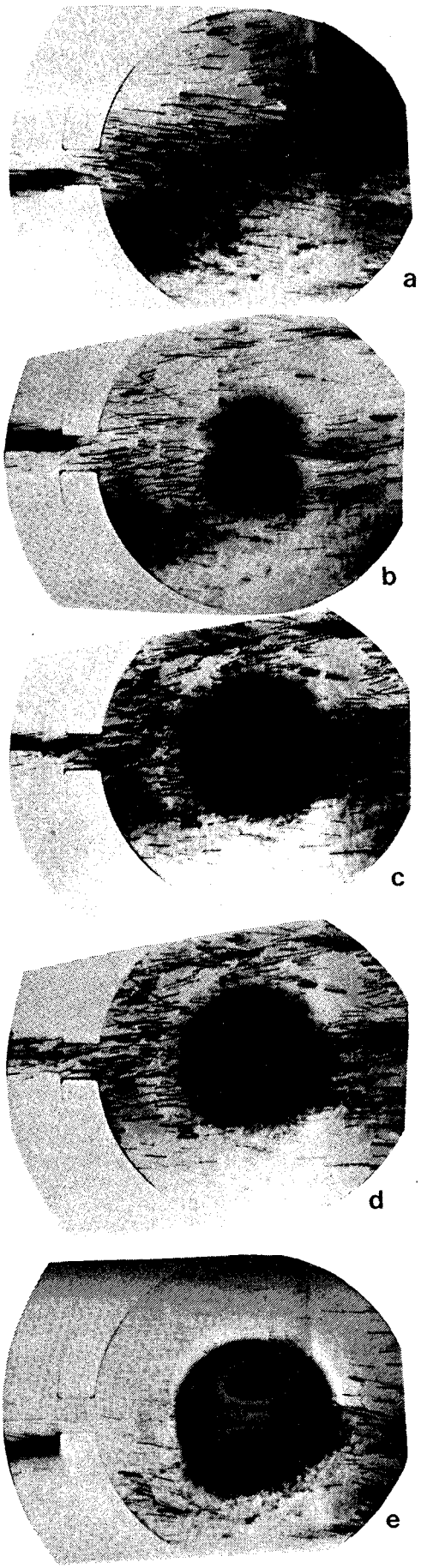


Figure 7: QAS resonator made with premium Q commercial quartz.

several resonators, to the apparition of a non linear effect (the "isochrosnism defect"). The bistable states appearing in highly excited crystals seem to be related with modes presenting different states of coupling between the 5.0.0. mode and other modes which as, in figure 7b, look very like anharmonics.

On the whole, the observations which have been made on the quartz resonators, concerning the influence of dislocations, have shown that they have a detrimental effect on the shape of the thickness shear vibration mode. The most important feature is the loss of the intrinsic symmetries of the modes. This has much probably important consequences on the electrical parameters of the resonators and also on the effects of some environmental variables such as the forces and the acceleration on the frequency stability.

IV.2 - Berlinite resonators.

In the case of berlinite crystal, it was observed enough similar contrast, although the quality of the sample is obviously less good than for the best quartz crystals. Nevertheless, some correlations between the mode shape of the resonators under vibrations and the defects were observed. In Figure 8 is presented a topograph of a slice of berlinite. It may be observed that is an important number of dislocations which have been created on the seed at the start of the growth. The two white contrasts correspond to mounting clips. This crystal is not vibrating. A second example is given respectively in Fig. 9a, 9b and 9c where we can observe the influence of the dislocations on the vibration mode. In the first topograph (Fig. 9a) numerous dislocations are observed on the two sides of the seed. Progressively from Figure 9a to Figure 9c the amplitude of the vibration is increased. We can observe that is a clear interaction between the defects and the vibration, as in the case of quartz crystals.

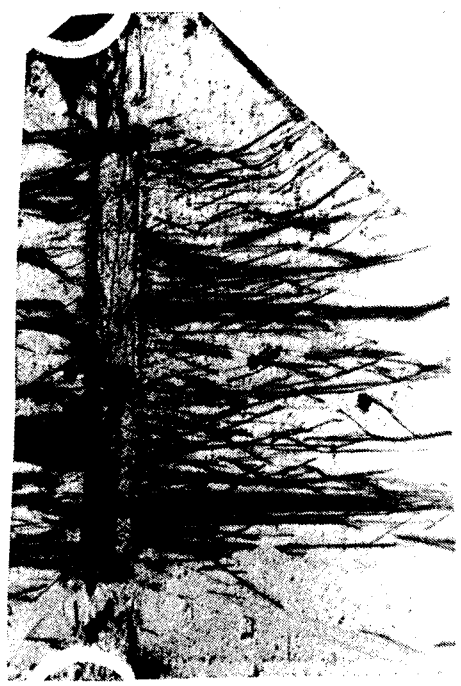


Figure 8: Plane berlinite plate (without vibration) .

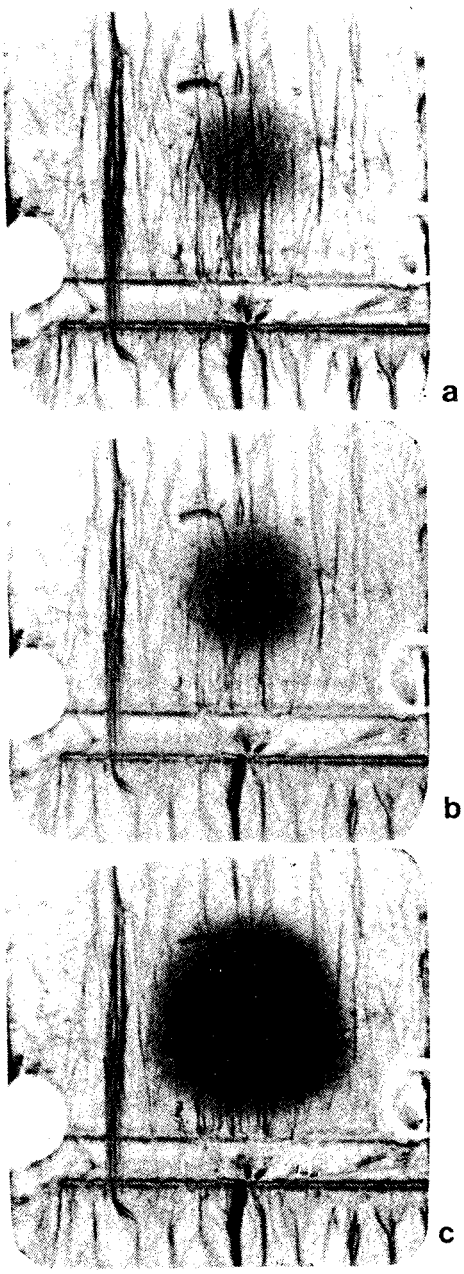


Figure 9: Plane berlinite resonator.
9a: Very low amplitude of vibration.
9b: Low amplitude of vibration.
9c: Large amplitude of vibration.



Figure 10: Plane AT berlinite resonator made across the seed.

Another example is given in Figure 10. In this case the electrodes have been set in the seed region and we can also observe a supplementary perturbation probably due to a difference between the electroacoustical properties of the seed and of the over-grown crystal.

The last example is given in the Figure 11. In this topograph we can observe numerous dislocations in the resonator and several types of others defects due to the mechanical process (Fig.11a). This plano-convex resonator was made to vibrate at the fifth overtone. In the following conventional topograph (Fig. 11b), which corresponds to one reflection of a same Laue pattern, the negative effect of the dislocations on the vibrating mode can be again observed. We can particularly notice the very asymmetric form of the central vibrating zone, and the interaction of the dislocations with the vibration.

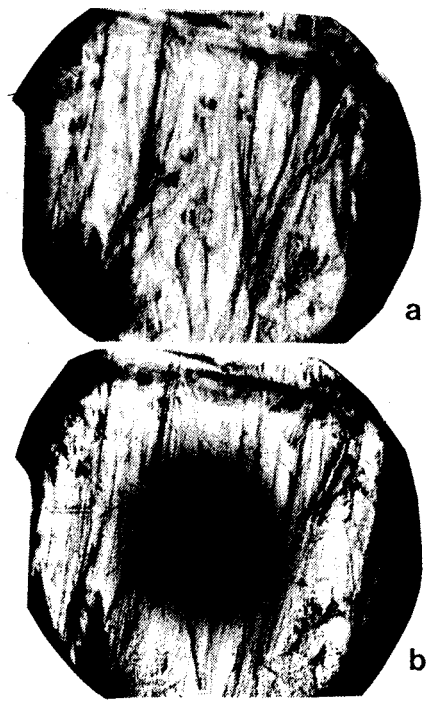


Figure 11: Plano -convex AT berlinite resonator.
11a: without vibration. 11b: vibrating.

V - ROLE OF THE GROWTH BANDS

In the case of growth bands of natural quartz crystals which are extended defects, stroboscopic experiments were performed on traverse topographs to get a global view over a resonator. The reflection parameters used are similar to the previous cases and the studied plano-convex resonators operated also on the fifth overtone of the thickness shear mode (5.0.0). Figures 12a-d show a series of traverse topographs of resonators made with natural crystals having growth bands, and, submitted to increasing vibration amplitudes. In Figure 12a the image of the growth bands without acoustic vibration can be observed. In Fig.12b and 12c the vibration amplitude is low and characteristic "elliptic curves" of the acoustic vibration exist even in presence of growth bands, but they are deformed in the present case. Figure 12d, taken with a higher vibration amplitude, shows a dark contrast on the central part of the resonator, as expected. On the edges of this contrast, which represents the envelop of the vibration amplitudes, contrasts of irregular forms can be seen. This indicates that growth bands have an important effect and that they perturb the spatial coherence of the acoustic wave. Another example of the effects of the growth bands is given in Figure 13a and Figure 13b. These are also topographs of a resonator made using natural quartz. In Figure 13a, where the crystal is out of vibration, it can be observed that the resonator presents in the central part growth bands. In Figure 13b, the contrast due to the vibration is greatly deformed by the growth bands.

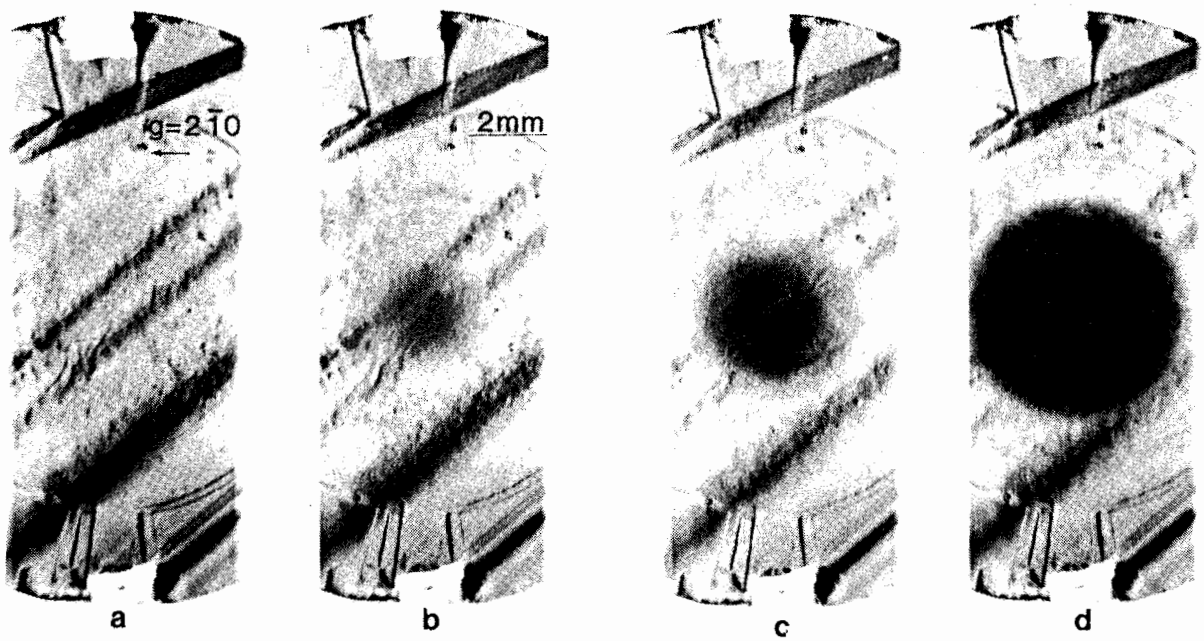


Figure 12: Interaction of the 5th overtone of thickness shear with growth bands in a plano-convex resonator.
12a: without vibration ; 12b, 12c, 12d increasing vibration amplitude.

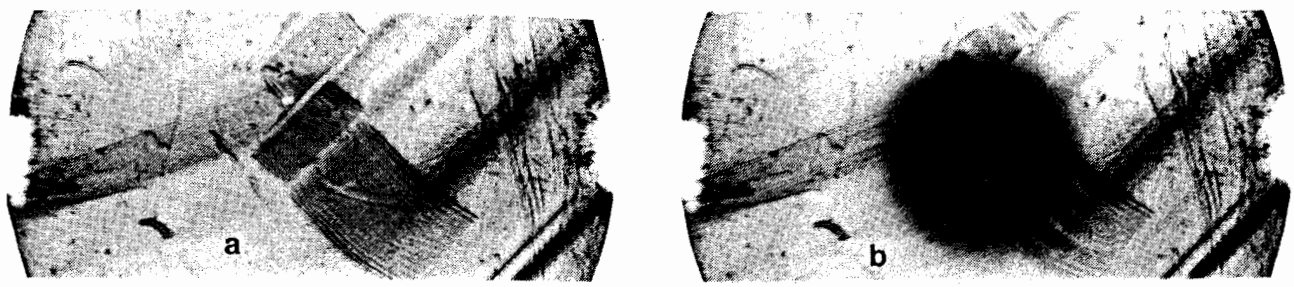


Figure 13: Interaction of the 5th overtone of thickness shear with an intricate structure of growth bands.
13a: without vibration. 13b: vibrating.

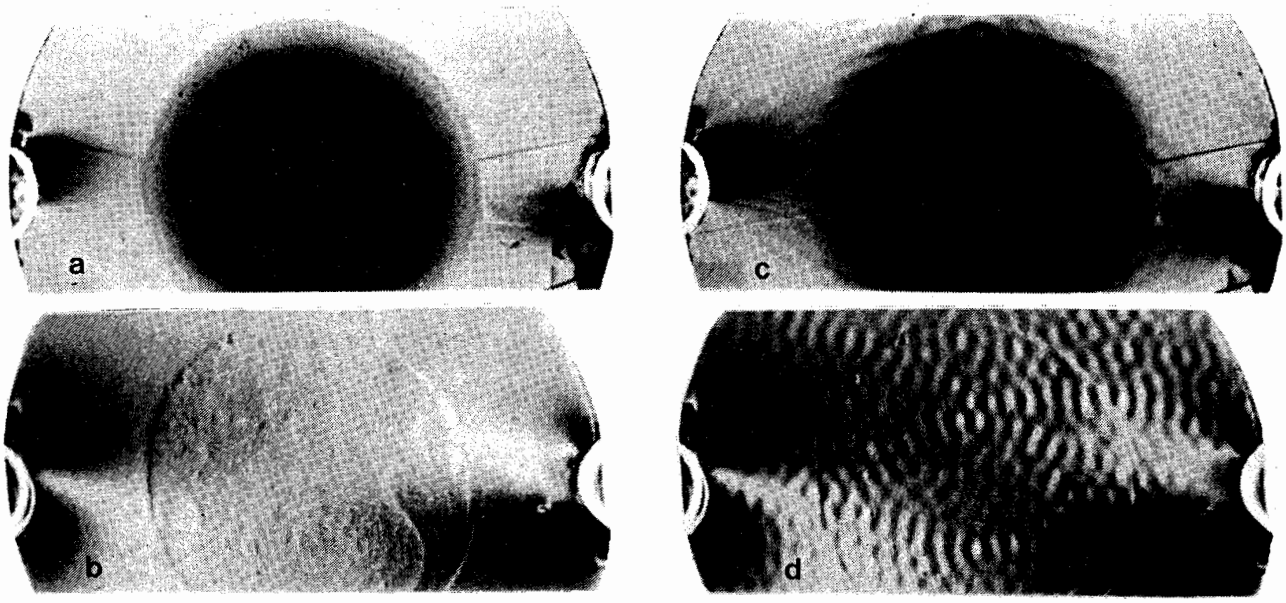


Figure 14: Coupling of the fundamental thickness shear to progressive plate modes.

14a: time integrated topograph of the u_1 component.
14b: time integrated topograph of a u_2+u_3 mix of components (mostly u_3).
14c: time resolved topograph of the u_1 component.
14d: time resolved topograph of the same mix of u_2+u_3 as in Fig.14b.

VI - COUPLED MODES

VI.1 -Trapped thickness shear coupled to plate modes with non stationary components

The non stationary coupled components cannot be observed by the classic topography because the time integration of the contrast on the film. Figures 14a and 14b are two "conventional" topographs of an AT-cut quartz plane resonator operating on the fundamental thickness shear mode, They show respectively the shear component (u_1) and a mix of the u_2 and u_3 (mostly u_3) components. These conventional (time integrated) topographs show that the displacement in this resonator has very nearly only a trapped stationary u_1 component (pure shear), with no detectable u_2 and u_3 components. With the stroboscopic topography on this resonator, operating in the same conditions, it may be observed that in reality the displacement includes new time progressive components (u_1, u_2 and mostly u_3). This non stationary components were unknown, unexpected in bulk wave energy trapping resonators, and not taken in account in the theoretical modelization (Figures 14c and 14d). Further experiments have shown their existence in most of the plane resonators vibrating on the fundamental mode. The energy carried out of the plate (towards the mounting clips and the crystal holder) is the limiting element of the Q factor for this type of resonator.

VI.2 -Intricate coupled modes in resonators with no energy trapping

In Figure 15 is given a stroboscopic topograph of an X-cut quartz resonator operating at the fundamental mode of thickness stretch which is known to be untrapped. This topograph represents a reflection of a Laue pattern corresponding to an in plane component of the displacement, and, it can be observed in this resonator that the mode extends all over the sample. The configuration of the contrast corresponds to an in plane component coupled with the thickness compression mode which extends up to the edges of the resonator. This crystal was operated at seven times the synchrotron frequency. The direction following the clips which gives the white contrast corresponds to the Y axis. Another example is given in Figures

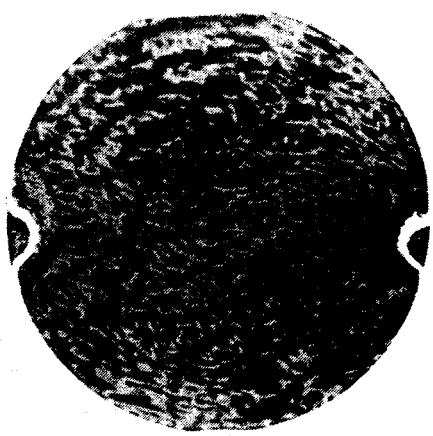


Figure 15: Topograph of an X-cut vibrating in untrapped thickness stretch.

16a and 16b which represent two topographs of a the same X-cut quartz crystal vibrating also on the fundamental mode. These two topographs correspond to two different reflections of the same Laue pattern that can image two different components of the vibration mode. It can be also observed that the mode extends up to the edge of the resonator. The existence of two in plane components of the displacement coupled to the expected component (normal to the plate) indicates that when there is no energy trapping, very intricate mode shape can be observed. The coupling of the thickness compression to other intricate components results much probably of the fact that the untrapped mode has to satisfy the boundary conditions at the edge of the plate (they contain terms permitting such coupling)

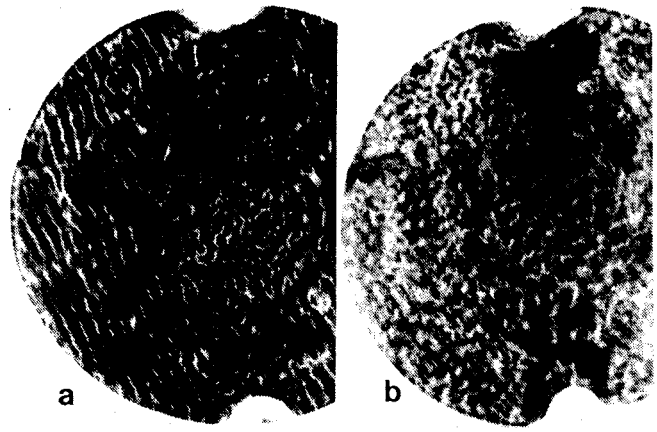


Figure 16: Two in plane components of the vibration mode of an X-cut fundamental resonator.

VII - CONCLUSION

Stroboscopic experiments performed on quartz and berlinite resonators containing dislocations or growth bands have shown interesting informations, namely on the existence of interaction between the acoustic wave and the crystal defects. This suggests that a non linear treatment is necessary to describe a more exact acoustic behaviour around the defects. This treatment will be useful to get local acoustic characteristics such as the deviation of acoustic wave propagation and global acoustic characteristics such as the energy dissipation through crystal defects. Classic X-ray topography using the synchrotron radiation has revealed the existence of strong coupled modes for certain well trapped modes. Further analysis concerning the mechanism of the mode coupling seem interesting for the basic understanding of the resonator theory and important for the practice.

Acknowledgments: The authors wish to acknowledge S.I.C.N. (France) and RADC/ESE (U.S.A.) for kindly providing several of the synthetic crystals used in this study. This work was supported by D.R.E.T. (France) through the contract n° 8734138004707501.

References

(1) BALLATO, A. ; KOSINSKI, J.A. ; LUKASZEK and MALLIKARJUN, S. Proceeding of the 5th. European Frequency and Time Forum - 1991 - Besançon - FRANCE.

(2) EERNISSE, E.P.; WARD, R.W. And WOOD, O.L. Proceeding of the 43rd Annual Symposium on Frequency Control - 1989 ; 388-395 - Denver - COLORADO (U.S.A.).

(3) CAPELLE, B. ; ZARKA, A. ; ZHENG, Y. ; DETAINT, J. And SCHWARTZEL, J. Proceeding of the 43rd Annual Symposium on Frequency Control - 1989 ; 470-476 - Denver - COLORADO (U.S.A.).

(4) ZARKA, A. ; CAPELLE, B. ; DETAINT, J. And SCHWARTZEL, J. Proceeding of the 41st Annual Symposium on Frequency Control - 1987 ; 236-240 - Philadelphia (U.S.A.).

(5) L.U.R.E. (Laboratoire pour l'Utilisation du Rayonnement Electromagnétique) - Orsay - FRANCE.

(6) TUOMI, T. ; NAUKKARINEN, K. And RABE, P. Phys. Stat. Sol - 1974 - A.25, 93.

(7) ZHENG, Y. ; ZARKA, A. ; CAPELLE, B. ; DETAINT, J. And SCHWARTZEL, J. - 1989 - Acta Cryst., A.45, 275-285.

(8) ZARKA, A. ; CAPELLE, B. ; DETAINT, J. And SCHWARTZEL, J. Proceeding of the 2nd European Frequency and Time Forum - 1988 ; 651-655 - Neuchâtel -SUISSE.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 7A Number 66

Quality Assessment of Quartz Crystal Units acc. to ISO 9000 with Implementation of Statistic Process Control

Dipl.Phys. Lutz Schulzke

KVG, R&D crystals, D-6924 Neckarbischofsheim

Abstract :

The CECC- System for the quality assessment of quartz crystal units acc. to ISO 9000 is described herein with the following documents for Quality management : ISO 9000, 9001, 9002, EN 29 000, 29 001, 29 002, Quality assessment procedures : CECC 00 114, Parts I and III and Generic, Sectional and Blank Detail Specifications for Quartz Crystal Units : CECC 68 000, 68 100, 68 101, 68 200 and 68 201.
In order to meet the requirements of the above standards and to get the capability approval as a manufacturer of quartz crystal units KVG has set up an appropriate program. Some essential elements of the computer aided quality system with the implementation of statistic process control are given in this paper.

Abbreviations :

KVG	: Kristall- Verarbeitung Neckarbischofsheim GmbH
ISO	: International Standards Organization
IEC	: International Electrotechnical Commission
EN	: European Standards
CENELEC	: European Committee for Electrotechnical Standardization
ECQAC	: Electronic Components Quality Assurance Committee
CECC	: Cenelec Electronic Components Committee
MUXHAG	: CECC Military Users Ad Hoc Group
ONH	: National Authorized Institution
ONS	: National Supervising Inspectorate
RP	: Rule and Procedure
QPL	: Qualified Products List (CECC 00 200)
CODUS	: CECC online database for specifications, manufacturers and others
WG 17	: WG Piezoelectric Devices for Frequency Control and Selection

Introduction :

The quality of a product is not given by the reliability of the final tests only but by the quality assessment from design / development, production, installation, servicing to operation. Standards for the harmonization of specifications and quality assessment procedures facilitate the international trade. In this field European organizations have achieved a leadership in the world through the documents issued for electronic components by CENELEC / ECQAC.
CENELEC is supported by actually 15 ONH's, such from

- Germany : DKE Deutsche Elektrotechnische Kommission in DIN und VDE
- France : UTE Union technique de l'Electricité
- United Kingdom : British Electrotechnical Committee
  BSI British Standards Institution

For quartz crystal components a complete system of standards is now issued. KVG as a major manufacturer with a worldwide market for its professional quartz crystal units, filters, oscillators and sensors has started a program to achieve capability approval in the CECC- System with benefits for both producer and customer (acc. to CECC 00 500) :

- assessment of stable quality level of the components
- many sources in different countries with equally qualified producers
- certification of release for each shipment
- facilitation of investigations
- less efforts in quality control for the producer and incone inspection
- increased reliability
- lower costs over life time
- simpler conditions for purchasing and marketing
- higher confidence towards manufacturer and product
- availability of components with quality approval in small quantities from approved distributors

In this paper the CECC- System is shown for quartz crystal units in some major items without any intention for completeness.
Anyhow KVG is a strong supporter of international standards both for technical details and quality assessment.
ISO 9000 was chosen for the title in this paper because it is better known in the world than the European CECC- System, which lacked of completeness in documents specially for the quality assessment of quartz crystal units. This deficit is now overcone by the big efforts of WG 17 in a manner that this program is agreed as a pilot project for IECQ, the quality system of IEC.

Standards for Quality Management :

Scope of the CECC-System is the ease of international trade with electronic components through
- publication of standardized specifications and quality assessment prodedures
- release of internationally accepted mark or certificate of conformance (COC) for components produced according to the CECC System without a further necessity of (income) tests

ISO 9000 - 2 : technically identical with EN 29 000 series

EN 29 000 : Quality Management and Quality Assurance Standards,Guidelines for Selection and Use

EN 29 001 : Quality Systems - Model for Quality Assurance in Design / Development, Production, Installation and Servicing

EN 29 002 : Quality Systems - Model for Quality Assurance in Production and Installation

CECC 00 500 Issue 3 INTRODUCTION INTO THE CECC SYSTEM
Approval :
Any organization can achieve an approval from CECC, either for capability or for qualification, when certain exacty defined and rigorous conditions are fulfilled. The approval can be given to manufacturers and distributors of electronic components certificated according to internationally accepted specifications and standards.

Product :
All components acc. to this system were subjected to a severe quality inspection and fulfill the requirements of a complete program for tests and release, surveyed by an independent inspectorate.

Advantages : given above in the introduction

Customer :
Anybody can get direct information upon organizations approved by CECC, specifications published and all components produced acc. to this system through
- publications (Qualifed Products List QPL : CECC 00 200)
- database CODUS

Flexibility :
For the requirements of the rapidly changing market two types of approvals (capability and qualification) are established, for the requirements of different quality levels the AFA-System (Assessed Process Average, acc. to CECC 00 014) and the PPM- System (Parts Per Million, acc. to CECC 00 800).

CECC 00 505 Issue 2 SPECIFICATIONS OF THE CECC- SYSTEM
See figure 1 for quartz crystal units.

CECC 00 501 Issue 1 GUIDE FOR MANUFACTURERS OF ELECTRONIC EQUIPMENT

Purchasing of Quartz Crystal Units with Qualification Approval :
Qualification approval means that the component was tested by an third independent organization according to an CECC- specification. The CECC System encourages a close cooperation between manufacturer and customer.
See figure 2 for quartz crystal units.

Purchasing of Quartz Crystal Units without Qualification Approval :
According to the new requirements either an existing specification can be amended or a new specification becomes necessary. The amended or new specification indicates the relation to an existing one, the additional requirements or levels of quality assessment and test plans. For the choice of the two possibilities the "structural similarity" given in the sectional specification CECC 68 200 should be taken into account.
See figure 3 for quartz crystal units.

Purchase of Quartz Crystal Units with Capability Approval :
An increasing number of standard / catalogue and customer designed components is produced with capability instead of qualification approval. Capability approval means that a manufacturer applies technologies with precisely defined boundaries. The approval is then valid for all components produced with this technology and is subjected to the surveillance of an independant organization.
See figure 4 for quartz crystal units.

CECC 00 502 Issue 1 GUIDE FOR MANUFACTURERS OF ELECTRONIC COMPONENTS

Advantages of the CECC- System :
- participation on a system accepted in Europe
- conformity to European standards as a condition for the European market
- qualification for use of CECC- Mark or COC
- reduction of different national, public or private quality assessment systems to the uniform CECC- System
- reduction of different specifications for practically equal components
- reduction of the multiplicity of enclosures
- uniform tests
- reduction of the variety of certifications
- addition to the CECC- QPL- List
- possibility of addition to QPL of MUAHAG
- increase in production of equal component types
- more economic production

Qualification of a manufacturer : acc. to RP 14/I, given below
Qualification approval : acc. to RP 14/II, not yet published
Capability approval : acc. to RP 14/III, given below

CECC 00 200 Issue 1 QUALIFIED PRODUCTS LIST (QPL) :

recommended suppliers for french military applications :

FAMILLE 68 000
SOUS - SOUS - FAMILLE 68 101, 68 102, 68 110 :
CEPE, COE, KVG, LPE, MATEL, OX-PIEZO, SAI, SEI, SPE/MBM
SOUS - SOUS - FAMILLE 68 111 : MICRO-CRYSTAL, STATEK
SOUS - SOUS - FAMILLE 68 112 : Q-TECH

MUAHAG Volume 6 Issue 2 PREFERRED PRODUCTS LIST Piezo Electric Devices :
CEPE, COPELEC, COE, LPE, M.B.M, OX-P, QUARTZKERAMIK, S.A.I., SCC, SQC

CECC 00 114/I Issue 1 1990 QUALITY ASSESSMENT PROCEDURES PART I
APPROVAL OF MANUFACTURERS AND OTHER ORGANIZATIONS

Requirements for the approval of manufacturers :

- generic specification : CECC 68 000
- sectional specifications : CECC 68 100 (Capability Approval)
CECC 68 200 (Qualification Approval)

- design control : KVG : yes
- Chief Inspector : KVG : Figure 5 Organigram

- application : EN 29 001 (Quality Systems - Model for Quality Assurance in Design/Development, Production, Installation and Servicing

- inspection organization : KVG : Figure 6 Quality Manual
Figure 7 Quality Organization
Figure 8 Quality System

- production and test facilities : KVG : State of the Art
add.: Independant Test Laboraty

The manufacturer's Chief Inspector :

- inspection tests :
CECC 68 100 : Figure 10, COCs
KVG : Figure 11, Quality Control, figure 8, SPC- Files

- review customer enquiries, drawings, technical documents and contracts :
KVG : Figure 8, Configuration Control with CAE for Specifications

- compliance of all piece parts and materials from suppliers :
KVG : Figure 8, Survey of Suppliers and CAQ for Incone Inspection

- records for the effectiveness of the inspection organization (2 years) :
KVG : Compilation and Analysis of Audits
Reports of Inspections every Day, Week, Month and Year
Reports of Final Tests every Month and Year

Maintenance of the approval of manufacturers :
Figure 9 : CECC 68 100, Table 2, Periodic Tests
Verification at Intervals of 2 Years

CECC 00 114/III Issue 1 1989 QUALITY ASSESSMENT PROCEDURES PART III
CAPABILITY APPROVAL OF AN ELECTRONIC COMPONENT MANUFACTURING ACTIVITY

The quality assessment by capability approval is equivalent to that by qualification approval and was introduced by market needs of

- small quantities or lots
- production in short times

Definitions :

- Capability approval :
manufacturing processes and quality control methods covering a specific component technology acc. to the generic specification

- Capability qualifying component (CQC) :
test specimen either specially produced or taken from normal production

- Rework :
rectification of processing errors, permitted by generic specification

Procedures for capability approval :

- Eligibility for capability approval :
manufacturers approval acc. to RP 14/I

manufacturing processes : CECC 68 100 : Figure 10
KVG : Figure 11
primary stage : CECC 68 100 : Surface Finishing of Blank

- Subcontracting : CECC 68 000 : none from Primary Stage

- Incorporated components :
KVG : Synthetic Quartz Bars (KVG has an own Blank Manufacture)
Bases, Cans and Mounting Springs
Conductive Cements and Electrode Materials (Gold, Silver)

Components not covered by CECC- Specifications,
Verification of Quality by the Chief Inspector through
- Figure 8, CAQ Incone Inspection
- Quality Assurance Contracts with Suppliers

- Application for capability approval :
definition of the technologies and ranges of components :
CECC 68 100 : Figure 10, necessary Technologies for the Processes
KVG : Figure 11, necessary Technologies for the Processes
AT- and similar Cut- Crystals from 1 to 300 MHz
Different Modes of Vibration and Enclosure Groups
Different Temperature, Tolerance and National Parameter Ranges
acc. to the KVG Crystal Catalogue

indication of all processes, tests and measurements :
CECC 68 100 : Figure 10, Processes, Tests : Section IV
KVG : Figure 11, Processes
Testplans by CAD acc. to Figure 8

- Descriptions of capability :
design rules and main features of construction : KVG : CAE

process flow chart : CECC 68 100 : Figure 10
KVG : Figures 11a/b

specifications for the CQCs and materials and parts used :
acc. to CECC 68 101 : Blank Detail Specification

inspection during the manufacturing process :
CECC 68 100 : Figure 10, CQCs
KVG : Figures 11a/b, Controls and figure 8, Audits

modifications : KVG : CAE with Indication of Issue, Date,
Responsible and Cause for Modification

- Demonstration and verification of capability :
program by the Chief Inspector with references to specifications of the CQCs, test and inspection requirements and / or process controls records of the test results

- Certificate for Capability Approval : VDE : Figure 12
by the responsible national authority
addition to CECC 00 200 Qualified Products List

- Maintenance of capability approval : as in RP 14/I

CECC 00 114/IV QUALITY ASSESSMENT PROCEDURES PART IV
PROCEDURE FOR ENHANCED ASSESSMENT OF QUALITY

Until publication RP 7/II shall be used.

Standards for Quality Assessment of Quartz Crystal Units :

Here are given only major items not mentioned in the above procedures.

CECC 68 000 Issue 1 1989 GENERIC SPECIFICATION : QUARTZ CRYSTAL UNITS

Related documents :
- IEC 50 : International Electrotechnical Vocabulary
- IEC 68 : Basic Environmental Testing Procedures
- IEC 122-1 : Quartz Crystal Units, Standard Values and Conditions
- IEC 122-3 : Quartz Crystal Units, Standard Outlines and Pin Connections
- IEC 302 : Standard Definitions and Methods of Measurement for Piezoelectric Vibrators up to 30 MHz
- IEC 410 : Sampling Plans and Procedures for Inspection by Attributes
- IEC 444 : Measurements of Quartz Crystal Units

Units, symbols and terminology : acc. to IEC- publications

Preferred ratings and characteristics :
- Temperature ranges for ambient operation and oven control
- Frequency tolerance
- Circuit conditions
- Levels of drive and drive level dependancy
- Climatic category
- Severities for bump, vibration and shock
- Leak rate

Marking (Standard) :
- Type designation as defined in the detail specification
- Nominal frequency
- Year and week of manufacture
- Factory identification code, manufacturer's name or trade mark
- Mark of conformity

Delayed delivery :
units held more than 2 years to be reinspected for frequency, resistance and solderability

Unchecked parameters :
may change from one to the other component

Test and measurement procedures :
- Procedures acc. to the Detail Specification
- Precision of measurement : inaccuracies shall be taken into account
- Standard conditions for testing
- Visual inspection, dimensions
- Electrical test procedures : acc. to IEC- publications
- Mechanical and environmental test procedures : acc. to IEC- publications
- Endurance test procedures : Ageing (30 days at 85 °C), extended aging

CECC 68 100 Issue 1 1989 SECTIONAL SPECIFICATION : QUARTZ CRYSTAL UNITS (CAPABILITY APPROVAL)

Information to be prescribed in detail specifications (CECC 68 101) :
- Outline drawing, dimensions and mounting
- Severities for environmental tests
- Marking, Ordering Information, Additional information
- Ordering information, additional information

Structural similarity (for extension of approvals) :
- common materials
- common methods of sealing or dimensional range for the enclosure
- common methods of design or processing

Processes :
- deposition of electrode material
- mounting methods
- frequency adjustment
- sealing of enclosure
- marking (durability)

Boundaries and limits :
- frequency range, range of motional parameters, frequency tolerance,
- temperature range
- range of drive levels
- range of unwanted responses
- ageing limits
- types of enclosures
- modes of vibration, overtone order and angle of cut
- climatic category, mechanical test severities

Quality conformance inspection : acc. to Detail Specification DS 68 xxx

ANNEX A : Example layout : abstract of description of capability for inclusion in CECC 00 200

ANNEX B : Layout of the front page of a CQC Specification for process control

ANNEX C : Layout of the front page of a CQC Specification to demonstrate a boundary or limit

CECC 68 101 Issue 1 1989 BLANK DETAIL SPECIFICATION : QUARTZ CRYSTAL UNITS (CAPABILITY APPROVAL)

Supplementary document to CECC 68 100, contains requirements for style and layout and minimum content of detail specifications, not mandatory for custom-built components

Characteristics :
- maximum resonance resistance
- maximum shunt capacitance, frequency pulling range, motional parameters

Certified test records : acc. to CECC 00 109

Requirements for quality conformance inspection

Inspection Levels (IL) and Accepted Quality Levels (AQL) : acc. to IEC 410

CECC 68 200 Issue 1 1990 GENERIC SPECIFICATION : QUARTZ CRYSTAL UNITS (QUALIFICATION APPROVAL)

ANNEX A : Test schedule for qualification approval

CECC 68 201 Issue 1 1990 BLANK DETAIL SPECIFICATION : QUARTZ CRYSTAL UNITS (QUALIFICATION APPROVAL)

Supplementary document to CECC 68 200, contains requirements for style and layout and minimum content of detail specifications, not appropriate for small batch production, where capability approval is recommended.

KVG Quality System

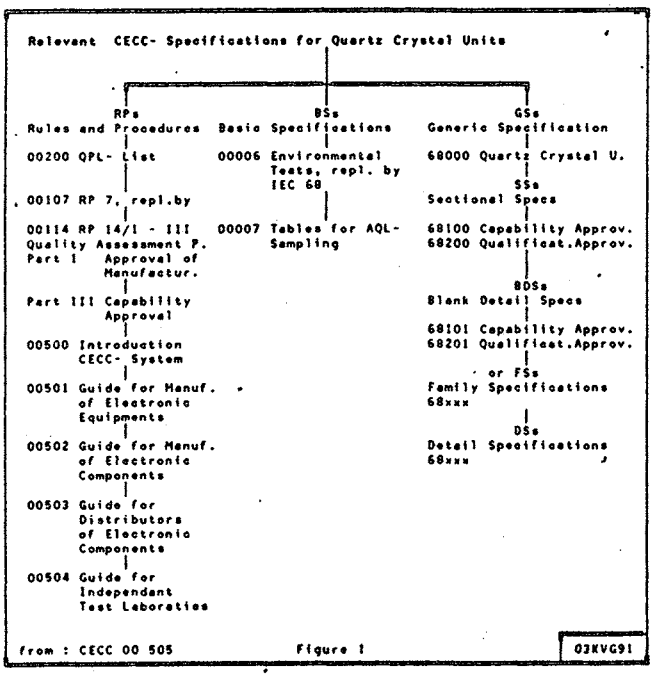
The quality system of KVG was introduced yet above in its relationship with the procedures for manufacturers and capability approval.
It is based on an ever growing ensemble of computer equipment :
- mainframe with
  30 online terminals in design, purchase, marketing to quality depts
  20 online data input stations in income, fabrication to shipment depts
  gateway to PC- network
- PC- network with server and
  24 working stations in production to quality depts
- other isolated PCs
- appropriate software for CIM, CAE, CAD, CAP, CAC, CAM and CAQ
  either bought or made in KVG

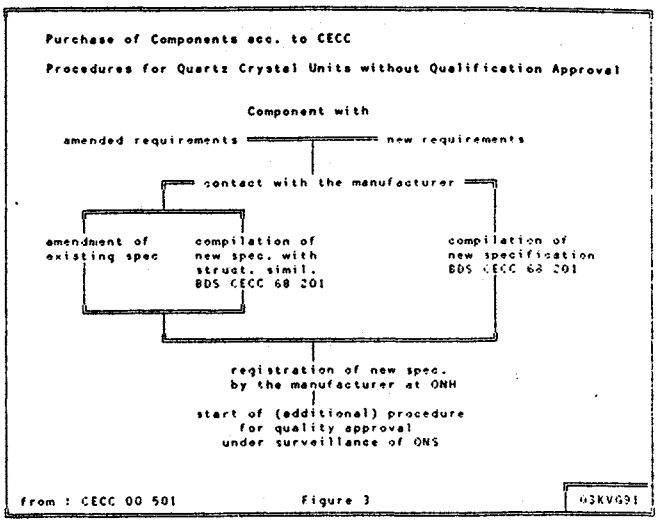
This means that more than 25 percent of the about 400 employees is guided, controlled and supported by the (dis) advantages of computers in the job, day by day, whether present or not.
For processes controlled by computer the system gives either standard (1.5) or product- (specification-) or quality-level- dependant values for the levels of inspection (IL) and acceptable quality (AQL). For critical processes and tests, such as operability, then total process control (TPC) takes the place of the statistical process control (SPC).
The newly defined SPC- database in figure 8 accumulates all the data of the above computer ensemble for each manufacturing order in KVG with a lot of activities from engineering, purchasing, income of materials, production, shipment, marketing to quality assessment.
With the quality system different loops can be closed or locked :
- real time loops for processes (CAM)
- medium and long time loops for other computer- aided activities

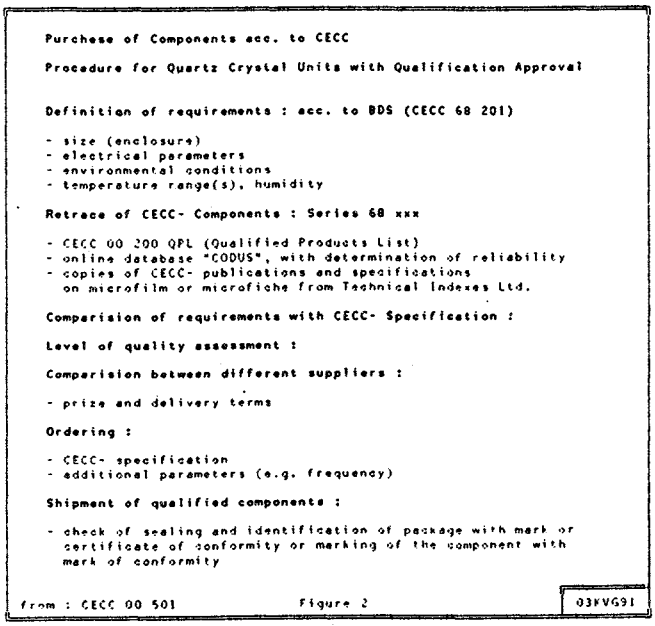
During and after completion of an order data are transferred to databases for appropriate statistics and management in order to maintain and increase the overall quality level of KVG.

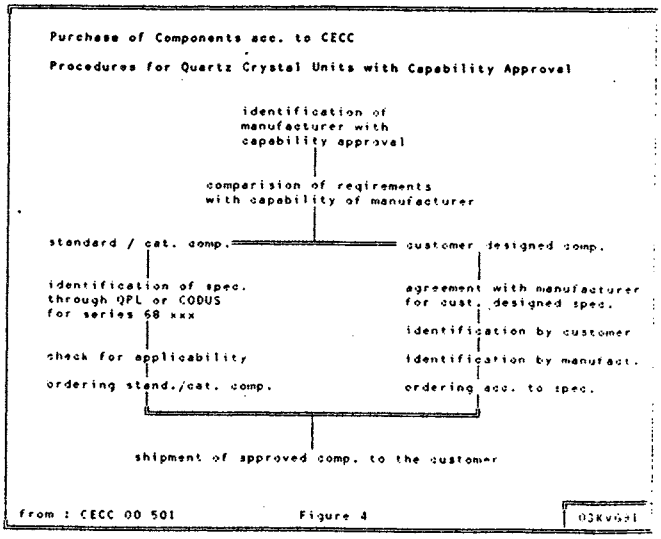
Conclusion

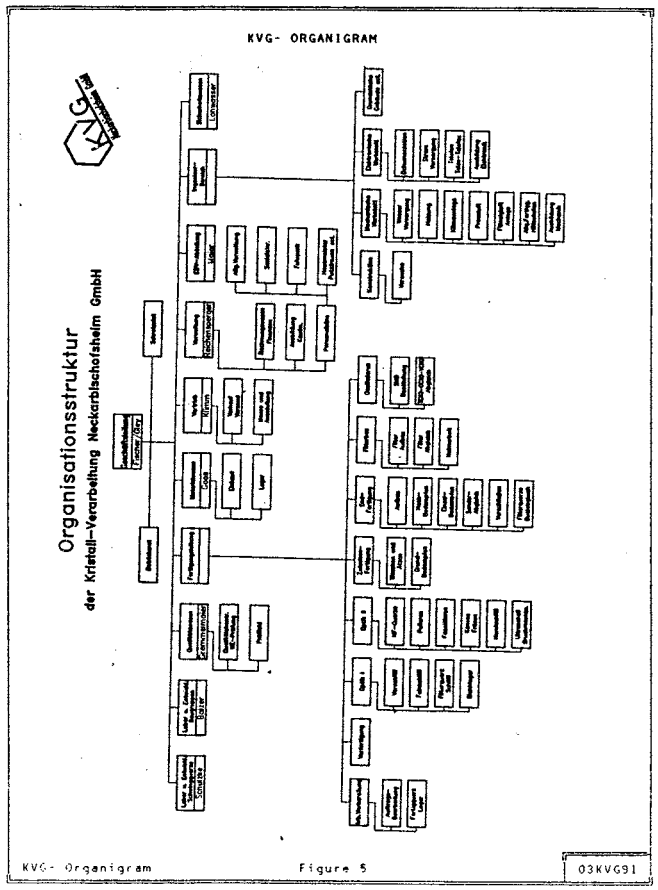
Nobody is perfect.
Perfection can be aimed to only by the good philosophy and cooperation of people, materials, equipment and - computers, where we can accumulate CECC- Procedures, data, levels and - ideas.

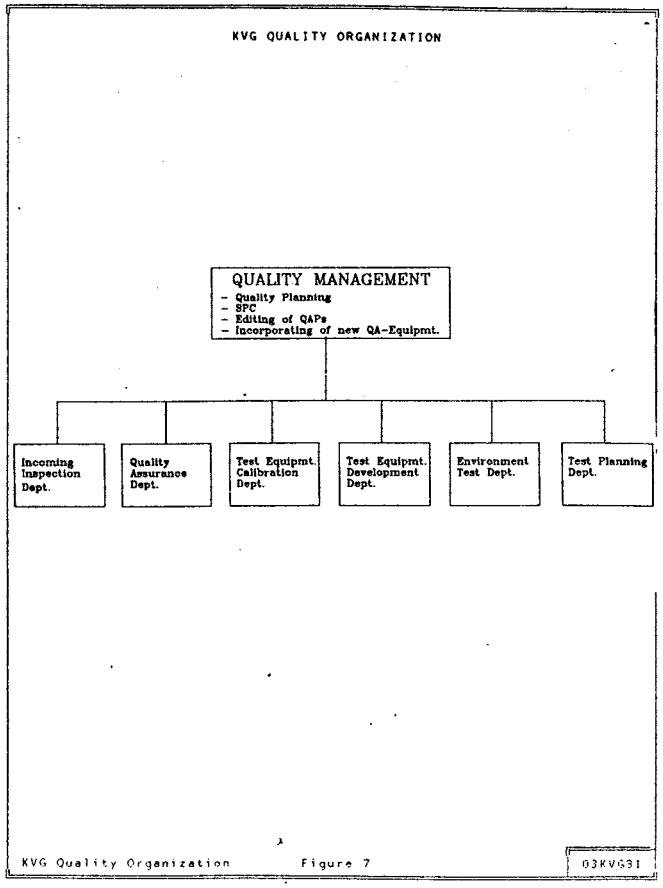

Relevant CECC- Specifications for Quartz Crystal Units
RPs
Rules and Procedures
00200 QPL- List
00107 RP 7, repl.by
00114 RP 14/1 - III Quality Assessment P. Part 1 Approval of Manufactur.
Part III Capability Approval
00500 Introduction CECC- System
00501 Guide for Manuf. of Electronic Equipments
00502 Guide for Manuf. of Electronic Components
00503 Guide for Distributors of Electronic Components
00504 Guide for Independant Test Laboraties
BSs
Basic Specifications
00006 Environmental Teats, repl. by IEC 68
00007 Tables for AQL-Sampling
GSs
Generic Specification
68000 Quartz Crystal U.
SSs
Sectional Specs
68100 Capability Approv.
68200 Qualificat.Approv.
BDSs
Blank Detail Specs
68101 Capability Approv.
68201 Qualificat.Approv.
or FSs
Family Specifications
68xxx
DSs
Detail Specifications
68xxx
from : CECC 00 505
Figure 1
03KVG91

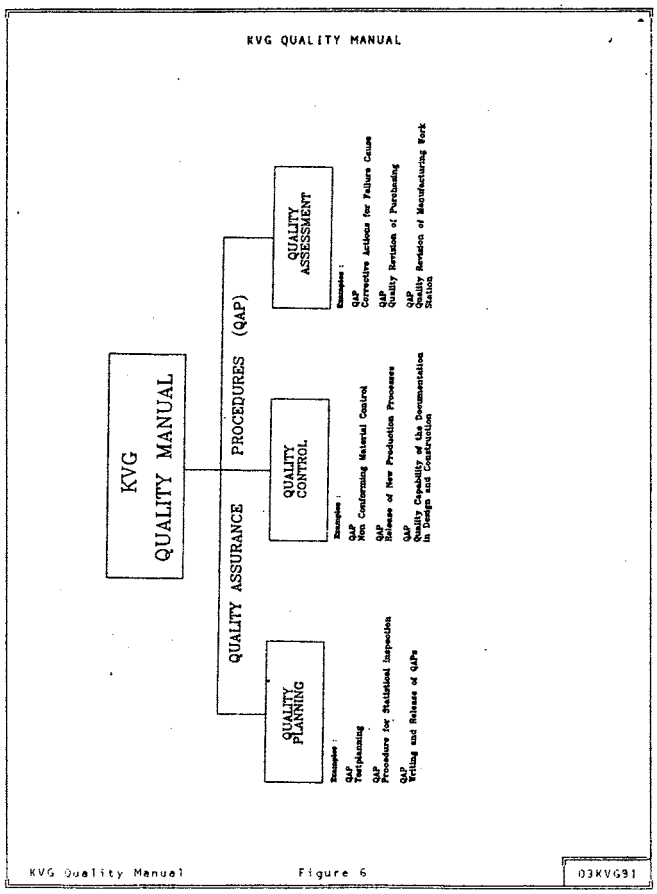

Purchase of Components acc. to CECC
Procedures for Quartz Crystal Units without Qualification Approval
Component with
amended requirements
new requirements
contact with the manufacturer
amendment of existing spec
compilation of new spec. with struct. simil. BDS CECC 68 201
compilation of new specification BDS CECC 68 201
registration of new spec. by the manufacturer at ONH
start of (additional) procedure for quality approval under surveillance of ONS
from : CECC 00 501
Figure 3
03KVG91

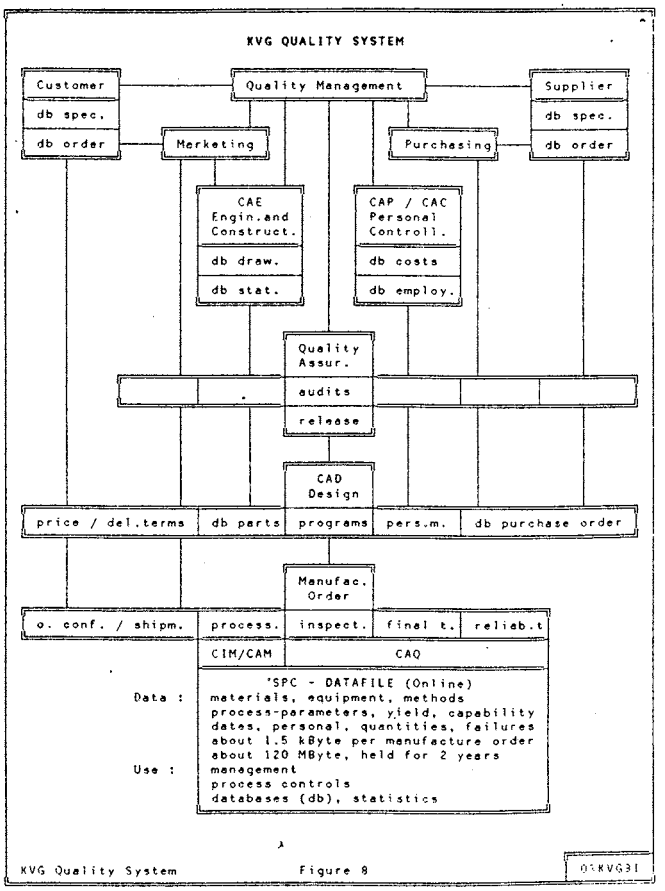

Purchase of Components acc. to CECC
Procedure for Quartz Crystal Units with Qualification Approval
Definition of requirements : acc. to BDS (CECC 68 201)
- size (enclosure)
- electrical parameters
- environmental conditions
- temperature range(s), humidity
Retrace of CECC- Components : Series 68 xxx
- CECC 00 200 QPL (Qualified Products List)
- online database "CODUS", with determination of reliability
- copies of CECC- publications and specifications on microfilm or microfiche from Technical Indexes Ltd.
Comparision of requirements with CECC- Specification :
Level of quality assessment :
Comparision between different suppliers :
- prize and delivery terms
Ordering :
- CECC- specification
- additional parameters (e.g. frequency)
Shipment of qualified components :
- check of sealing and identification of package with mark or certificate of conformity or marking of the component with mark of conformity
from : CECC 00 501
Figure 2
03KVG91


Purchase of Components acc. to CECC
Procedures for Quartz Crystal Units with Capability Approval
identification of manufacturer with capability approval
comperision of reqirements with capability of manufacturer
standard / cat. comp.
customer designed comp.
identification of spec. through QPL or CODUS for series 68 xxx
check for applicability
ordering stand./cat. comp.
agreement with manufacturer for cust. designed spec.
identification by customer
identification by manufact.
ordering acc. to spec.
shipment of approved comp. to the customer
from : CECC 00 501
Figure 4
03KVG91


KVG- ORGANIGRAM
Organisationsstruktur
der Kristall-Verarbeitung Neckarbischofsheim GmbH
KVG- Organigram
Figure 5
03KVG91


KVG QUALITY ORGANIZATION
QUALITY MANAGEMENT
- Quality Planning
- SPC
- Editing of QAPs
- Incorporating of new QA-Equipmt.
Incoming Inspection Dept.
Quality Assurance Dept.
Test Equipmt. Calibration Dept.
Test Equipmt. Development Dept.
Environment Test Dept.
Test Planning Dept.
KVG Quality Organization
Figure 7
03KVG91


KVG QUALITY MANUAL
KVG QUALITY MANUAL
QUALITY ASSURANCE PROCEDURES (QAP)
QUALITY ASSESSMENT
QUALITY CONTROL
QUALITY PLANNING
KVG Quality Manual
Figure 6
03KVG91

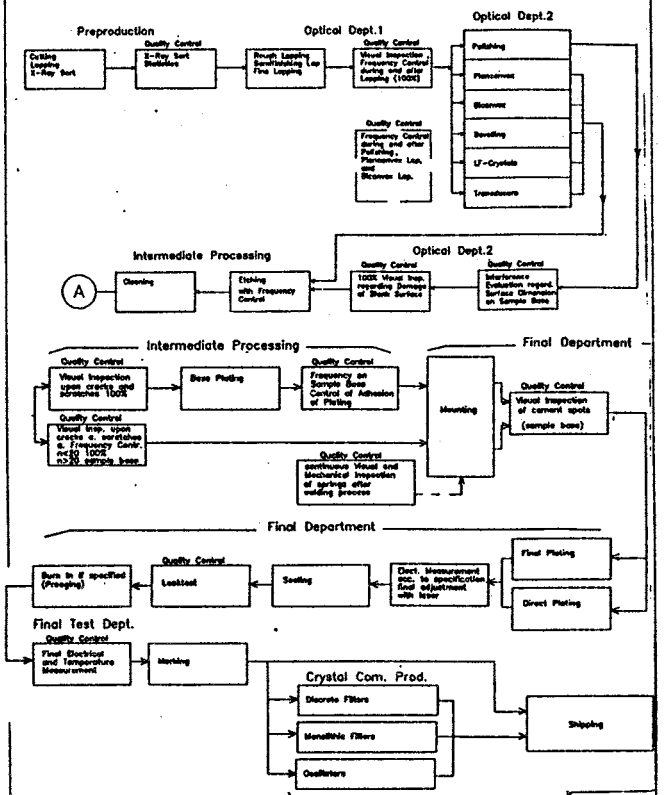

KVG QUALITY SYSTEM
Customer
db spec.
db order
Quality Management
Supplier
db spec.
db order
Marketing
Purchasing
CAE Engin.and Construct.
db draw.
db stat.
CAP / CAC Personal Controll.
db costs
db employ.
Quality Assur.
audits
release
CAD Design
price / del.terms
db parts
programs
pers.m.
db purchase order
Manufac. Order
o. conf. / shipm.
process.
inspect.
final t.
reliab.t
CIM/CAM
CAQ
*SPC - DATAFILE (Online)
Data :
materials, equipment, methods
process-parameters, yield, capability
datas, personal, quantities, failures
about 1.5 kByte per manufacture order
about 120 MByte, held for 2 years
Use :
management
process controls
databases (db), statistics
KVG Quality System
Figure 8
03KVG91

CECC 68 100 : PERIODIC TESTS FOR MAINTENANCE OF CAPABILITY APPROVAL

CQC Test Plan	Test Group	Periodicity	CQCs
Deposition of electrode material on the crystal blank	- Drive level dependancy	1 week	8
	- Electrode film thickness	4 weeks	4
	- Electrode film adhesion	4 weeks	4
Bonding	- Electrical resistance	4 weeks	8
	- Cement adhesion		
Mounting	- Resistance to solder heat, Visual test	2 years	8
	- Bump, Vibration, Shock, Visual test		
Frequency adjustment accuracy	- Frequency and resonance resistance	1 week	16
Enclosure	- Visual test, Dimensions, Insulation, Leak tests	1 week	8
	- Rapid change of temp., Visual test, Leak tests	2 years	4
	- Immersion in cleaning solvents, Solderability, Tensile and Thrust on terminations, Flexibility of wire terminations, Terminal bend test, Visual test, Leak tests		4
Frequency and temperature range	- Visual test, Dimensions, Insulation, Leak tests	2 years	12
	- Drive level dependancy, Frequency and resonance resistance at room temp. and in the temp. range		
Motional parameters and unwanted resp.	- Visual test, Dimensions, Insulation, Leak tests	2 years	8
	- Co, Cl, unwanted resp.		
Climatic performance	- Climatic sequence, Ins. resistance, Frequency and resonance resist. in temperature range, df, dR, Visual tests	2 years	8
	- Damp heat, Insulation resistance, Frequency and resonance resist. in temperature range, df, dR, Visual tests		
Durability	- Aging, Extended aging	2 years	10
	- Shelf life, df, dR, Frequency and resonance resistance at room temp. and in the temp. range		10

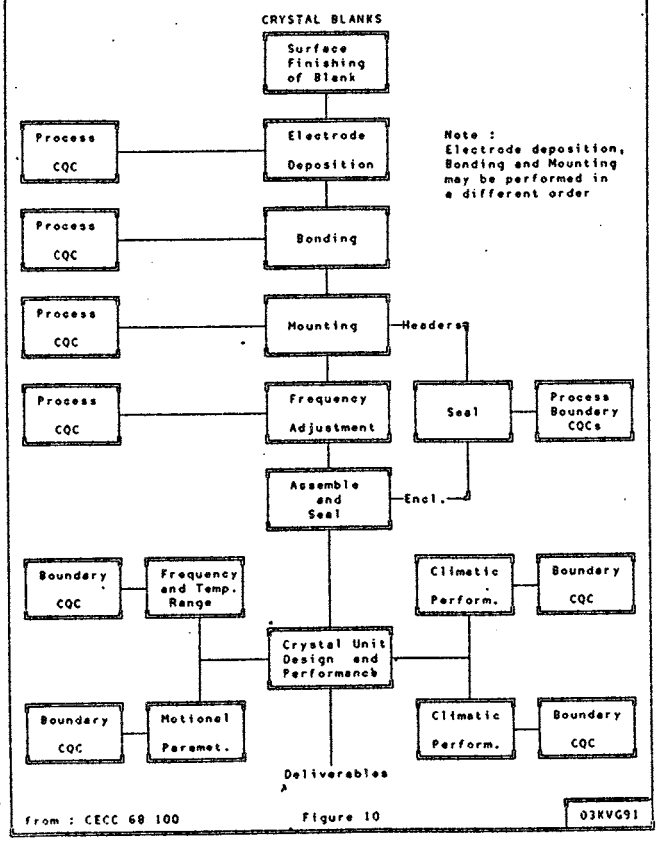
from : CECC 68 100 Figure 9 03KVG91

KVG FABRICATION AND QUALITY CONTROL PROCESSES part 1



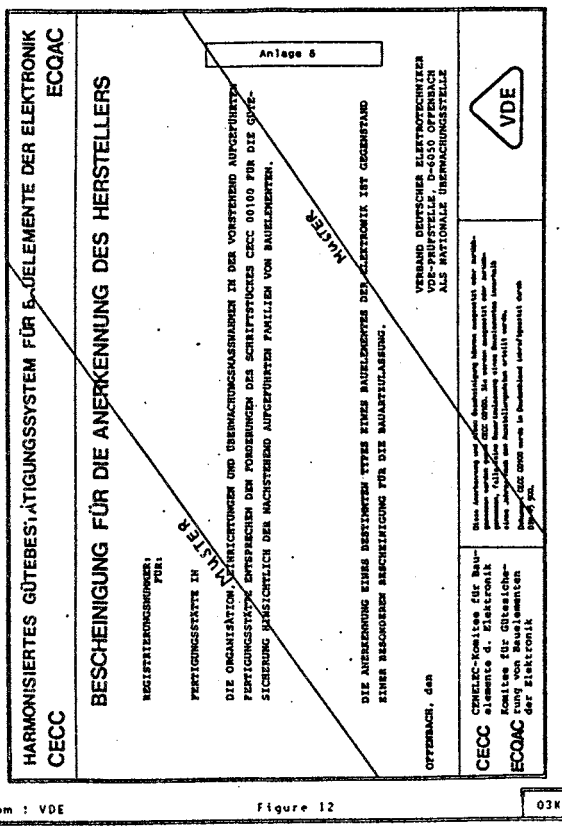
KVG Process- Flow- Chart Figure 11a 03KVG91

CECC 68 100 : PROCESSES and PLAN for the SELECTION of CQCs



from : CECC 68 100 Figure 10 03KVG91

VDE : CERTIFICATE FOR CAPABILITY APPROVAL



from : VDE Figure 12 03KVG91

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 7A Number 67

NEW POSSIBILITIES OFFERED BY SILICON STRUCTURES FOR RESONATORS AND SENSOR DESIGN

Prof. Nico F. de Rooij
Institute of Microtechnology
University of Neuchâtel
Rue A.-L. Breguet 2

CH-2000 Neuchâtel

Micromachining of single crystal silicon (bulk micromachining) as well as thin films, in particular polycrystalline silicon (surface micromachining) has resulted in a great variety of novel types of sensors and actuators. Recently a major interest has grown in developing silicon resonators for sensor applications, using single crystal silicon as well as polycrystalline silicon.

In this paper, basic aspects of bulk and surface micromachining will be reviewed.

In addition, various methods of excitation of silicon resonators will be discussed.

Finally, representative examples of a variety of sensors, for pressure, force and flow measurements, using silicon resonators will be presented.

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 7A Number 68

NONLINEAR STRUCTURE OF THE PHASE MOTION IN OSCILLATORS

Michel PLANAT

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

Abstract

Until now, our understanding of phase fluctuations in quartz crystals has been mainly phenomenological. In this paper we analyze the closed to resonance phase motion from the point of view of chaotic dynamics with the help of Düffing's equation. We use the so-called Poincaré surface of section technique to compute amplitude and phase jumps at each period of the forced oscillation and to draw the diagram of phase locked states. In the high frequency excitation range we observe an anomalous response with long periodic nonchaotic orbits and broad band spectra. This is compared with our conventional view of resonances.

Introduction

Our understanding of resonance in nonlinear systems is still fragmentary. On an experimental point of view the closed to resonance spectrum in high Q resonators exhibit a 1/f shape which has not yet been explained [1, 2]. On a theoretical point of view the low frequency, i.e. the long time behavior, in nonlinear differential systems is difficult to tackle numerically. Tradition in understanding periodic phenomena and resonances goes back to Poincaré and main fundamental results are expressed in the form of Poincaré-Birkoff and Kolmogorov-Arnold-Moser (KAM) theories [3]. Only a brief account of these approaches is given here. In a linear system resonance can take place when there is a rational frequency ratio Ω between two degrees of freedom, i.e. two frequencies ω_1 and ω_2 such that $\Omega = \omega_1/\omega_2 = p/q \in Q$. In such a case the phase space trajectory lies on a 2-torus and the section of the torus shows discrete points on a circle. The same point of view is taken in the perturbed nonlinear system by computing the phase θn ($n \in N$) at successive crossings of the section (called a Poincaré section) and defining the winding number $w = \lim_{n \to \infty} (\theta_n - \theta_0)/n$. A rational value of w means nonlinear resonance and points in parameter space giving a unique rational value belongs to an "Arnold tongue". Low frequency chaotic orbits appears where two tongues begin to overlap. KAM theory predicts the structural stability of closed to resonance, but nonresonant trajectories $\omega_1/\omega_2 \simeq p/q \in Q$ with respect to small hamiltonien conservative perturbations (slight deformation of tori). Resonant trajectories are expected to break the invariant tori according to the Poincaré-Birkhoff theorem : a set of stable islands are formed near fixed points of the unperturbed Poincaré mapping and the stable islands are bounded by a stochastic separatrix connecting hyperbolic fixed points [3,4,5].

The simplest form for the Poincaré mapping of a perturbed two dimensional torus is the so-called Arnold circle map

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi} \sin 2\pi\theta_n \quad (mod\ 1) \qquad (1)$$

It is easy to observe that for $K = 0$ we obtain $w = \Omega$ so that a slight variation of Ω leads to a new resonance and no stability should be expected. When $K \neq 0$, the $w = f(\Omega)$ curve is a devil's staircase. At $K = 1$ (Fig. 1) the devil's staircase is complete (the length of all intervalls corresponding to unlocked regions has a zero Lebesgue measure) ; when $K > 1$, locked tongues begin to overlap and this is the threshold for chaotic trajectories in phase space. Locked regions in parameter spece (Ω, K) (i.e. the dictionnary of all devil's staircases) are given in Fig. 2.

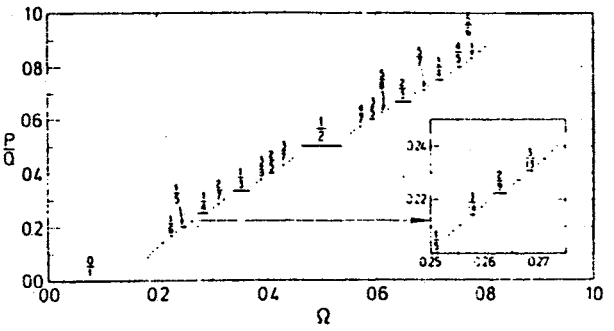


Fig. 1 : The mode-locking structure at $K = 1$ for the map (1) The devil's staircase is complete, and the complement of the mode-locked windings is of Lebesgue measure zero and universal fractal dimension D [Ref. 6]

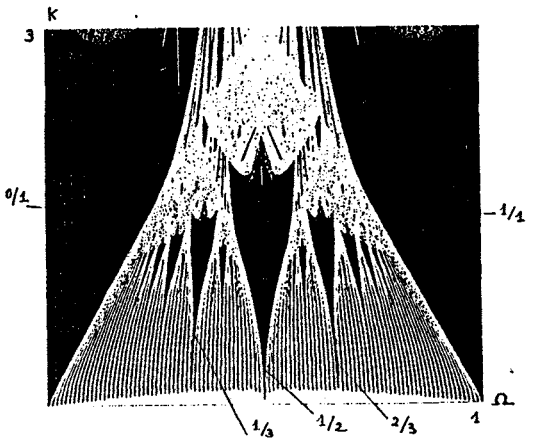


Fig. 2 : Locked regions for the circle map (1)

Eq. (1) only involves the phase θ_n. More generally it is necessary to consider the coupling between the amplitude I_n and phase θ_n on the Poincaré surface of section. The simplest generalisation of (1) is the so-called standard map

$$I_{n+1} = I_n - \frac{K}{2\pi} \sin 2\pi\theta_n$$
$$\theta_{n+1} = \theta_n + \Omega + I_{n+1} \qquad (2)$$

In the limit of zero amplitude the circle map (1) is recovered. For this mapping the dictionnary of locked regions is given on Fig. 3.

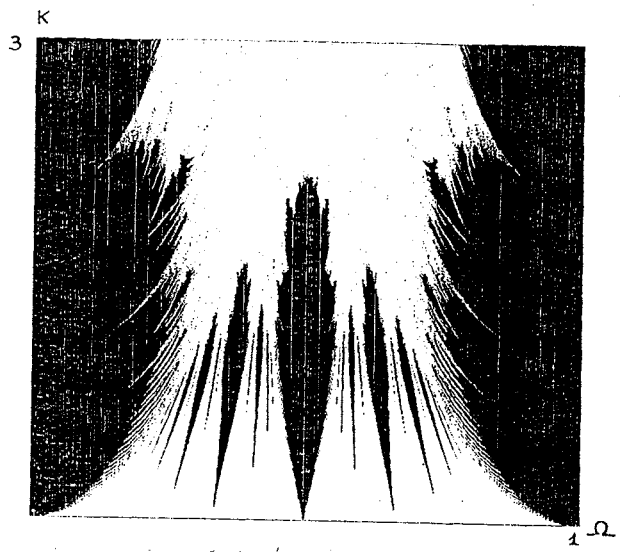


Fig. 3 : Locked regions for the standard map

Now main tongues have nonsmooth boundaries and secondary thin tongues have disappeared : the action-angle coupling has a selective stabilizing effect.

As a third example of locking we consider an analytic map : the Mandelbrot's map

$$z_{n+1} = z_n^2 + c \ ; \ z_n \in \mathbb{C} \ and \ c = p + iq \in \mathbb{C} \qquad (3)$$

The diagram of locked states is identical to the well known Mandelbrot set (Fig. 4). However for maps (1) and (2) a given tongue corresponds to a unique rational value while for map (3) all tongues are connected and the set is a fractal [7].

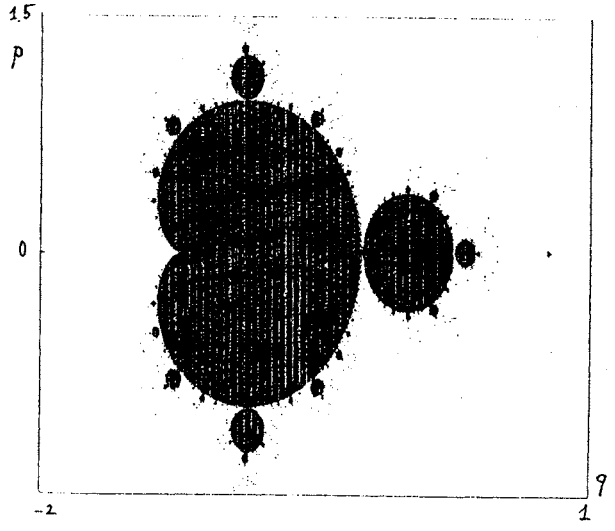


Fig. 4 : Locked regions for the Mandelbrot's map

Nonlinear resonances of Düffing's equation

We go to the concrete problem of modeling resonances in a 5 MHz AT cut quartz crystal resonator with two parallel stress free boundaries located at $x = 0$ and h. If we restrict for mechanical vibrations, the shear horizontal displacement $u(x,t)$ is given by solving

$$u_{tt} = V_T^2 u_{xx} + (3\Delta/\rho) u_x^2 u_{xx} + \eta u_{xxt}$$
$$u_x [1 + (\Delta/\rho V_T^2) u_x^2] = 0 \quad when \ x = 0 \ and \ h \qquad (4)$$

where

$$V_T = (C_{66}/\rho)^{1/2} \simeq 4 \times 10^3 \ m\,s^{-1}$$

$$\Delta = (1/2) C_{22} + C_{266} + (1/6) C_{6666} \simeq 10^{12} N/m^2$$

$$\eta = h V_T / (N Q \pi) \simeq 10^{-7} \ m^2 s^{-1}$$

the C's are elastic constants, ρ is the mass density, $Q \simeq 2 \times 10^6$ is the quality factor of the resonator and N is the number of half periods of the cavity [8].

In the linear nondissipative approximation, standing waves in the resonator take the form

$$u = a_o \cos(\omega_o t) \cos(k_o x) \qquad (5)$$

where $\omega_0 = k_0 V_T$ and $k_0 h = (n+1)\pi$, $n = 1, 2 \ldots$

Accounting for nonlinearity and dissipation, the small amplitude standing waves can be found as

$$u = a(t) \cos(k_o x) \qquad (6)$$

where $a(t)$ obeys the differential equation

$$a'' + \mu a' + \omega_o^2 a + \delta a^3 = 0 \qquad (7)$$

with

$$\delta = 3\Delta k_o^4 / 4\rho \ and \ \mu = k_o^2 \eta$$

If the resonator is excited by an alternating signal of the form $b \cos \omega t$, with $\omega = \omega_0 + \nu$, $\nu \ll \omega_0$, eq. (7) becomes the popular Düffing's equation. Its solutions are studied as :

$$a(t) = \rho(t) \cos[\omega t + \theta(t)] \qquad (8)$$

Using the multiple scale or the averaging approach, it is found that in the stationnary region, the frequency response is given as [9]

$$\nu = \frac{3}{8} \frac{\delta}{\omega_o} \rho^2 \pm \left(\frac{b^2}{4 \omega_o^2 \rho^2} - \mu^2 \right)^{1/2}$$
$$tg\,\theta = \mu / \left(\frac{3}{8} \frac{\delta \rho^2}{\omega_o} - \nu \right) \qquad (9)$$

Solution (9) explains the effect of bistability and hysteresis which is observed in high Q resonators. However it cannot account for the closed to resonance unstability. In an attempt to approach this problem we study the nonlinear resonance with the method presented in the introduction.

Düffing's equation is rewritten in normalized form as

$$a'' + \mu a' + a + \delta a^3 = a_1 \cos(\Omega \tau + \Phi_o) \qquad (10)$$

where $\tau = \omega_0 t$ is the new time scale ; $\Omega = \omega/\omega_0$; $\mu = \mu/\omega_0$; $\delta = \delta/\omega_0^2$; $a_1 = b/\omega_0^2$ and we have introduced the initial phase Φ_0 of the excitation signal.

Eq. (10) can be rewritten in the form of a first order differential system. Introducing polar coordinates $a = r \cos \Phi$, $a' = r \sin \Phi$ we obtain

$$r' = -\mu r \sin^2\phi - \delta r^3 \sin\phi \cos^3\phi + a_1 \sin\phi \cos(\Omega \tau + \Phi_o) \tag{11}$$

$$\delta' = -1 - \mu \sin\phi \cos\phi - \delta r^2 \cos^4\phi + \frac{a_1}{r} \cos\phi \cos(\Omega \tau + \Phi_o)$$

To model the closed to resonances behavior we assume that during one period of the excitation signal the phase ϕ follows the forcing frequency ω and that amplitude remains constant, i.e.

$$\phi = \Omega \tau \ , \ r = r_o \quad if \quad 0 < \tau < 2\pi/\Omega \tag{12}$$

Integrating system (11) from $\tau = 0$ to $\tau = 2\pi/\Omega$ we obtain amplitude and phase jumps as

$$\Delta r = -\pi \Omega^{-1} (\mu r_o + a_1 \sin \Phi_o) \tag{13}$$

$$\Delta\phi = -2\pi \Omega^{-1} \left(1 + \frac{3}{8} \delta r_o^2 - \frac{1}{2} \frac{a_1}{r_o} \cos \Phi_o\right)$$

Generalizing the jump procedure from the nth to the $n+1^{th}$ step we obtain the Düffing's equation mapping as follows

$$s_{n+1} = s_n |1 - \varepsilon(\Omega)| + K(\Omega) \sin(2\pi\theta_n) \tag{14}$$

$$\theta_{n+1} = \theta_n + \Omega^{-1} |1 + s_{n+1}^2| - \frac{K(\Omega)}{2\pi s_{n+1}} \cos(2\pi\theta_n) \ (mod\ 1)$$

where we have taken $\phi = -2\pi\theta$; $s = K|r$; $K_1 = 1/2\ (3\delta/2)^{1/2}$ $\varepsilon(\Omega) = 2\pi^2 / (Q\Omega)$ and $K(\Omega) = q\,\Omega^{-1}$ and $q = \pi a_1 K_1$.

From now we restrict for the case of high Q resonators, i.e. $\varepsilon(\Omega) \simeq 0$ so that the mapping (14) is conservative. At high amplitudes map (14) becomes a generalized form of standard mapping (2) as follows :

$$s_{n+1} = s_n + q\,\Omega^{-1} \sin(2\pi\theta_n) \tag{15}$$

$$\theta_{n+1} = \theta_n + \Omega^{-1} |1 + s_{n+1}^2|$$

where the amplitude term s_{n+1} is replaced by its square s_{n+1}^2.

It is responsable for the amplitude-frequency effect which is observed in high Q resonators. Moreover the nonlinear parameter depends on the excitation frequency.

Map (14) is a generalized form of Fermi mapping which models the dynamics of a ball between a fixed and an oscillating wall [4]. Period 1 fixed points (θ_0, s_0) are given by

$$2\pi\theta_o = \begin{matrix} 0 \\ \pi \end{matrix} \tag{16}$$

$$\Omega^{-1} |1 + s_o^2 \pm a_1 K_1 / (2 s_o)| = m \in \mathbb{N}$$

Linear stability is studied by linearizing the map at the fixed points to obtain

$$\Delta s_{n+1} = \Delta s_n \pm 2\pi K(\Omega)\, \Delta\theta_n \tag{17}$$

$$\Delta\theta_{n+1} = \Delta\theta_n + \left(2\Omega^{-1} s_o \pm \frac{K(\Omega)}{2\pi s_o^2}\right)\left(\Delta s_n \pm 2\pi K(\Omega)\, \Delta\theta_n\right)$$

The associated jacobien matrix A has a unit determinant (conservative mapping) and its trace is

$$Tr A = 2 \pm 2\pi K(\Omega) \left|2\Omega^{-1} s_o \pm \frac{K(\Omega)}{2\pi s_o^2}\right| \tag{18}$$

The point $\theta = 0$ corresponding to the normal use of the resonator is linearly unstable since $\mathrm{tr}\, A > 2$.

The point $\theta = \pi$ is stable when $|\mathrm{Tr}\, A| < 2$.

To study the nonlinear stability of fixed point $\theta = 0$, we linearize the amplitude around the fixed point and assume that the excitation frequency is such that $\Omega^{-1} = m \in \mathbb{N}$. From (16) this gives

$$s_o = \left(\frac{1}{2} a_1 K_1\right)^{1/3} \ , \ s_n = s_o + t_n \tag{19}$$

Nonlinear stability around the fixed point $\theta = 0$ has now to be studied from

$$t_{n+1} = t_n + 2\pi s_o^3 \Omega^{-1} \sin(2\pi\theta_n) \tag{20}$$

$$\theta_{n+1} = \theta_n + \Omega^{-1} |1 + t_{n+1}^2 + t_{n+1} \cos(2\pi\theta_n)(3 s_o - t_{n+1})|$$

The diagram for locked states in parameter space (Ω^{-1}, s_0) is given on Fig. 5.

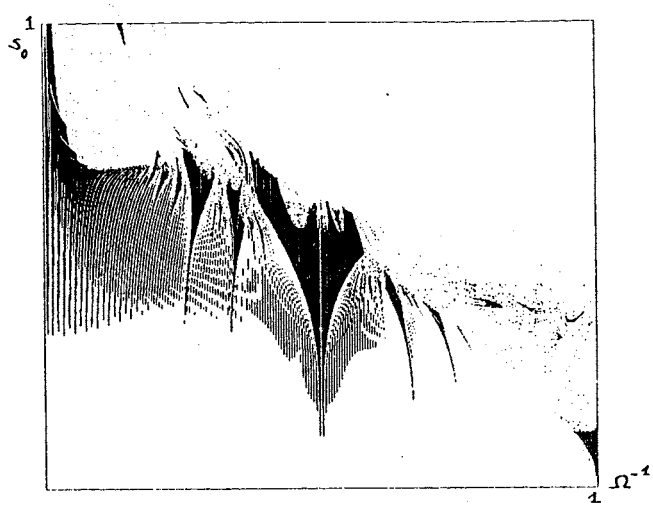


Fig. 5 : Locked regions for Düffing's eq. mapping near the fixed point of zero phase

The diagram clearly shows similarities with Figs. 2 and 3 at the main tongues 1/1, 1/2, 1/3 ... The first difference lies in the fact that the main tongue 1/1 is at low amplitude and that harmonic tongues are located at higher amplitudes in accordance with perturbation calculations. This is due to the frequency dependence of nonlinear terms in map (20). The second difference is seen in the high frequency low winding number excitation range : instead of the 0/1 tongue a double set of thin tongues are observed for $0.35 \leq s_0 \leq 0.70$ together with a high amplitude tail.

To study this new behavior we plotted the winding number in Fig. 6 for different values of the nonlinear parameters. At low values of s_0 no steps are seen ($s_0 = 0{,}4$) ; with increasing s_0 ($s_0 = 0.67$) we observe slippery steps in the high frequency range and some flat steps characteristic of a devil's staircase. The slippery staircase is time dependent : they are more and more steps as discrete time n goes on. this new behavior is not yet understood. The corresponding phase space trajectory is shown on Fig. 7. Instead of a slightly deformed circle corresponding to the section of a two dimensional torus we get a double closed curve characteristic of a four dimensional phase space (Fig. 7a). With decreasing excitation frequency the double torus breaks into a long periodic trajectory (Fig. 7b) which is the precursor of a non strange attractor (Fig. 7c).

One way to characterize the underlying structure of these attractors is through Fourier spectra of the radius t : they are many lines in the spectra leading to broadband noise in the lower frequency part. Another way to study the appearance of long period and chaotic trajectories is through one-dimensional bifurcation diagrams. Since chaos develops through double tori in Düffing's equation mapping it is tempting to draw a bifurcation diagram for the radius t versus the nonlinear parameter s_0. Such a diagram is given in Fig. 8 for $\Omega^{-1} = 0.26$ (doubling is clearly seen near $s_0 = 0.6$) and $\Omega^{-1} = 0.50$.

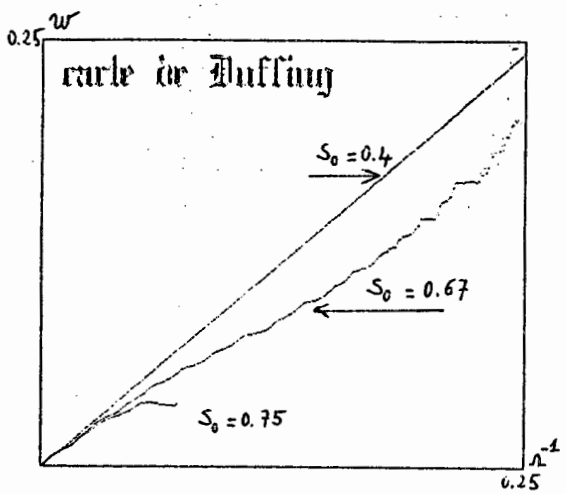
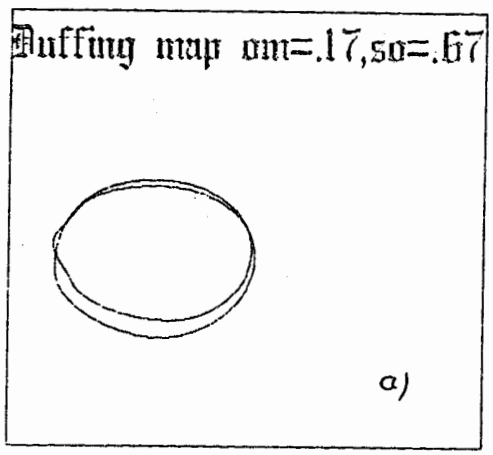
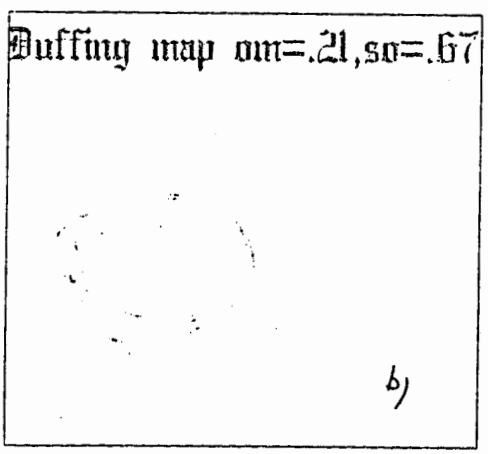


Fig. 6 : Stair case of locked states in Düffing's eq. mapping





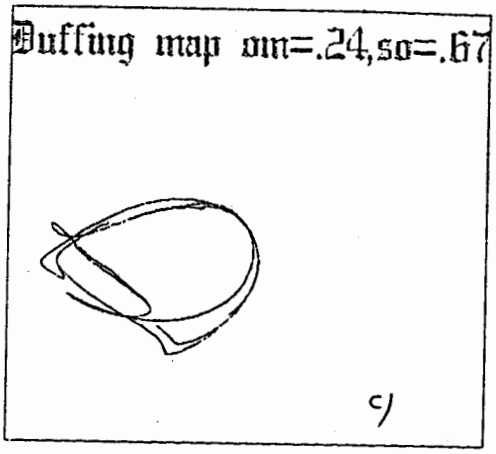


Fig. 7 : Poincaré section of Düffing's eq. mapping at low winding numbers

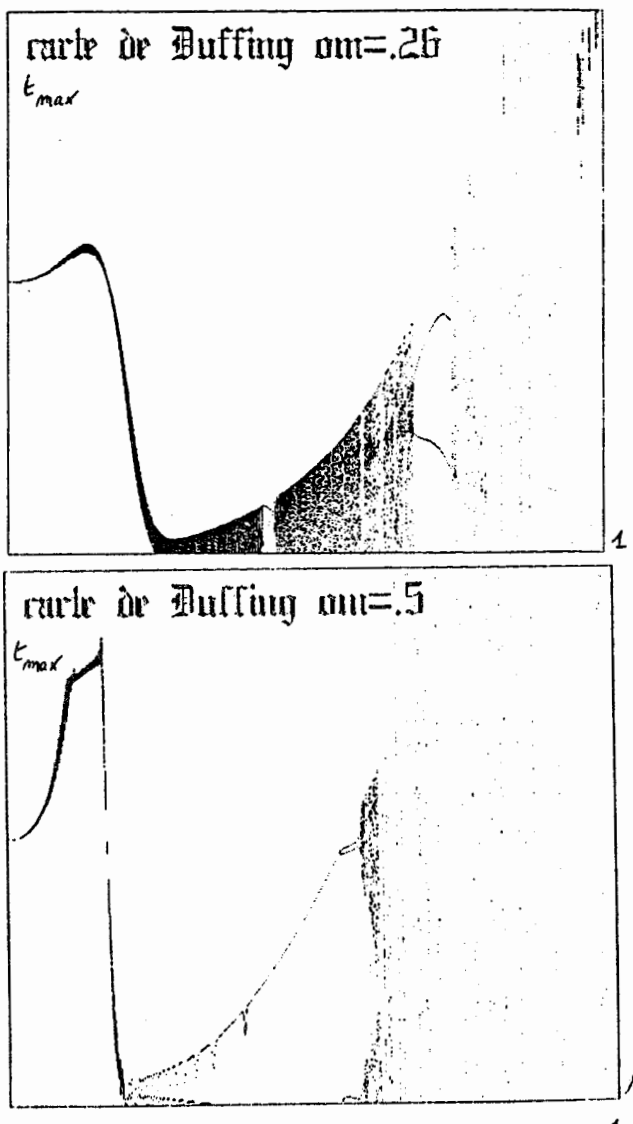


Fig. 8 : Bifurcation diagram for the radius of the closed curve

The high amplitude limit

In the high amplitude limit the map has the simpler form (15). The diagram for locked states in parameter space (Ω^{-1}, q) is given in Fig. 9.

Let us focus on the behavior at low winding numbers and high amplitude. A slippery staircase is recovered as shown on Fig. 10 for several values of the nonlinear parameter. The phase space trajectory is given in Fig. 11 : a drastic increase of dimensionnality is observed : in the fully developed chaos the phase space is built with many stochastic trajectories spread around closed curves at discrete intervals. We looked at Fourier spectra for the radius and found low frequency broad band noise with approximately $1/F^{1.5}$ dependence.

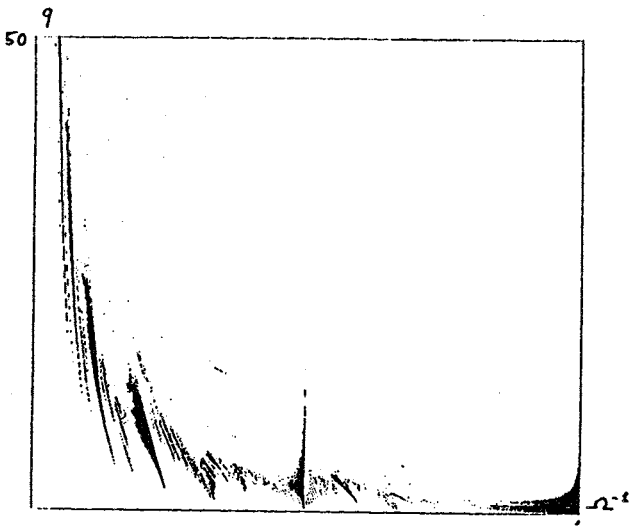


Fig. 9 : Locked regions for the high amplitude form of Düffing's equation

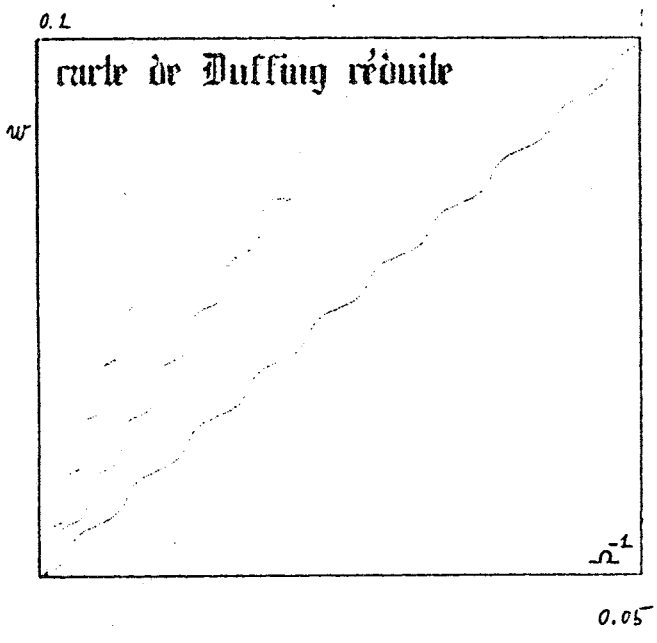


Fig. 10 : Staircase of locked states for the high amplitude limit of Düffing's equation

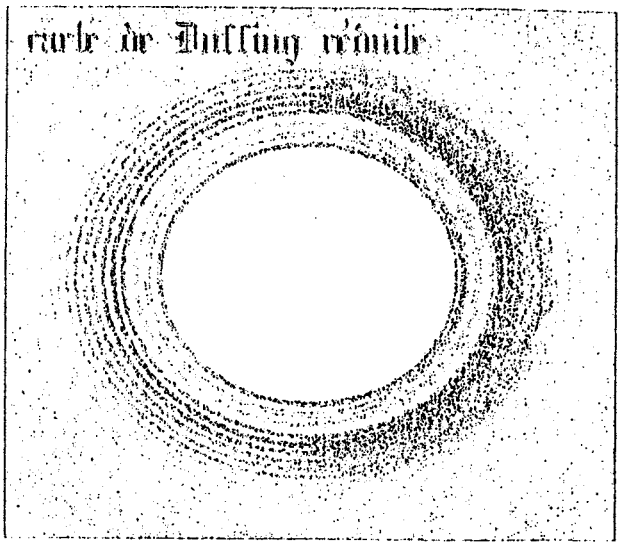


Fig. 11 : Poincaré section for the high amplitude limit of Düffing's equation

Conclusion

Starting with Poincaré-Birkhoff and KAM view of resonance we found an analytic expression for the section of trajectories in nonlinear differential equations. By studying Düffing's equation we found several new features in the long time (low winding number) limit, which were not expected from the conventional resonance theory. Instead of structural stability and locking with devil's staircases, the low winding number limit shows slippery time dependent staircases : they are associated to double tori or stochastic trajectories around tori located at discrete intervals.

Since the model can be immediately applied to high Q quartz resonators we expect significant advance can be found soon in the study of low frequency instabilities.

References

[1] T.E. Parker, Appl. Phys. Lett. 46, 246 (1985).

[2] M. Planat, J.J. Gagnepain, Appl. Phys. Lett. 50, 510 (1987).

[3] V.I. Arnold, "Mathematical methods of classical mechanics", Springer Verlag, N.Y. (1978).

[4] A.J. Lichtenberg, M.A. Liebermann, "Regular and stochastic motion", Springer Verlag, N.Y. (1983).

[5] J. Guckenheimer and P. Holmes, "Nonlinear oscillations, dynamical systems and bifurcation of vector fields", Springer Verlag, N.Y. (1983).

[6] P. Cvitanovic, M.H. Jensen, L.P. Kadanoff and I. Procaccia, Phys. Rev. Lett. 55, 343 (1985).

[7] H.O. Peitgen and P.H. Richter, "The beauty of fractals", Springer Verlag, N.Y. (1986).

[8] H.F. Tiersten, J. Acoust. Soc. Am. 59, 866 (1976).

[9] A.H. Nayfeh, D.T. Mook, "Nonlinear oscillations", John Wiley & Sons, N.Y. (1979).

SESSION 8A

OSCILLATORS II

Chairman : M. OLIVIER

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 8A Number 69

NONLINEAR MODELLING TECHNIQUES FOR QUARTZ CRYSTAL OSCILLATORS

R. Brendel*, G. Marianneau*, E. Robert**

*Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France
** CNES, DRT/TIT/TF, 18 avenue E. Belin, 31055 Toulouse Cedex

Introduction

The oscillator performances are continuously improved, at the same time, their applications in a number of fields need better and better stability. As a consequence the methods used in developing new devices have drastically changed. The method used in the past consisted in choosing a "well-suited" design for the expected goal and improving it by using a set of empirical rules often dictated by the designer know-how. The development of computer sciences in the engineering fields gives rise to new methodologies : today the Computer Aided Design is almost universally used. In electronic engineering, it exists a lot of softwares which can be used to help in the study of new electronic devices, these programs give the designer a complete view on the behavior of the future circuit, they allow to save time and money and leads to a much more efficient design. Most of these programs are intended to the development of digital electronics, a small number of them can be used for analog electronics and a still smaller number for oscillator design.

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.

Principle of an oscillator

In order to overcome the previously described difficulties, specific tools have been developed to simulate oscillators in general and especially quartz crystal oscillators.

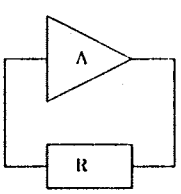


Fig. 1

An oscillator (Fig. 1) is conceptually composed of an amplifier A and a resonant feedback loop R 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 : digital circuit, bipolar of field effect transistor and its polarisation 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.

Specialized softwares

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 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].

Applications of simulators

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.

The simple Colpitts oscillator

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. 2. It will be used throughout this paper to present the various methods used.

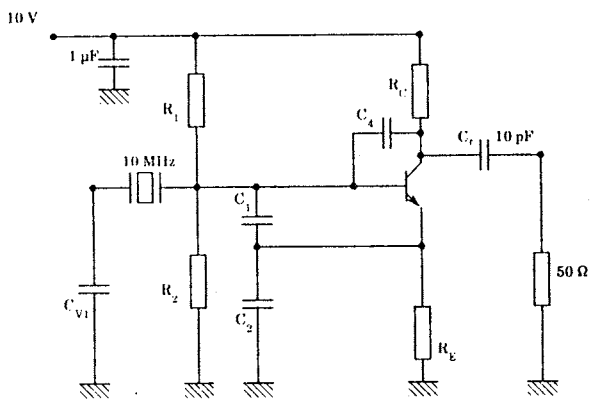


Fig. 2

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 polarisation 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 3 represents the equivalent network of a bipolar transistor using y parameters, the four complex admittances y_n can be put on the form

$$y_n = g_n + pC_n \qquad (n = i, f, r \text{ or } o) \tag{1}$$

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.

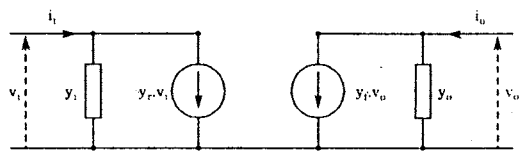


Fig. 3

There is 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

$$|y_i| = \left|\frac{i_i}{v_i}\right|_{v_o=0} = Y_i$$

$$\angle y_i = (\angle i_i - \angle v_i)_{v_o=0} = \Phi_i \tag{3}$$

So the conductance g_i and capacitance C_i defined by Eq. 1 can be derived

$$g_i = Y_i \cos\phi_i$$
$$C_i = \frac{Y_i \sin\phi_i}{\omega} \quad (4)$$

The six remaining parameters are obtained in a similar way.

Circuits and method used for the determination of 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_f and y_o. In both cases, the dc biasing circuit is the same as in the oscillator. In the first case (Fig. 4) 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 (Fig. 5) 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 $v_i = 0$.

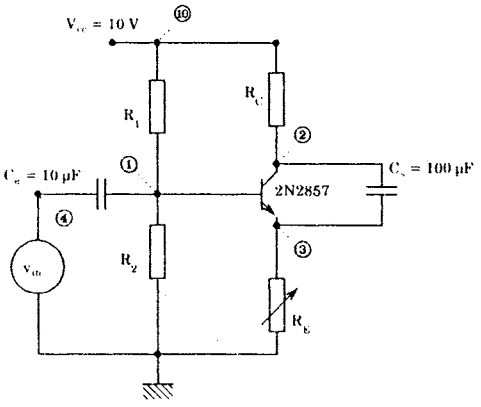


Fig. 4

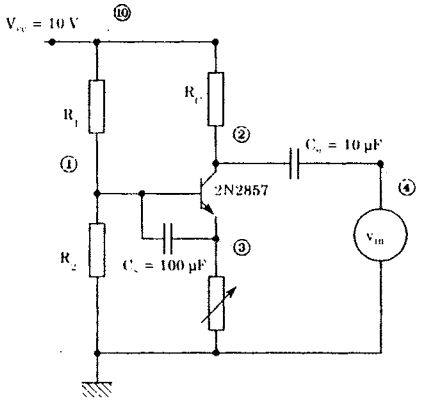


Fig. 5

Simulations have been achieved by using input signals of increasing amplitude. When large amplitudes are used, the signals in the circuit are no longer sinusoïdal (Fig. 6) 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 serie from which the parts g_n and C_n of the different parameters can be derived through Eqs. (3) and (4).

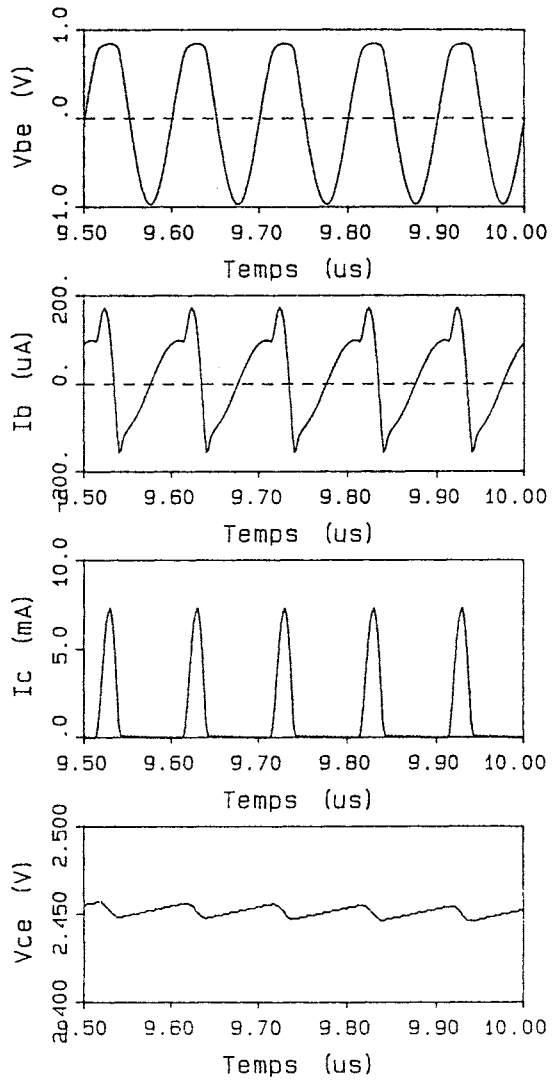


Fig. 6

Different simulations have been carried out in various biasing and temperature conditions. Fig. 7, 8, 9 and 10 shows the results obtained.

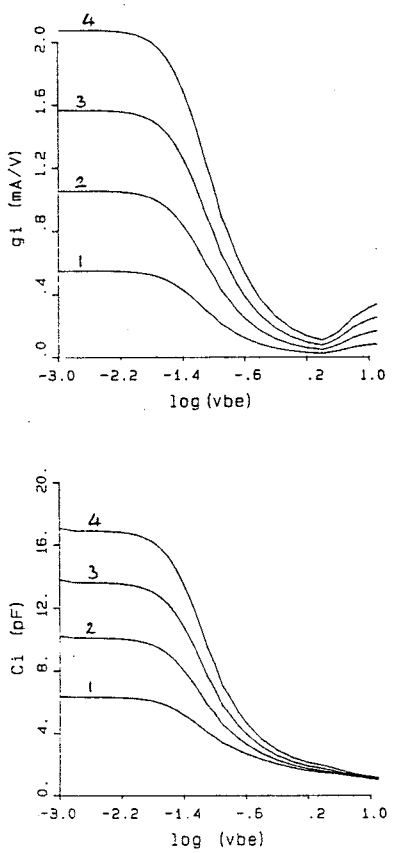


Fig. 7 Parameter y_i. V_{CE} = 3.3 V

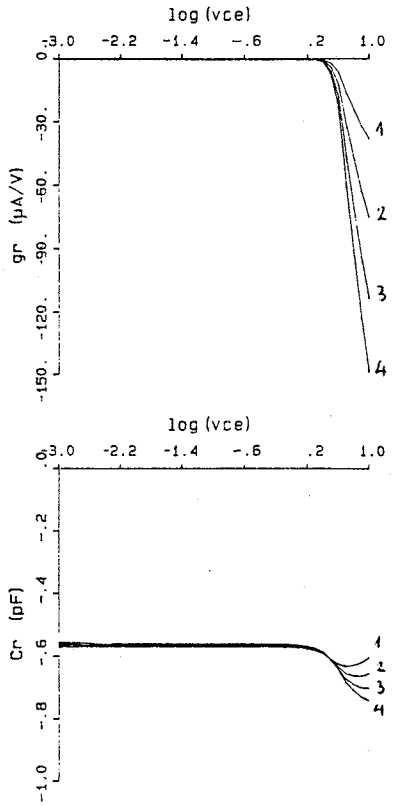


Fig. 9 Parameter y_r. Same conditions as for Fig. 7

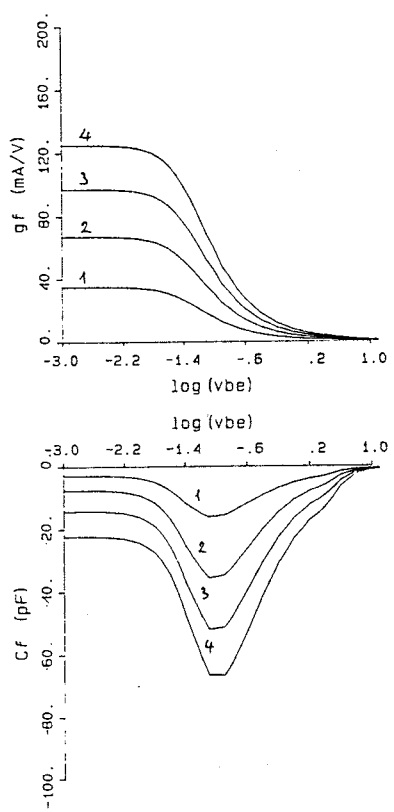


Fig. 8 Parameter y_f. Same conditions as for Fig. 7

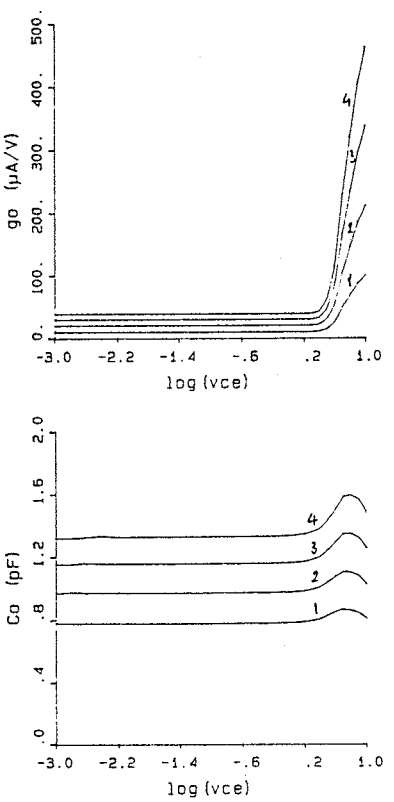
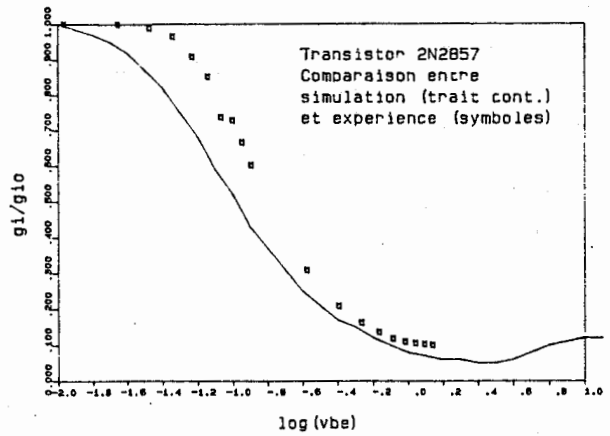
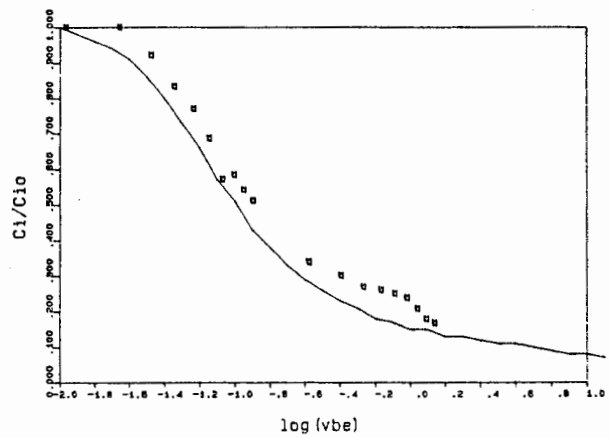
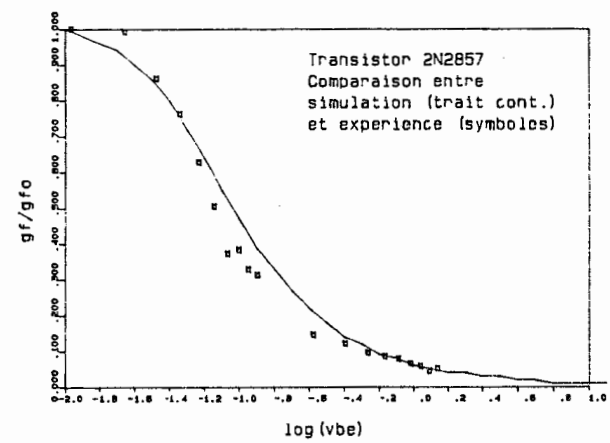


Fig. 10 Parameter y_o. Same conditions as for Fig. 7







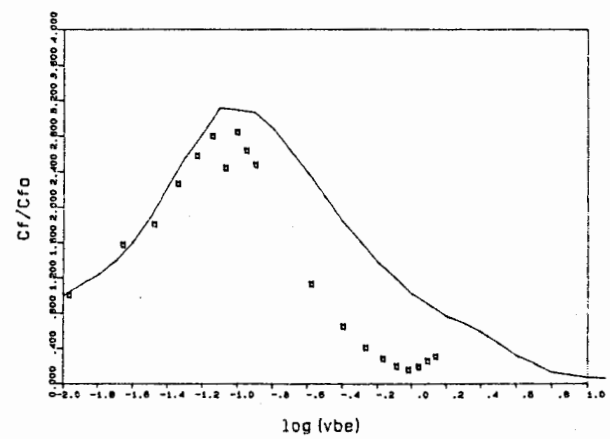


Fig. 11

Measurement of the large signal y parameters

The experimental verification of the simulation results has been carried out by using a circuit like the one shown on Fig. 4. 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. 11 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. 12 shows the equivalent circuit for the Colpitts oscillator described on Fig. 2.

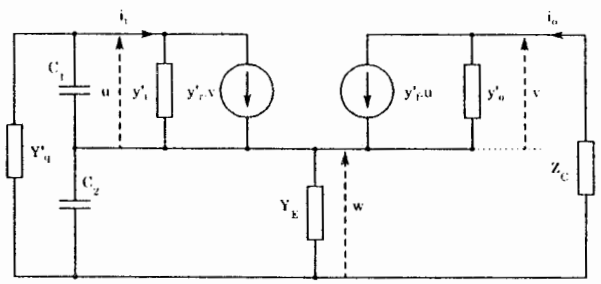


Fig. 12

The admittance Y'_q is formed by the crystal, the matching capacitor C_{v1} and the bias resistors R_1 and R_2 (Fig. 13).

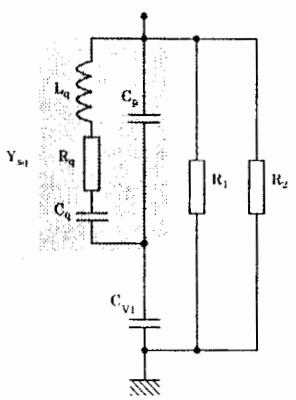


Fig. 13

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 (Fig. 14).

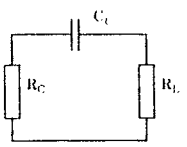


Fig. 14

$$Z_c = \frac{R_c(1 + p R_L C_\ell)}{1 + p C_\ell (R_L + R_c)} \tag{6}$$

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. 12. 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. 15 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. 15

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:

$$\sum_{k=1}^{N1} a_{2k}\,\omega^{2k} = 0$$
$$\sum_{k=1}^{N2} \beta_{2k+1}\,\omega^{2k+1} = 0 \qquad (11)$$

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. 7-10). 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. 16 shows the flow chart of the program used.

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.

Algorithm

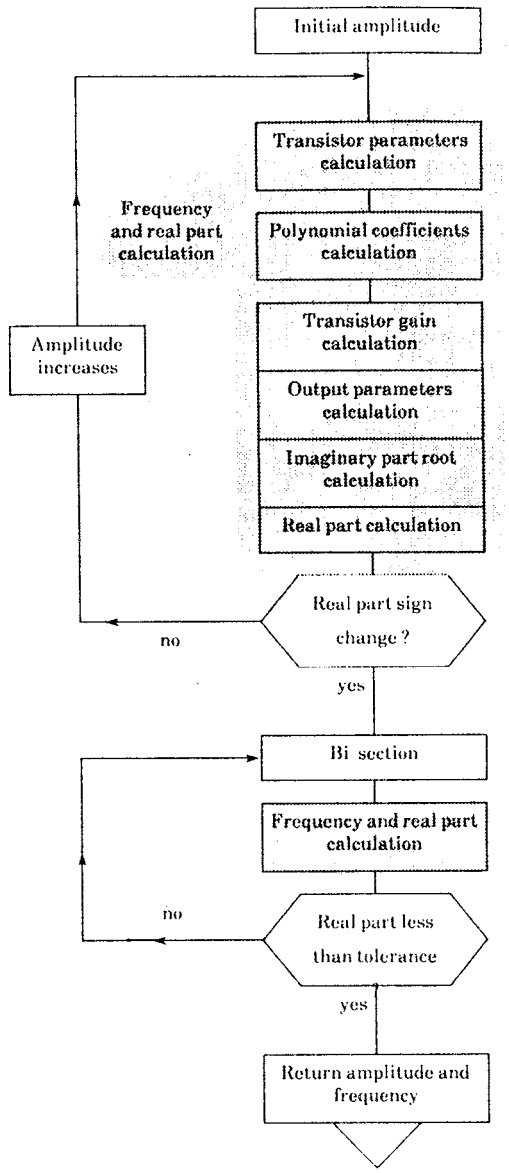


Fig. 16

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).

Session 8A Number 70

Crystal controlled oscillators in the UHF-range by PLL-technique

Adler, Bernd and Neubig, Bernd W.

TELE QUARZ GROUP, D-6924 Neckarbischofsheim/Germany

1. Introduction

Modern communication requires highly stable frequency sources at elevated frequencies up to 1 GZ and above with high spectral purity and low phase noise. Therefore the VHF/UHF frequency has to be derived from or locked to a crystal oscillator at lower frequency. This can be done either by frequency multiplication or by phase-locked loop (PLL-) techniques.

The paper compares the possible solutions and describes a modular approach to a family of crystal controlled oscillators such as PXO's, VCXO's and OCXO's in the VHF/UHF-range.

Results for a TCXO at 1.2 GHz and a VCXO at 278 MHz are presented.

2. Alternative concepts

2.1 Frequency multiplication

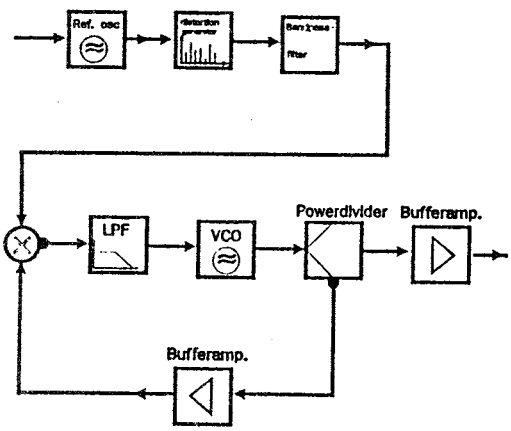
The simpliest principle for the generation of stabel VHF- or UHF-signals is the successive multiplication of the output frequency of the crystal-controlled reference oscillator.

The higher the final frequency, the more complex become the necessary efforts for effectively generating and filtering of the desired harmonics. A large number of tuned L-C-filters or crystal filters (or other types of filters) are required, making the adjustment work of the whole unit more and more tedious. The output spectrum shows a whole "fence" of unwanted frequencies (harmonics and subharmonics), whose attenuation is determined by the amount of selection means, the Q of the filter elements and the accuracy of the adjustment. The advantage however is, that the phase noise is not much degraded and is mainly determined by the reference oscillator itself, the multiplication factor, and the inherent noise of the frequency multipliers.

In the PLL-based concepts, a voltage controlled oscillator (VCO) based on an oscillator with lumped L-C-elements, stripline components, SAW resonators, or ceramic resp. dielectric resonators, is locked to the crystal-controlled reference oscillator. Two alternative approaches may be used: harmonic sampling or PLL with frequency division.

2.2 Harmonic sampling

The schematic diagram is given in Fig. 1. Strong harmonics of the reference oscillator output signals are generated by distortion into very short pulses. The desired harmonic is filtered out and fed to a phase detector, in which its phase is compared to the signal of the VHF/UHF VCO. Through a low-pass filter, the output of the phase detector controls the VCO and locks it to frequency.



The advantage of this method is, that the phase loop operated on a high frequency, and therefore the response characteristics can be extremely fast, and the phase noise close to the carrier is improved.

The disadvantage is, that good filtering of the desired harmonic is necessary, as the unwanted harmonics appear as side lines in the output spectrum of the VCO. For higher frequencies, much circuitry is necessary for the frequency multiplication.

2.3 PLL with frequency division

The principle is depicted in Fig. 2. The phase detector input is either connected directly to the reference oscillator or is fed by the reference frequency divided by factor T. This frequency is compared to the VCO output frequency divided by a larger factor N. Therefore the output frequency of the VCO is locked to

$$fo = f_{Ref} * N/T$$

Advantageous is, that the phase detector loop works on a rather low frequency, and thus can be realized by commonly available integrated circuits. For the high frequency division by N also a variety of standard ICs is available, mainly based on ECL resp. GaAs technology.

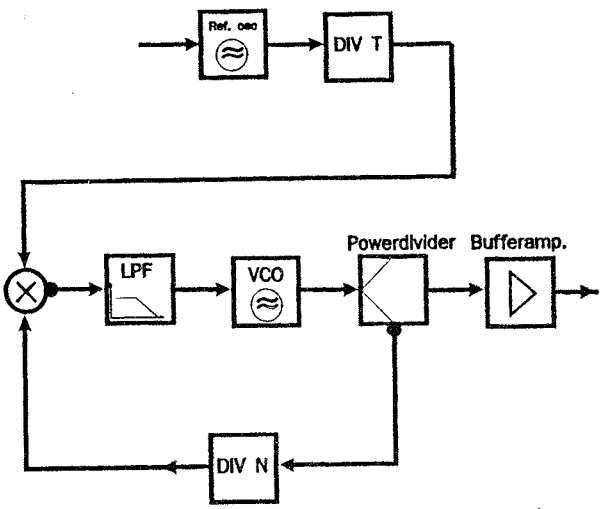


Fig. 2: PLL with frequency division

The output spectrum shows mainly two sidelines at fo+-f_{Ref}/T. In poor designs, also several multiples of f_{Ref}/T may occur.

The disadvantage is, that the phase-locked loop works at a rather low frequency, which yields a slower response characteristic. The phase noise of the reference oscillator is degraded by the factor N, the overall phase noise is strongly influenced by the control loop design.

3. Phase noise considerations

The influence of different noise sources on the phase noise performance can be analyzed by insertion of noise sources into the phase control loop, as shown in Fig. 3.

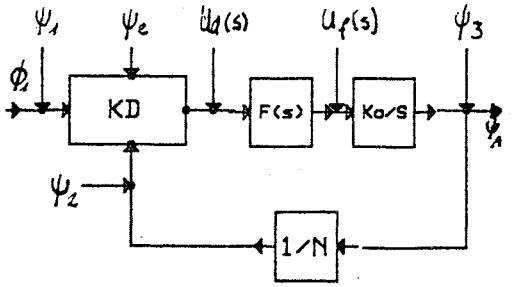


Fig. 3: Noise sources in the phase control loop
K_D is the transfer constant of the phase detector, F(s) is the frequency response of the loop lowpass filter, the Psi-values are the phase fluctuations of the different stages.

The phase fluctuation at the output of the PLL-circuit can be derived from this picture to be

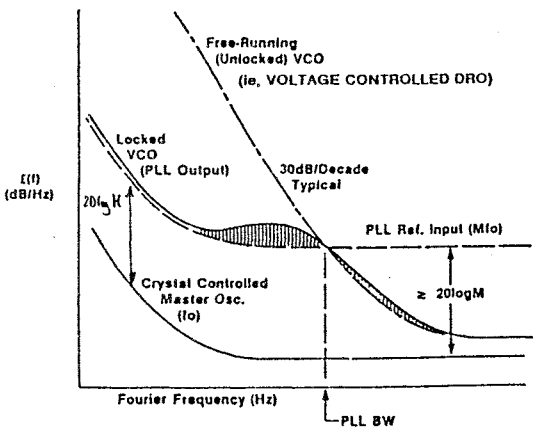
$$\psi_A = (\psi 1 + \psi 2 + \psi_\Theta) \cdot N \cdot \frac{1}{1 + 1/(\omega_n^2 \tau_1) \cdot s + 1/\omega_n^2 \cdot s^2} + \psi 3 \frac{N s + \tau_1 s^2}{K_D K_O N + N \tau_1 s}$$

The first part of this expression shows a lowpass characteristics. Noise at the frequency control input port of the VCO appears at the output with limited bandwidth. The phase fluctuations of divider, mixer, and reference oscillator are summed up and a factor, which increases this term proportionally. The lower the inherent phase noise of these components, the lower is the overall noise. The noise contribution of this term is also proportional to the frequency division factor N, which therefore should be as low as possible. The phase noise close to the carrier, within the loop bandwidth of the PLL, is mainly determined by the noise of the PLL components, especially by the close-to-carrier phase noise of the reference oscillator.

The second part of the formula represents a highpass characteristics. The phase noise at Forier frequencies larger then the loop bandwidth are mainly governed by the VCO. Therefore it should be realized with a lownoise circuit design using a high effective-Q of the resonator element.

Furthermore it can be seen from the formula, that a low pole frequency w_n is advantageous for good phase noise performance.

Typical phase noise spectra for a free-running and a locked VCO are given in Fig.4.



4. Realization

Based on the PLL technique with frequency division, a modular approach was developped which allows to realize crystal-controlled oscillators of different kinds in the VHF-/UHF range up to 1.3 GHz. The upper limit is currently only set by availability of inexpensive frequency divider IC's and the maximum recommended power dissipation in a miniaturized package, which can be accepted without degradation of stability and warm-up characteristics.

The basic unit in this approach is a module consisting of the phase detector, the lowpass filter, the VCO with its output stage, and the VHF/UHF frequency divider.

This module, which is located on a pc-board of 33 mm *24 mm, can be combined with any suitable reference oscillator, which may be a PXO (packaged crystal oscillator), a VCXO (voltage controlled crystal oscillator), a TCXO (temperature controlled crystal oscillator), a DTCXO (digitally compensated crystal oscillator), a TCVCXO (combination of TCXO and VCXO), or even an OCXO (oven controlled crystal oscillator). Thus, depending on the combination, we directly get a PXO, VCXO, TXCO etc. in the VHF-/UHF frequency range.
In this modular system, the complete unit is housed in a standard oscillator enclosure with the dimensions 36.1mm*27.2mm*19.4mm (L*W*H*). The VHF-/UHF-pc-board is directly mounted onto the inside part of the bottom plate of the oscillator enclosure. On top of this module, the second pc-board, containing the reference oscillator, is attached (see Fig.5). This oscillator can be selected out of a variety of existing oscillator models (PXO's, VCXO's, TCXO's etc.), which were already designed to fit into the same housing.

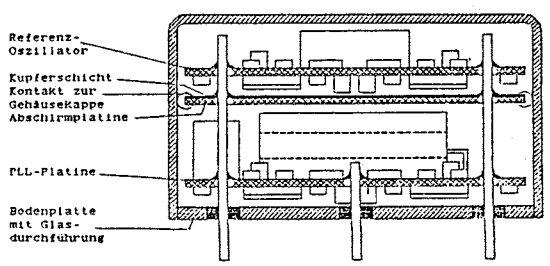


Fig.5: modular construction

Fig. 6 shows as an example the UHF-part of a TCXO operating at approximately 1.2 GHz.

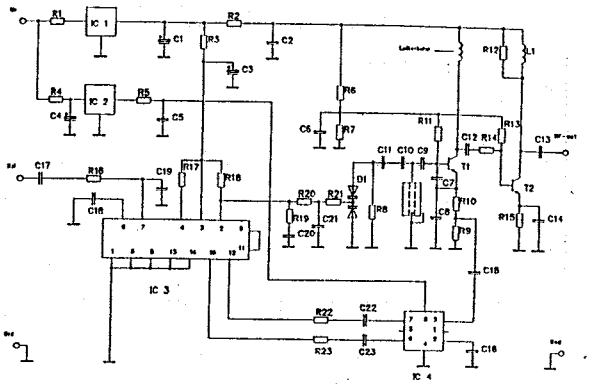


Fig.6: UHF part of a 1.2 GHz TCXO

The VCO is realized with a dielectric resonator with an open-circuit Q of approximately 1000. The temperature coefficient of the free-running VCO is shown in Fig. 7.

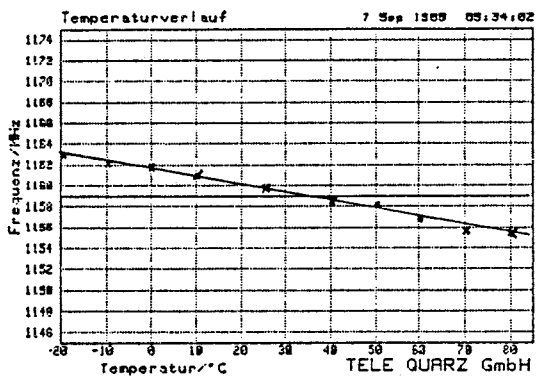
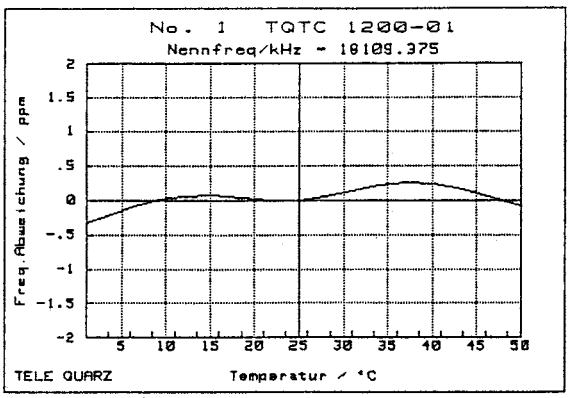


Fig. 7: temperature coefficient of the 1.2 GHz VCO (free-running)

The frequency divider is a two modulus prescaler in ECL technology.

The TCXO has a temperature stability of better than 0.5 ppm in 0°C to +50°C (other stabilities in other temperature ranges can be easily derived). The current consumption is 70mA at 15V supply voltage with an output signal of typical 0dBm.

Fig. 8 shows typical curves for the frequency stability vs. temperature. In Fig. 9 the output spectrum of the oscillator unit is depicted. The two sidelines at f_o+-f_{Ref} are -70dB or more down, the spectrum far-off is extremely clear. The phase noise is below -100dBc at 100 Hz and above.



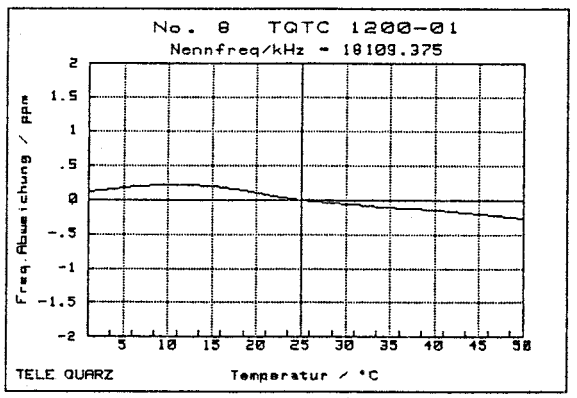


Fig. 8: frequency stability vs. temperature 1.2 GHz TCXO

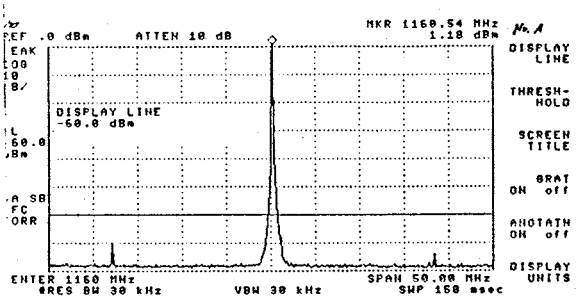


Fig. 9: Output spectrum of the 1.2 GHz TCXO

Another example is an VCXO at 278.528 MHz, used for frequency multiplex applications. In this case, the whole circuit, including all components, could be placed on one pc-board 33mm*24mm. Thus the housing could be reduced in height to 12.8mm.

The VCXO works with a supply voltage of -5.2V and draws 70mA. It has two complementary ECL outputs with rise and fall times of approx. 350 ps. The frequency pulling range is +-200 ppm +-10%.

The 278 MHz-VCO uses a tuned LC-circuit, which is optimized for low phase noise.

In Fig.10 the measured pulling characteristic is shown.

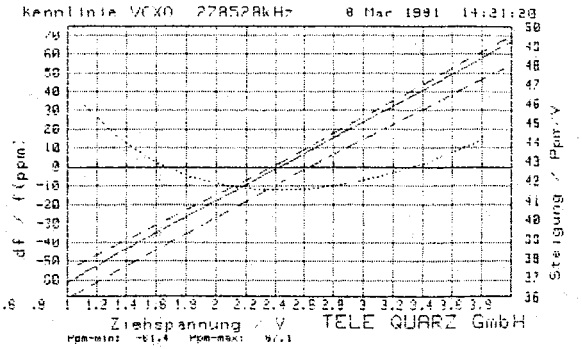


Fig. 10: Pulling characteristic 278.528 MHz VCXO

In the output spectrum, the lower side line at fo-fRef is attenuated only by -54 dB, whereas the upper side line is 70dB down. This is due to the compact construction of the unit.

The first harmonic is 32dB below the carrier (see Fig.11).

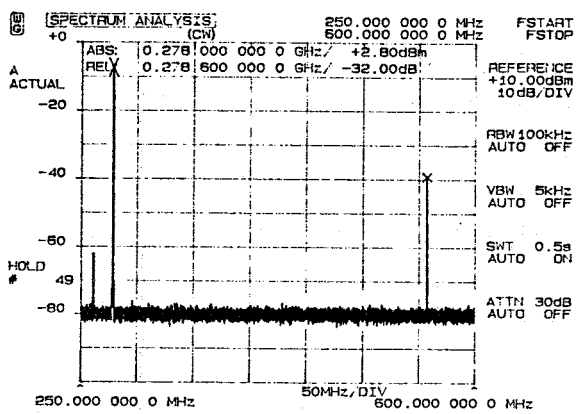


Fig.11: output spectrum VCXO 278.528 MHz

Fig.12 is a close-up view of the noise sidebands of the carrier frequency taken from a spectrum analyzer.

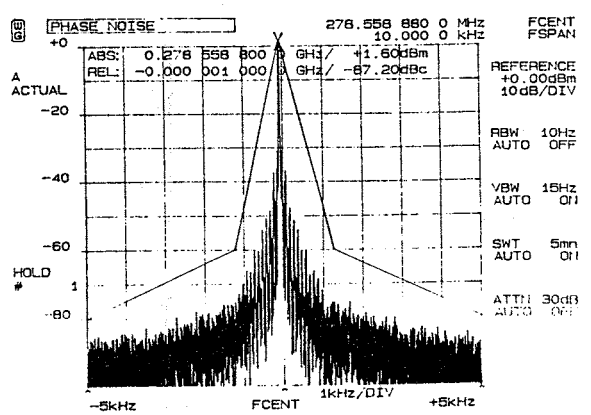


Fig.12: Noise sidebands of the 278.528MHz signal

In Fig.13 the noise sidebands of the VCO itself are shown, both in free-running mode and when it is locked to the crystal oscillator.

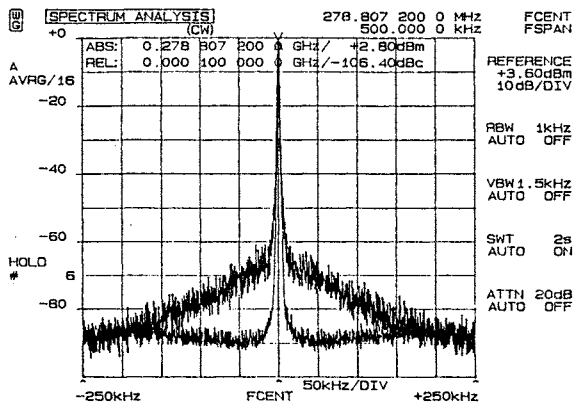


Fig.13: VCO noise sidebands free-running/locked mode

The dynamic behaviour of the 278.528MHz VCXO frequency is very important, because in the application the unit is integrated into an other PLL-loop. Fig.14 displays the over-all time response of the output frequency for a pulling voltage step from 0.2V to 2.3V. It can be seen that the response is very smooth.

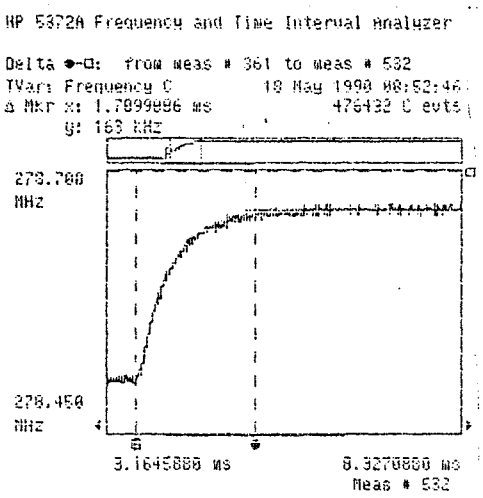


Fig.14: time responce of the output freqeuncy to a pulling voltage change

Finally Fig.15 shows the frequency response of the modulation, i.e. how the frequency deviation of the FM signal changes at higher modulation frequencies. The 3dB corner (modulation bandwidth) is above 8kHz, which is better then the system requirements.

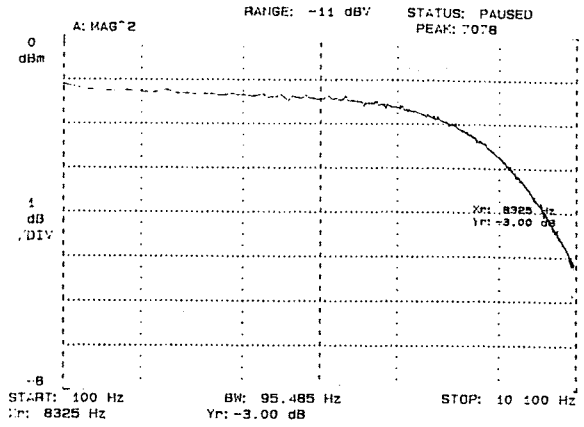


Fig.15: frequency response of the modulation

Session 8A Number 71

MAGNETIC SENSITIVITY OF OSCILLATOR COMPONENTS

R. Brendel*, C. El Hassani*, P. Krux*, E. Robert**

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S. associé à l'Université de Franche-Comté-Besançon
32, avenue de l'Observatoire - 25000 Besançon - France
** CNES, DRT/TIT/TF, 18 avenue E. Belin, 31055 Toulouse Cedex

Introduction

Orbitographic missions such as DORIS project intended by CNES requiers high stability oscillators (USO) on board as well as on ground. To this end these devices have to exhibit high immunity against environmental perturbations such as temperature, irradiations, vibrations, etc. Recently, attention has been focused on the oscillator magnetic sensitivity whose order of magnitude may reach in some case the order of magnitude 10^{-11}/gauss. Such a value is much greater than the DORIS specification (10^{-13}/gauss).

An oscillator is an assembly of a more or less large number of components whose value defines the oscillating frequency. For that reason, individual magnetic sensitivity of each component may contribute to the sensitivity of the whole.

In order to understand the phenomenon, CNES has initiated a large number of experiments in various laboratories (ONERA, CNES, CNRS), the first experiments have already demonstrated the responsability of some components like inductors or even the resonator itself.

This work presents the results of recent experiments carried out in our laboratory. The magnetic sensitivity of each component of the oscillator has been investigated : capacitor, varicap, transistor, etc. The various experimental setups are described and their sensitivity are compared.

It is shown in this paper that many of the studied devices present a magnetic sensitivity which in most cases strongly depends on the orientation of the component with respect to the magnetic field. The incidence of the individual magnetic sensitivity on the overall oscillator sensitivity is discussed.

1. Oscillator sensitivity

The magnetic sensitivity of the oscillator as a whole has been evidenced on many occasions in the past at CNES [1,2], at ONERA [3, 4] and at CNRS [5]. All experiments have demonstrated a pronounced anisotropy of the oscillator's magnetic sensitivity [1-5]. Today, the main question is : which is or which are the component(s) responsible of the magnetic sensitivity ? Experiments done at ONERA have already evidenced the significant role of some ferrite core inductors used as mode or overtone selectors in many oscillators [3, 4]. However it is not sure that other components could not exhibit their own sensitivity, this is why a battery of experiments has been performed in order to assess the magnetic sensitivity of the various oscillator's components starting with the resonator.

2. Resonator sensitivity

By using an experimental oscillator in which only the resonator is submitted to the magnetic field [5], it has been possible to show that the resonator has its own sensitivity to the magnetic field. It has been explained by the fact the crystal is generally supported by two or more thin strips of nickel alloy which are bent by the magnetic field. This bending induces strains in the crystal so that its frequency changes.

Fig. 1 shows three records obtained with three different orientations of the magnetic field with respect to the resonator (regular structure) in all cases the magnetic field was about 1,300 gauss, for this intensity, the magnetic sensitivity can be roughly estimated from 10^{-12}/gauss for the less sensitive direction up to 10^{-11}/gauss for the most sensitive direction.

Fig. 2 shows the results obtained with another resonator of a same structure, whose sensitivity is slightly smaller ranging from 6.10^{-12}/gauss to 10^{-11}/gauss.

Fig. 3 shows the sensitivity of a BVA resonator, which is fairly smaller than for other resonators and its anisotropy is much less pronounced : from $1.5\ 10^{-12}$/gauss to $3.6\ 10^{-12}$/gauss. The mechanism of the BVA's sensitivity has not yet been explained.

Here should be noted that all measurements have been performed with rather high magnetic fields.

The sensitivity curves represented on Figs. 1, 2, 3 for both regular and BVA show that for small fields the sensitivity can be different than for high fields, although it is probably not drastically different.

3. Experimental setups for other components

Two methods have been used to study the sensitivity of other components like capacitors or inductors. The first method involving a phase bridge method is illustrated on Fig. 5. The component under test is placed in the magnetic field while the remainder of the circuit is kept far from the magnetic influence. In the second method, the component under test is a part of an oscillator (Fig. 6), Fig. 7a shows the oscillator used for inductor sensitivity measurement and Fig. 7b the one used for capacitor sensitivity measurement. Among the different possible solutions in both cases the oscillator has been chosen in such a way that one component terminals can be grounded so the component under test can be put at the end of a coaxial cable.

In some cases a third method shown on Fig. 8 has been used in order to enhance the measurement sensitivity of the experimental setup. It involves the same oscillator as used in the second method described above but instead of directly recording the frequency variations, the synthesizer is locked on the oscillator frequency and the error voltage of the phase lock loop feeds a Fast Fourier Transform analyzer. In that case, if the device under test is sensitive to the magnetic field, the error voltage exhibits lines at the magnetic field excitation frequency and its higher order harmonic components.

Here should be noted that in all cases, the component under test and the oscillator are separately ovenized. In spite of the drift due to the temperature change, some experiments performed after the component oven was cut off have proved it cannot be incriminated in the observed magnetic sensitivity, in addition, the strong anisotropy of some component behavior shows that the phenomenon cannot be attributed to a thermal effect.

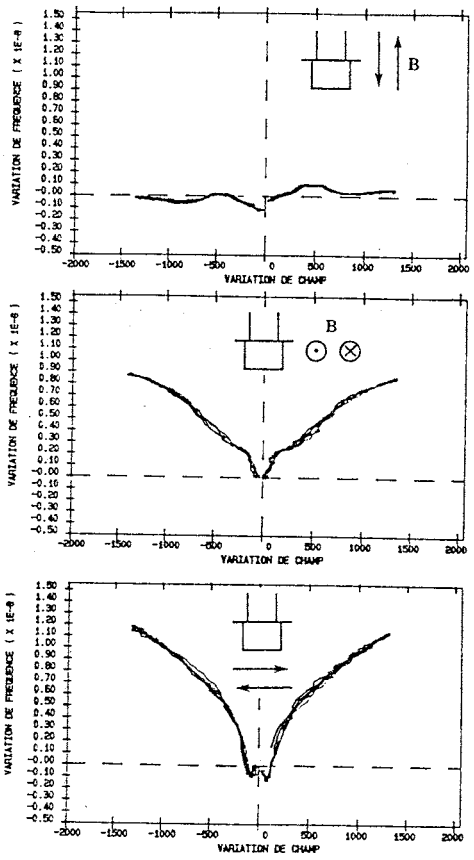


Fig. 1

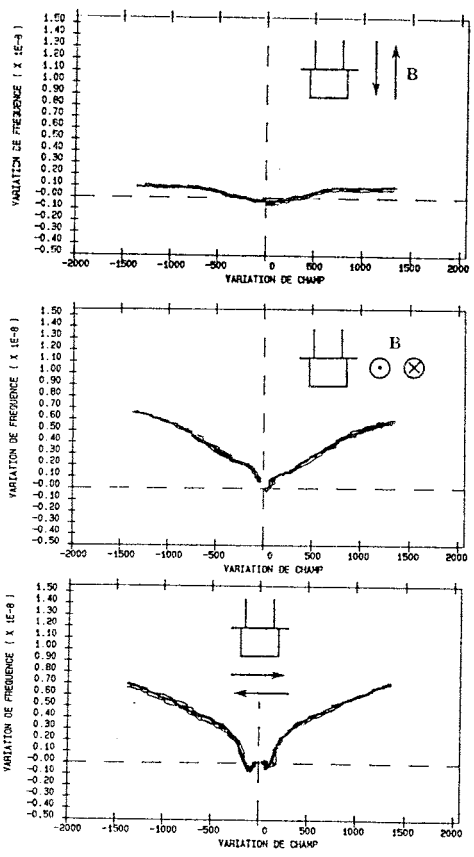


Fig. 2

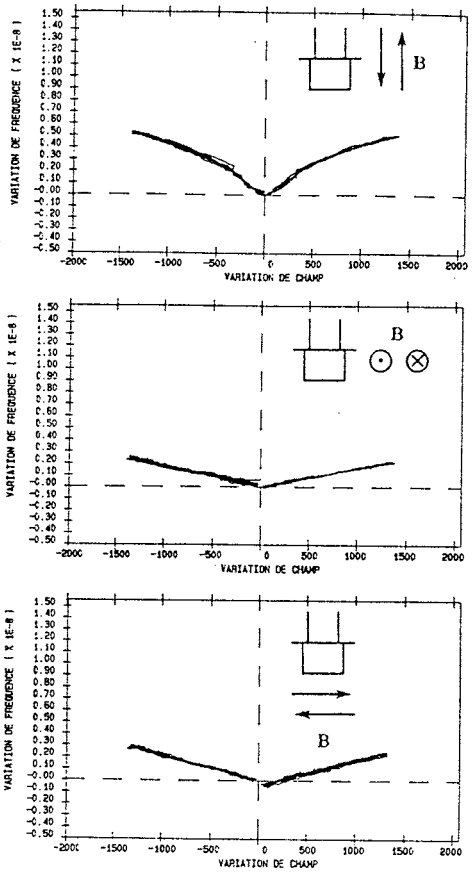


Fig. 3

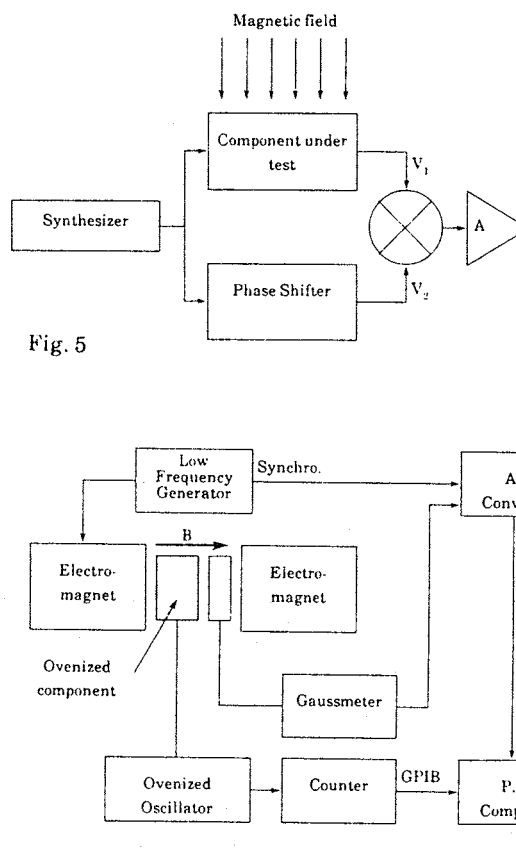


Fig. 5

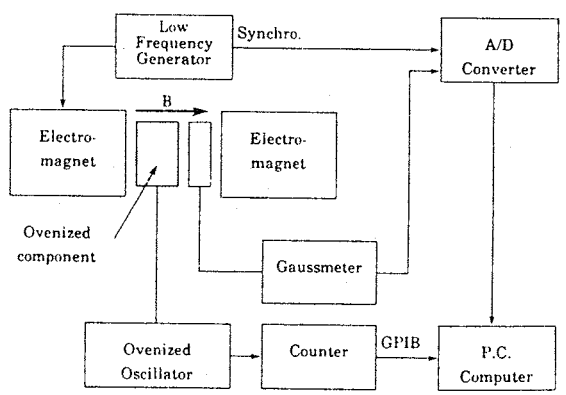


Fig. 6

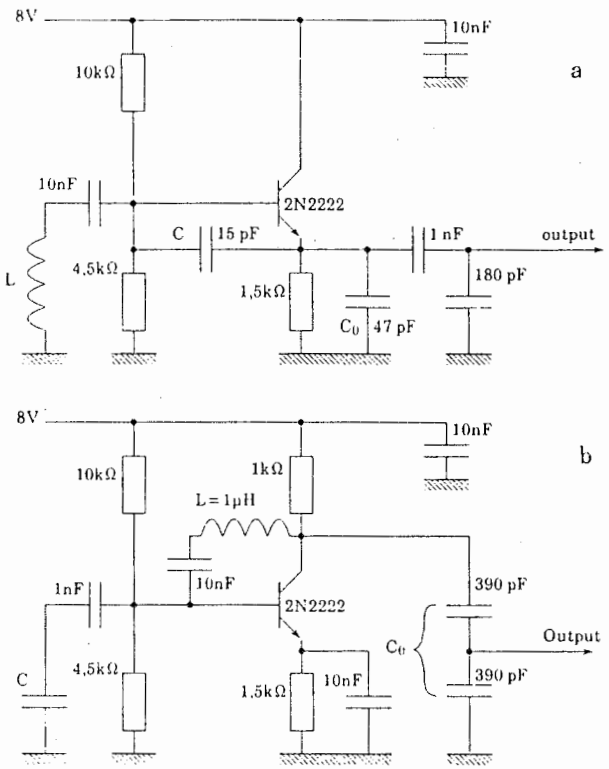


Fig. 7

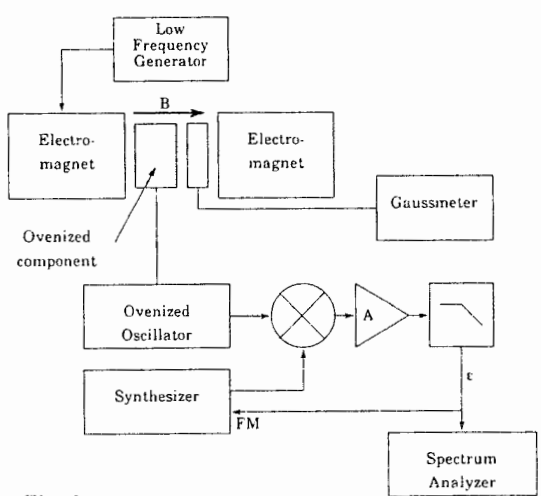


Fig. 8

4. Inductor sensitivity measurements

Two kinds of inductor have been investigated : ferrite core and air core inductors. Figs. 9 and 10 show the results obtained with a ferrite core inductor of 6.8 µH by using the phase bridge method (Fig. 9) and the oscillator method (Fig. 10). Both experiments confirm the earlier published results on the magnetic sensitivity of that kind of component [4] : a strong sensitivity with an approximate quadratic behavior and a rather marked anisotropy. The largest sensitivity being observed when the magnetic field is directed along the coil axis.

The sensitivity of air core inductors is much less important than the former and cannot be detected by using the phase bridge method. Fig. 11 shows the results obtained with an 8.2 µH air core inductor by using the oscillator method.

The records show that the anisotropy is less marked than for ferrite core and the sensitivity is quite linear.

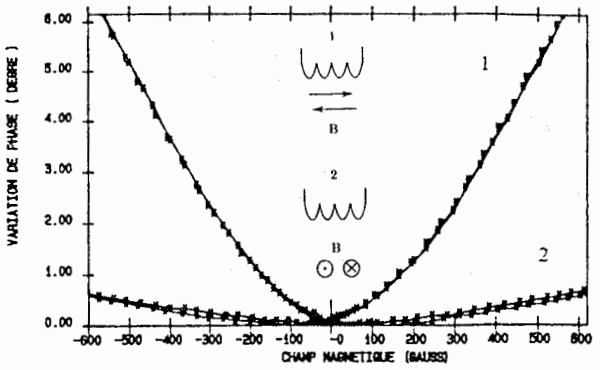


Fig. 9

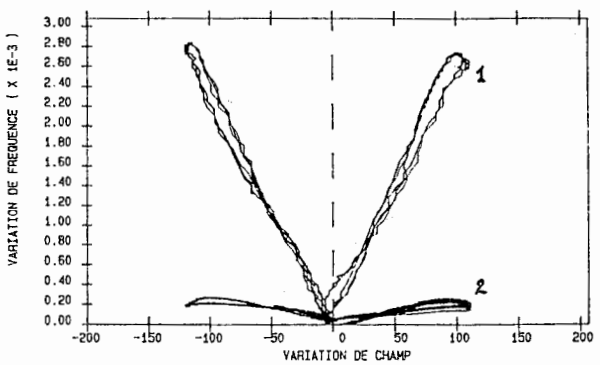
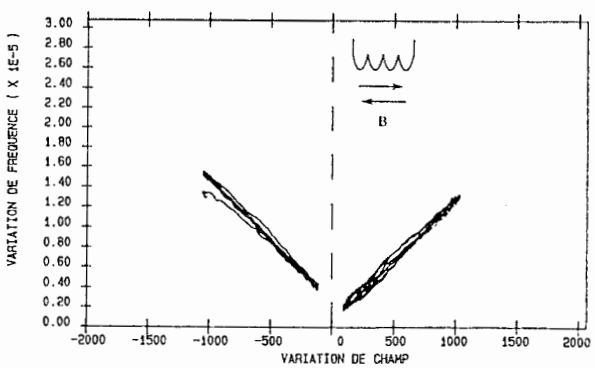


Fig. 10



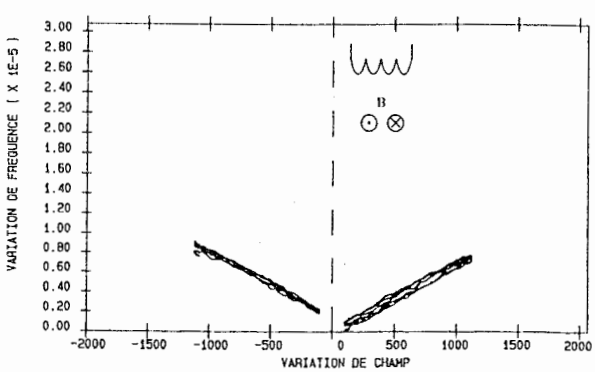
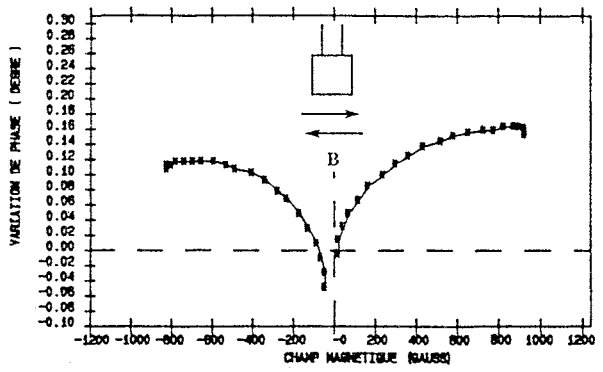


Fig. 11

5. Capacitor sensitivity measurements

Experiments have been concerned with regular or SMT ceramic capacitors. Fig. 12 shows the results obtained with a 100 pF ceramic capacitor by using the phase bridge method and Fig. 13 shows the records done with SMT capacitors by using the oscillator method. In the latter case, three capacitors have been connected in parallel so as to get a large enough capacitance permitting a good measurement sensitivity (see Sec. 8 below).

These figures show that capacitors also exhibit a magnetic sensitivity but unlike the inductors, their sensitivity doesn't noticeably depends on the magnetic field orientation and the variation looks like a squared root law. In all cases the capacitance decreases when the magnetic field increases regardless of its direction.



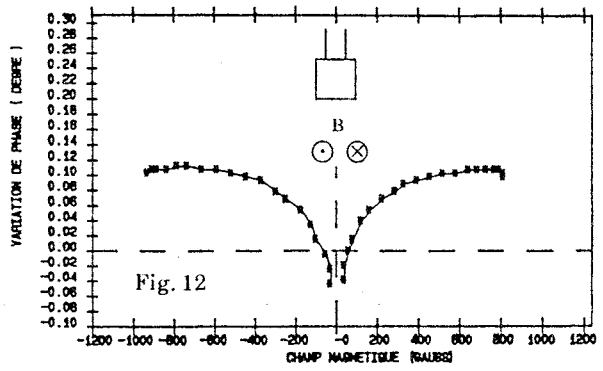
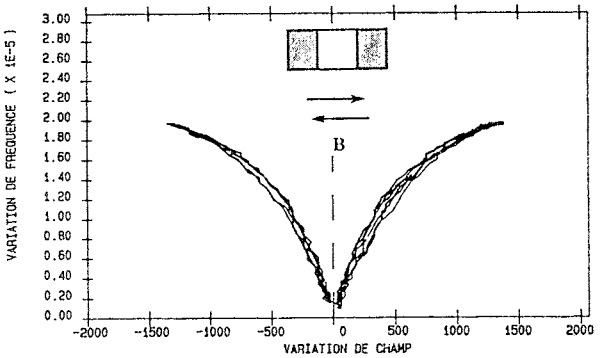


Fig. 12



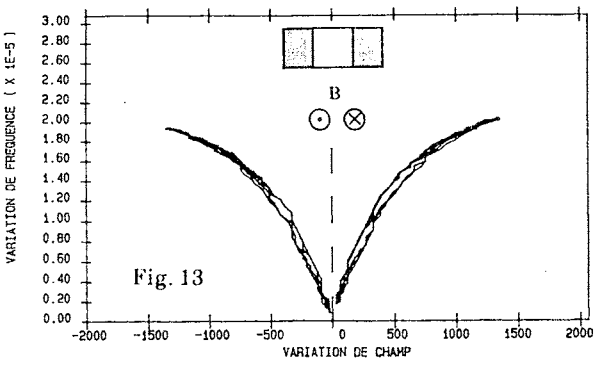


Fig. 13

6. Sensitivity of voltage variable capacitance diodes

The sensitivity of these devices has been studied by using the oscillator method. The experimental setup is the same as on Fig. 7b. The diode is put at the end of the coaxial cable at the capacitor place while the bias circuit is connected to the other end outside the magnetic influence. The records on Fig. 14 show that once again this component is sensitive to the magnetic field and this sensitivity depends on the orientation of the magnetic field. Fig. 15 represents the sensitivity of a voltage variable capacitance diode 1N5483 versus magnetic field orientation obtained by using the phase lock loop method described on Fig. 2. The direction angle is null when the magnetic field is directed along the diode axis.

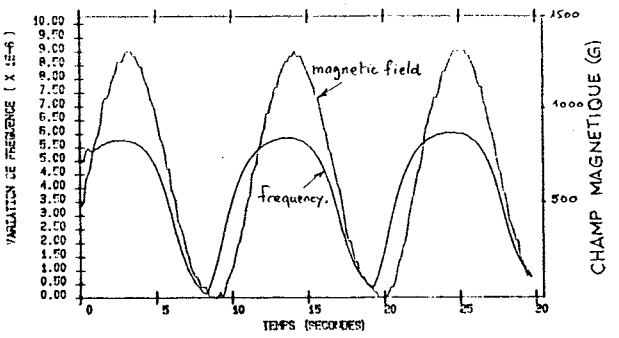


Fig. 14

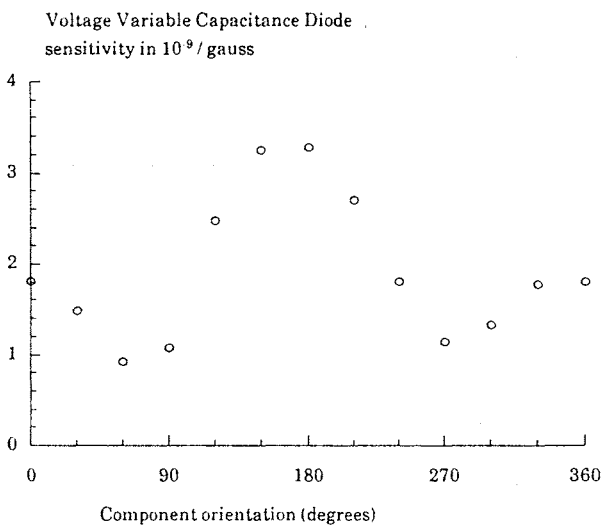
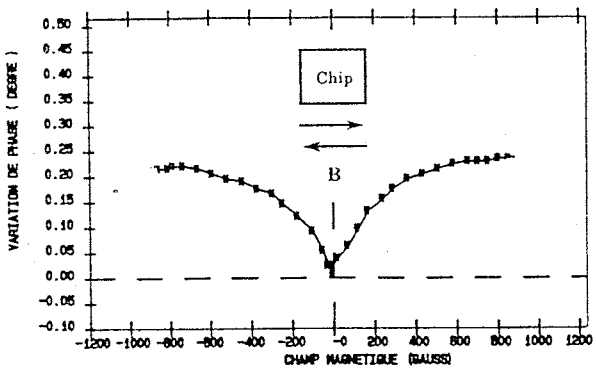


Fig. 15

7. Transistor sensitivity

The transistor sensitivity is among the most difficult measurements to deal with because it is practically impossible to place it alone in the magnetic field. So, experiments have been done by using the phase bridge method in which the transistor is put in the magnetic field as well as the resistors of the biasing circuit. However, the DC voltage was kept outside the magnetic field. Although there remains a little doubt, it can be considered that the observed sensitivity is mainly due to the transistor, in fact, all experiments intended to evidence the resistor sensitivity have failed.

Fig. 16 shows the results obtained for two different orientations of a 2N2857 transistor, the maximum sensitivity is observed when the magnetic field is directed perpendicularly to the chip plane.



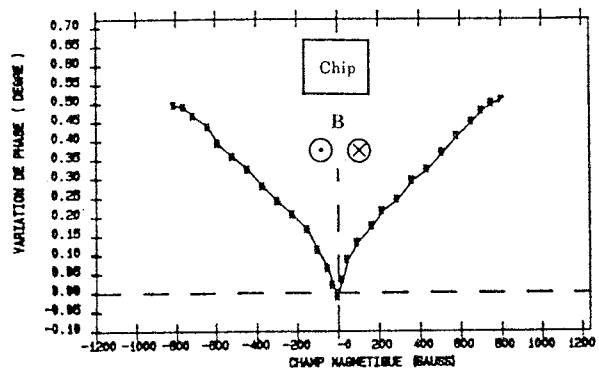


Fig. 16

8. Discussion on the measurement sensitivity

a) Phase bridge method

In this experiment (Fig. 5) the double balanced mixer is fed by two signals V_1 and V_2 shifted by about 90° so its output voltage is proportional to the phase difference between V_1 and V_2.

The synthesizer output impedance and the mixer input impedance being near 50 Ω, the equivalent circuits from the mixer input point of view looks like Fig. 17. Then the measurement sensitivity takes the form :

$$\frac{dV_s}{dC} = \frac{-a R \omega_o}{1 + (R C \omega_o)^2}$$

where V_s is the amplifier output voltage (see Fig. 5) and a a constant which depends on the mixer slope and on the amplifier gain. This coefficient can be easily measured by giving the phase shifter two adjacent values and measuring the following voltage difference. In the present case a is about 3.3 mV/degree. Thus, a 100 pF capacitance and a 10 MHz working frequency, lead to an overall sensitivity of about 0.5 mV/pF. Then the resolution of the measurement setup is limited to about 0.1 pF because of the noise level.

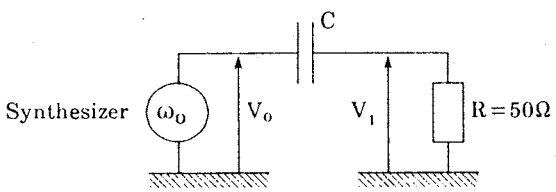


Fig. 17

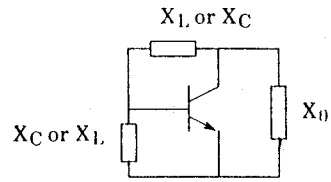


Fig. 18

b) Oscillator method

For both oscillators used (Figs. 7a and b and 18) the oscillation condition can be written :

$$X_c + X_L + X_o = 0$$

where X_c, X_L and X_0 are the oscillation loop reactances. Then, the oscillation frequency is

$$\omega^2 = \frac{1}{L}\left(\frac{1}{C} + \frac{1}{C_o}\right)$$

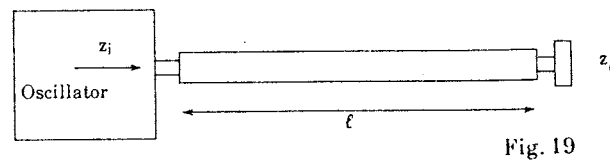


Fig. 19

This expression don't take into account the coaxial cable required to keep the oscillator circuit far from the magnetic field, in fact the impedance "seen" by the oscillator is modified by the wave propagation along the transmission line (Fig. 19) it is related to the load impedance by the well known relation :

$$z_i = \frac{1 + \rho e^{-2j\beta\ell}}{1 - \rho e^{-2j\beta\ell}}$$

$$\rho = \frac{z_\ell - 1}{z_\ell + 1}$$

$$\beta = \omega / v$$

z_i and z_ℓ are normalized impedances with respect to the line characteristic impedance R_0, ρ is the reflection coefficient at the load, β the propagation constant, v the propagation velocity and ℓ the line length.

Because of the line, the oscillation frequency and its variations are fixed by z_i rather than z_ℓ. It is given by one of the following equations :

- Inductor measurement :

$$Im\left(R_o \frac{1 + \rho e^{-2j\frac{\omega\ell}{v}}}{1 - \rho e^{-2j\frac{\omega\ell}{v}}}\right) - \frac{1}{\omega}\left(\frac{1}{C} + \frac{1}{C_o}\right) = 0$$

with

$$\rho = \frac{jL\omega - R_o}{jL\omega + R_o}$$

- Capacitor measurement :

$$Im\left(R_o \frac{1 + \rho e^{-2j\frac{\omega\ell}{v}}}{1 - \rho e^{-2j\frac{\omega\ell}{v}}}\right) + L\omega - \frac{1}{C_o\omega} = 0$$

with

$$\rho = \frac{1 - jR_oC\omega}{1 + jR_oC\omega}$$

These equations can be solved by using a computer, they also allow to get the sensitivity of the measurement system which can be expressed as :

$$\frac{\Delta\omega}{\omega} = k_L \frac{\Delta L}{L} \quad \textit{for inductors}$$

$$\frac{\Delta\omega}{\omega} = k_c \frac{\Delta C}{C} \quad \textit{for capacitors}$$

According to the experimental conditions, the correction factor may range from -0.05 to -0.5. Besides, the sensitivity of the measurement setup is limited by the oscillator stability. However, frequency variation of 10^{-6} are easily evidenced so that for a 100 pF capacitor, the corresponding capacitance variation are $\Delta C/C \approx 10^{-5}$ so that in the case of interest the measurement resolution will be about 0.001 pF that is hundred times better than the phase bridge resolution. Nevertheless the sensitivity strongly decreases when using small capacitances.

Table 1 summarizes the sensitivities of various components.

Component	Relative magnetic sensitivity (ppm/gauss)	Induced oscillator magnetic sensitivity (in 10^{-12}/gauss)
Ferrite core inductor	57.	200.
Air core inductor	0.028	0.1
SMT capacitor	0.24	0.8
VVC diode	0.003	0.0001
Transistor	-	4.
Regular resonator	-	12.
BVA resonator	-	4.4

9. Influence of the component sensitivity on the oscillator overall magnetic sensitivity

As for temperature effects, almost all oscillator components have their own magnetic sensitivity. Thus the overall influence on the oscillator behavior cannot be easily estimated by simple inspection. However, the oscillator magnetic sensitivity can be determined by using the simulator presented elsewhere in these proceedings [8].

The interface between environmental fluctuations like temperature and magnetic fields are at present being written but are not yet available, nevertheless Table 1 shows the individual effect some component sensitivities should have on a Colpitts oscillator frequency.

10. Conclusion

With a very few exceptions, almost all electronic oscillator's components exhibit a more or less strong magnetic sensitivity. In some cases that sensitivity has no noticeable influence on the oscillator performances but in other cases it may have awkward effects. This particularity has to be taken into account when designing new devices especially when they are submitted to varying environmental conditions as it is the case for on board USO's.

The magnetic sensitivity of some resonators as well as inductors has been explained. Although magnetic sensitivity of other components has been studied in the past [6, 7], it remains to be clarified.

As regard of the protection against the spurious magnetic effects on the oscillators, the problem looks like the temperature effects and some solutions used to minimize temperature effects can be transposed for the magnetic field effect protection. Three main methods can be considered :

- by developing new components insensitive to magnetic field : that solution may need a lot of time and efforts because the components have to be improved with respect to the magnetic field while keeping their other properties (mechanical, thermal, electrical ...).
- by using magnetic shielding : that solution has already been studied and used [9] but it is not always used in space applications because of the weight increase and of the lack of knowledge on the behavior of the magnetic shielding material during the launch and on its aging characteristics.
- a possible solution could be found by using a compensation circuit acting like those used in TCXO's for temperature correction.

Acknowledgement

This work was supported by CNES under contract 932 88 53 47400.

References

[1] A. Benezet, E. Robert, "Mesure de susceptibilité au champ magnétique de plusieurs OUS", Compte-rendu d'exécution d'essais n° L5160, CNES (1987).

[2] E. Robert, M. Brunet, "Susceptibilité magnétique des oscillateurs à quartz spéciaux, état des connaissances", Technical Report 88/354, CNES (1988).

[3] F. Deyzac, "Oscillateurs à quartz de bord en environnement sévère. Sensibilité magnétique", Technical Report n° 37/7132 PY, ONERA (1988).

[4] F. Deyzac, "Magnetic sensitivity of quartz oscillators", 4th EFTF (1990).

[5] R. Brendel, C. El Hassani, M. Brunet, E. Robert, "Influence of magnetic field on quartz crystal oscillators", 43rd Ann. Freq. Cont. Symp. (1989).

[6] N. Saleh, "h_{FE} degradation in the presence of intense magnetic field", J. of Appl. Phys., vol. 45, n° 2 (1974).

[7] N. Manohara Murthy, P. Mallikarjuna Reddy, S.V. Subrahmanyam, "Effect of magnetic field on the current gain of junction transistors", Indian Journal of pure and applied physics, vol. 18 (1980).

[8] R. Brendel, G. Marianneau, E. Robert, "Nonlinear modelling techniques for quartz crystal oscillator", These proceedings.

[9] J.P. Granger, "Blindage magnétique de l'OUS de DORIS", Rapport final d'étude S2-NT-10000-3800-MT - MATRA ESPACE (1989).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 8A Number 72

FIRST MEASUREMENTS ON OCIR CONTROLLED OVENS

B. ABBOUD, S. GALLIOU, M. MOUREY, J-P VALENTIN

ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Laboratoire de Chronométrie, Electronique et Piézoélectricité
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

The feasibility of O.C.I.R.* design was proved one year ago (1) ; so infrared ovens were performed for quartz resonators. A brief presentation of prototypes with corresponding theoretical analysis is made. A second part is reserved to the infrared ovens : thermal warm-up, thermal resistance between the heating source and external environment, measurements of power consumption versus the ambient temperature are presented. The third part shows the frequency behavior during the warm-up time, i.e. the necessary times to obtain frequencies to with-in 10^{-7}, next 10^{-8} of the final frequency ; the frequency stabilities are given at 25°C for the short and medium terms.

* O.C.I.R. : French abbreviation of Oscillateur à Chauffage Infra-Rouge.

INTRODUCTION

Quartz is an absorbent material in the infrared wavelength scale. This property was used to perform infrared ovens for Quartz resonator. Some models were studied by electrical simulation, showing the feasibility of these ovens and allowing an adequate choice of parameters and elements.

The principal qualities of this heating mode are :

1) Homogeneous heating of quartz resonator.
2) Fast warm-up time.
3) Low power consumption (Heating under vacuum).

PRESENTATION OF PROTOTYPES

The quartz resonator is the center of the first prototype called OCIR 1. Two deposited electrodes are blackened by a thin layer of Chromium oxyde. Heaters are two black bulbs. The thermal sensor is a thermistor stuck on heaters. These parts are settled inside two reflecting shells manufactured in aluminium's alloy. The whole system is under vacuum (Fig. 1)

DEVICE 1 : OCIR 1

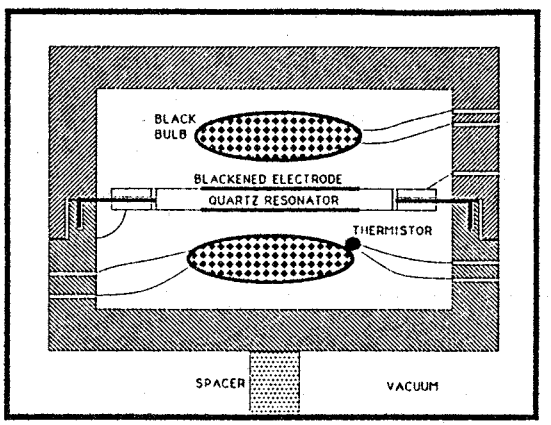


Fig. 1 : OCIR 1

Using the electrical analogy we built the equivalent circuit of this prototype. Quartz resonator and shells are represented by their thermal capacitors and resistors. In symetrical heating, black bulbs are represented by a current generator and a thermal capacitor. Similarly, two other current generators simulate radiative exchanges between heaters, quartz resonator and shells. Quartz supports and black bulbs wires are two thermal resistors with high values.

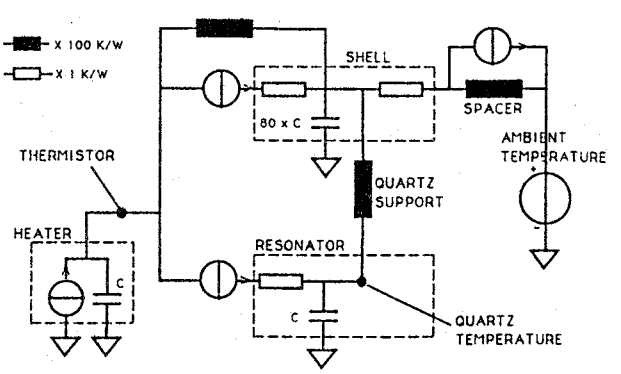


Fig. 2 : Electrical analogy of OCIR 1

Our second prototype is a performed version of OCIR 1. Aluminium alloy in shells was subtituted by insulating glass (80 % lead oxyde). Reflectivity and shape coefficient of these last one were performed to reduce radiative exchanges with the black bulbs (Fig. 3). This prototype is called OCIRIS (French abbreviation for : "OCIR Isolé").

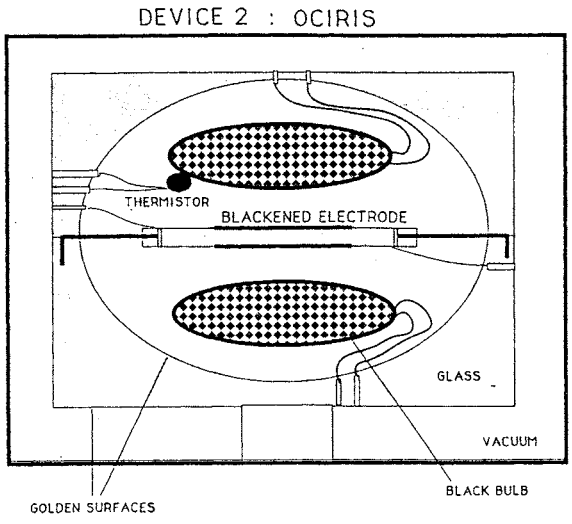


Fig. 3 : OCIRIS

In the equivalent electrical circuit the reduction of radiative exchanges between the black bulbs and the shells is represented by a smaller current generator. Thermal resistors representing the insulating glass of shells are higher than those of shells in aluminium alloy.

A very large thermal insulation is obtained between the quartz resonator and the external area (Fig. 4).

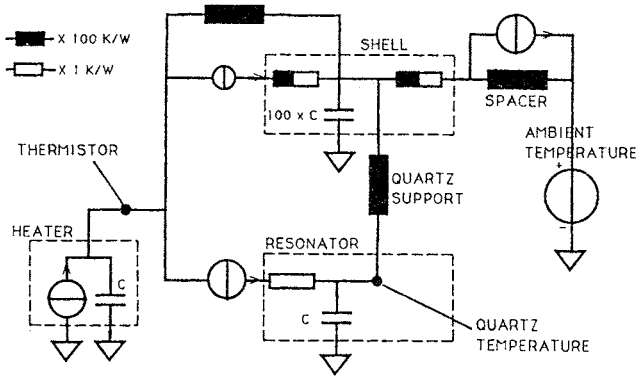


Fig. 4 : Electrical analogy of OCIRIS

The basic idea for the design of the third prototype was the establishment of a uniform temperature around the quartz resonator. So we have used very light shells in aluminium alloy on which black bulbs were stuck. This prototype was called OCIRAL (French abbreviation for "OCIR Allégé" - Fig. 5)

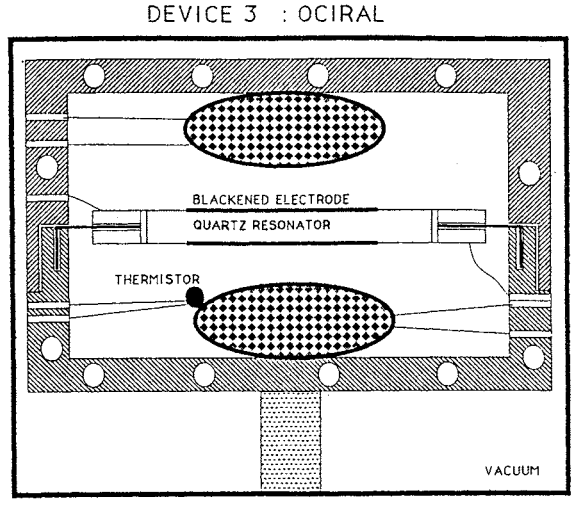


Fig. 5 : OCIRAL

In the equivalent electrical circuit (Fig. 6), light shells induce a lower value of the corresponding thermal capacitor and by sticking heaters on the shells the thermistor temperature becomes closed to the shells temperature.

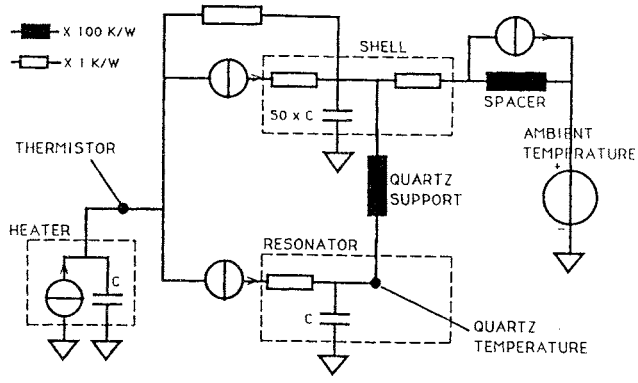


Fig. 6 : Electrical analogy of OCIRAL

THERMAL FEATURES

In order to qualify every prototype (and identify the thermal components of their equivalent network) each of them was submitted to a step of supply power. Their thermistor temperature and quartz temperature responses were registered. Last ones are obtained by means of B-mode frequencies (all resonators are standard 10 MHz-SC cuts). Fig. 7 and Fig. 8 show the evolution of these temperatures for OCIRIS and OCIRAL : Curves are almost identical but such a result needs about 500 mW for the OCIRAL device instead of only 125 mW for the OCIRIS device (here the ambient temperature is 25°C).

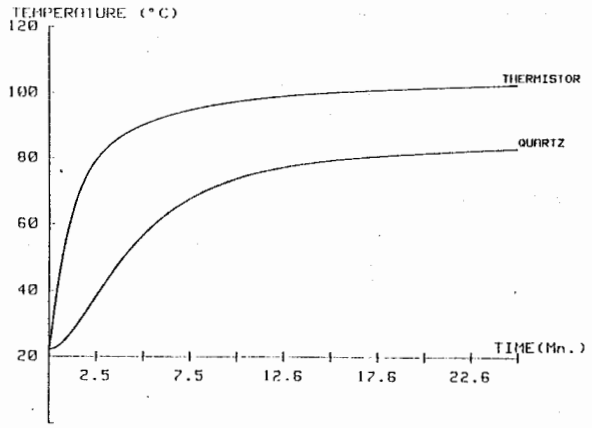


Fig. 7

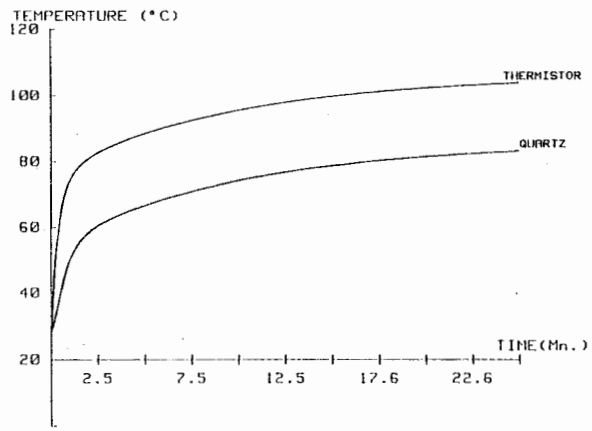


Fig. 8

This experiment was repeated for different ambient temperatures and everytime we have determined, after the establishment of the steady-state, the global thermal resistor between the quartz resonator and the external environment. The global thermal resistor depends on the ambient temperature as shown Fig. 9.

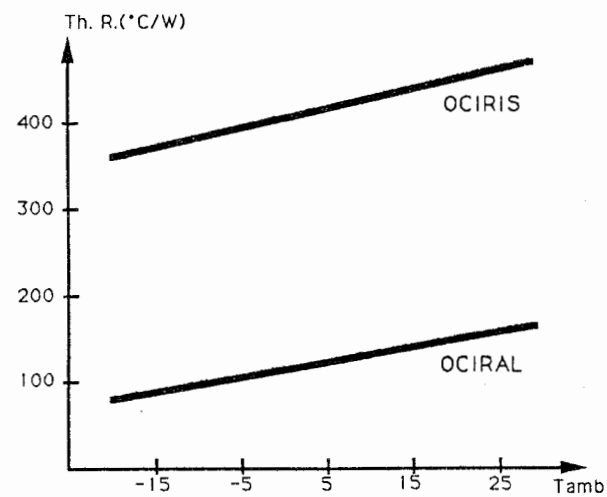


Fig. 9 : Global thermal resistor versus the ambient temperature

WARM-UP AND SHORT-TERM STABILITY

After the thermal characterization a closed-loop study was performed : the OCIR is used as a component of an oscillator. The oven temperature is controlled by means of a traditional electronic circuit using the thermistor stuck on one black bulb as a thermal sensor. For this step, experiments are done with standard electronic components which are not temperature controlled, keeping in mind that the electronic might be integrated and enclosed with the shells.

The heating power is applied symmetrically on both black bulbs. At 25°C, the warm-up time of OCIRAL, and OCIRIS systems we registered (Fig. 10). As above, the results are almost identical when the maximum available power at turn-on is three times lower for OCIRIS than for OCIRAL.

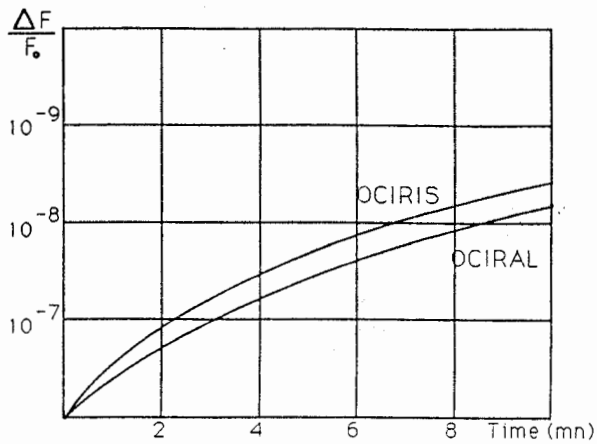


Fig. 10 : Warm-up

Results of short-term stability measurements are shown in Fig. 11. For large sampling time intervals (100 s) we may assume that our experimental conditions are not ideal because of the non-controlled temperature of the electronic.

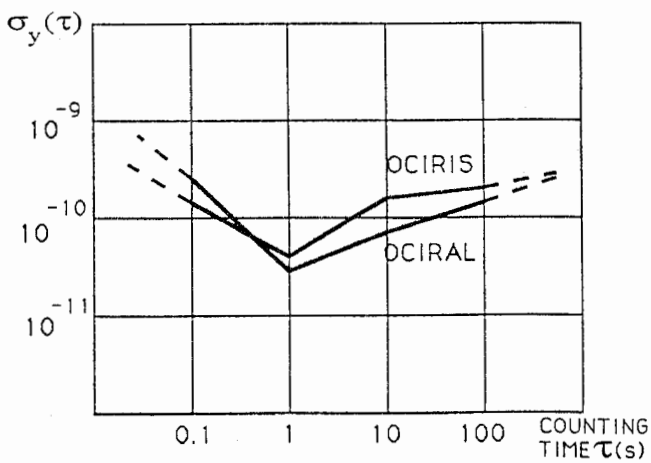


Fig. 11 : Short-term stability

TEMPERATURE CONTROL

OCIR system shows a peculiarity : a larger thermal gradient appears between the thermistor and the quartz resonator than in usual temperature controlled systems.

As in usual systems the controlled temperature is not the quartz temperature but the sensor temperature which is stuck on to the heater. For OCIR system the same principle was chosen because the stability conditions of the closed loop are easily reached. So the oven electronic is simplified. On the other hand the quartz temperature is not perfectly controlled and slightly depends on the ambient temperature.

An other way is to close the thermistor to the quartz crystal, or to use the B mode in the case of the SC cut (2). This last solution is reserved for the next OCIR prototypes. The drawback of this solution is to involve a large sophistication of the electronic control.

The best solution is to keep the sensor stuck on the bulb and to realize a compensation which cancels the effect of the thermal gradient between the thermistor and the quartz crystal. The compensation may be made by electronic.

A first one seems particularly adaptable to the OCIR design. It consists on using a non-symmetrical heating, the current through one of both bulbs being used as a sensor of the ambient temperature. A difference between the two applied powers appears when the ambient temperature is different from 25°C.

A second solution consists on keeping an usual temperature control circuit and making a correction of the reference voltage by means of a thermal sensor settled at the ambient temperature.

The establishment of a numerical function between the external temperature and the applied power to reach a constant quartz temperature is a third solution.

By means of one of these solutions the goal is to obtain a better result than that one obtained with TCXO devices.

COMPARISON - SUMMING UP

As soon as our prototypes will be perfected, they will be comparable to other systems classified in scale of low-power, fast warm-up, miniature oscillator ; that's why we have chosen to compare just now our first results with two other systems (U.S. device 1 and U.S. device 2 : both developed by american manufacturers (3)(4)(5)).

About the warm-up, two important parameters are taken into account : the maximum power at turn-on and the necessary time to reach 10^{-7} or 10^{-8} of the final frequency. In order to compare these two parameters simultaneously, we introduce a coefficient given by their product. This coefficient called warm-up coefficient is listed on Table 1.

	Warm-up coefficient (10^{-7}/10^{-8})	P_{Max} (turn-on)
U.S. Device 1	17/34 W.mn	8.4 W
U.S. Device 2	/ > 8 W.mn	8.8 W
OCIRAL	3.7/11.2 W.mn	1.5 W
OCIRIS	1.5/4 W.mn	0.5 W

Table 1

Concerning the power consumption good results are obtained with the thermal insulation of the OCIR devices. Table 2 lists compared power consumptions at 25°C (for OCIRAL and OCIRIS the turn-over point is about 75°C). With their thermal insulation, OCIR are also able to operate at low external temperature as shown in table 3, but the frequency change all over the temperature range must be improved.

U.S. Device 1	450 mW
U.S. Device 2	125 mW
OCIRAL	500 mW
OCIRIS	125 mW

Table 2

	Operating Temp. Range		P_{Max} (Turn-on)
U.S. Device 1	0° C	+ 50°C	8.4 W
U.S. Device 2	- 40°C	+ 80°C	8.8 W
OCIRAL	- 15°C	+ 65°C	1.5 W
OCIRIS	- 40°C	+ 65°C	0.5 W

Table 3

About the aging we have no result today. We hope a similar aging as an usual quartz resonator 10 MHz-SC cut (ϕ 13,2 mm), that means about 10^{-10}/day. The thin layer of Chromium oxyde deposited on electrodes might reduce the usual aging of the SC cut.

Short-term stabilities are compared in table 4

	0.1 s	1 s	10 s	100 s
U.S. Device 1	0.5	0.5	0.5	
U.S. Device 2	2	0.5	3	
OCIRAL	19	4	8	13
OCIRIS	14	4	10	14

Table 4 : $\sigma_y(\tau)$ in parts of 10^{-11}

Weight, volume and size of different systems are given in table 5

	Weight (g)	Volume (cm3)	Size (cm)
U.S. Device 1	40	36	3.3/3.3/3.3
U.S. Device 2		16	ϕ = 2.7 h = 3.0
OCIRAL	30	21	ϕ = 3.0 h = 3.0
OCIRIS	65	21	ϕ = 3.0 h = 3.0

Table 5

ACKNOWLEDGEMENTS

The authors wish to thank E. ANDREY for his contribution in the manufacturing of OCIR prototypes and J.J. BOY for his help in the measurements.

REFERENCES

(1) S. GALLIOU, M. MOUREY, R. DELAITE, J-P VALENTIN, "Theoretical analysis of infrared ovens for tactical oscillators", 4th European Frequency and Time Forum, p 487-494, 1990.

(2) S. GALLIOU, M. MOUREY, "Dual-mode temperature controlled quartz resonator", 3th European Frequency and Time Forum, p 147-151, 1989.

(3) BALL-EFRATOM Division - Technical notice : "ECXO series sub-miniature", sept. 1989.

(4) H.M. GREENHOUSE, R.L. Mc GILL and D.P. CLARK, "A fast warm-up quartz crystal oscillator", 27th AFCS, p 199-217, 1973.

(5) H.W. JACKSON, "Update on the tactical miniature crystal oscillator program", 36th AFCS, p 492-498, 1982.

Session 8A Number 73

A NEW WAY TO CONTROL PHASELOCKED FREQUENCY STANDARDS

Mannermaa Jari & Kalliomäki Kalevi

University of Oulu, Dept. of Electrical Eng.

90570 Oulu, Finland

Abstract

A conventional digital PI -type controller of the phaselocked crystal or rubidium frequency standard and an adaptive maximum entropy (MEM) controller are studied both theoretically and in practice. MEM and digital PI-type controllers are simulated to obtain the phase and frequency stabilities of a typical phaselocked frequency standard. The signal used in the theoretical tests is a phase ramp. The simulations point out that the MEM controller is faster and at least three times more effective than the PI controller. Practical tests with a phaselocked Loran C receiver confirm the results. The noise power ratio of the phase of the PI controller is at least four times that of the MEM, and the respective ratio for frequency is approximately from one to four. The system itself is known to be a low-pass filter for noise in the Loran C signal and a high-pass filter for that of the receiver crystal.

Introduction

Phaselocked crystals or rubidium frequency standards are usually controlled by conventional PI or PII -type, analogue or digital controllers, but the possibilities of using the Maximum Entropy Method (MEM) in many applications have been under investigation at the Electronic Measurements Lab. of the University of Oulu for a number of years. One application examined is the use of an adaptive MEM controller for this purpose. MEM and digital PI-type controllers have been simulated to study the phase and frequency stabilities of a typical phaselocked frequency standard. The test signal was a phase ramp with a slope of 25 ns/h saturating to a level of 1 microsecond, corresponding to a double jump of $7*10^{-12}$ in the frequency domain. The superimposed noise level was about 30 ns. The performance criterion was a product of the Allan variance of frequency and the time for which the frequency error lies outside the limits (in our case > 10^{-12}). A crystal phaselocked by means of Loran C signal was employed and controlled in the practical experiments. The frequency was determined from the phase. The criterion was the ratio of noise levels of the phase and frequency of these two controllers. Step responses were also obtained for the MEM controller.

Method

The principal properities of the whole systems, especially the sources of noise, were first mapped. The system itself is described in the form of a block diagram in Fig. 1.

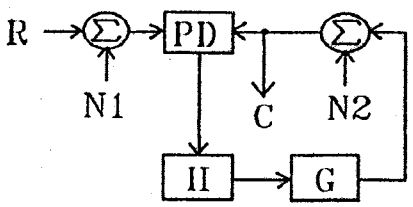


Fig 1. Block diagram of the whole system, in which C is the phase, G the crystal, H the controller, PD the phase detector, R the phase reference input, N1 the Loran C noise and N2 the noise of the oscillator.

Using the notations of Fig. 1, we can derive

$$C = \frac{HG(R+N1)}{1+HG} + \frac{N2}{1+HG} \qquad (1)$$

where the variables can be either Laplace or z -transformed. If R = 0 and N2 = 0, then

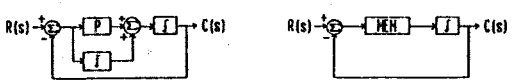
$$\frac{C}{N1} = \frac{HG}{1+HG} \qquad (2)$$

and if R = 0 and N1 = 0, then

$$\frac{C}{N2} = \frac{1}{1+HG} \qquad (3)$$

The noise processes of the system can described exactly enough by these simplified equations.

MEM and digital PI-type controllers were simulated to study the phase and frequency stabilities of a typical phaselocked frequency standard. The block diagrams of the controllers used are presented in Figure 2.



a) b)

Figure 2. Block diagrams of the control systems used: a) conventional PI -controller, and b) MEM controller.

The use of MEM in control systems is based on its ability to predict new values in the process, since it is a method for calculating coefficients for the autoregressive process having N recorded and known values:

$$y(n) = a(1)x(n-1)+a(2)x(n-2)+\ldots+a(M)x(n-M) \quad (4)$$

where

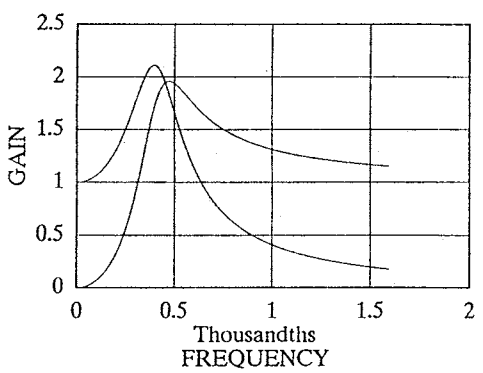
y(n) is a predicted value of the known time series,
x(i)s are measured values of the time series,
a(j)s are coefficients calculated by the MEM,
M is named as the degree of the autoregressive model or the length of the prediction error filter.

It must be noted that many succesive predictions can be calculated ahead based on the same initial time series and the already predicted values. This is done by inserting the last prediction as the first value in the time series and dropping the last (oldest) one off, since otherwise the time series would easily become too long. After N predictions the whole time series naturally contains only predicted values.

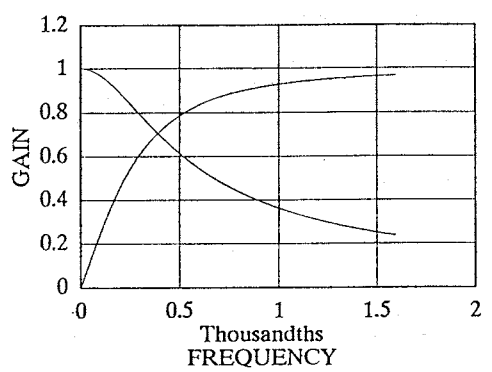
The advance achieved by using the MEM is that process is not controlled by the error signal of the moment of measurement but by a predicted one calculated from the measured ones. Thus adaptive controllers of this kind remain the desired number of steps ahead in approximating the state of the system as compared with conventional ones.

Results

The noise processes of the whole systems are approximated in Fig. 3, in which the frequency responses of a PI (a) and a MEM controller (b) are presented.



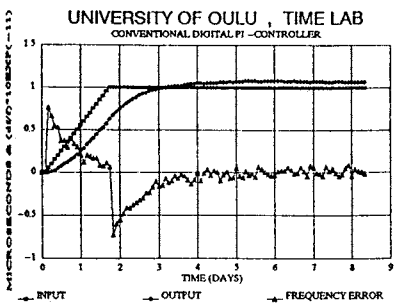
a)



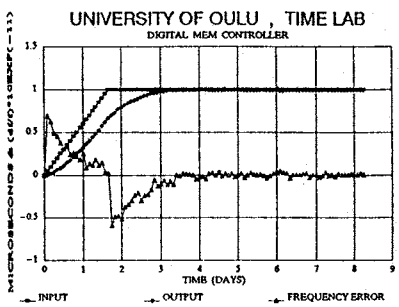
b)

Fig. 3. Frequency responses of the whole system when using a PI (a) or a MEM (b) controller with parameters: N=5, M=2, a_1=0.3 and a_2=0.5.

Some theoretical results are shown in Figure 4. The result for the PI -type controller is the best when determined experimental, but that for the MEM controller is not optimized. As can be seen from the diagrams, the MEM controller is faster than the PI controller.



a)



b)

Figure 4. Results obtained with the controllers given a simulated phaselocked frequency standard: a) conventional PI controller and, b) MEM controller.

It should be added that a PID -type controller was also tested, the D (derivative) -element being in the feedback branch (as a tachometer), but this system gave poorer results than without the D element.

The following tables are presented some results obtained in January and February 1991 in tests employing a crystal in a Loran C receiver and a Loran signal:

Table 1.
Test results for the PI and MEM controllers. N and M are the parameters of the MEM and TC the control interval. The measuring interval is 2 hours. The values of "Phase" and "Frequency" are calculated by using the formula of the standard deviation.

Controller	Time	Phase ns	Frequency ns/d	N M TC
PI	25.-31.12	171	1926	652
MEM	09.-14.01	16	268	5 2 1200
PI	26.-29.01	84	1337	652
MEM	30.-31.01	51	499	5 2 1200

Table 2.
Test results for the PI and MEM controllers. N and M are the parameters of the MEM and TC the control interval. The measuring interval is 10 minutes. The values of "Phase" and "Frequency" are calculated by using the formula of the standard deviation.

Controller	Time	Phase ns	Frequency ns/d	N	M	TC
MEM	14.-16.01	92	6711	5	2	1200
PI	17.01	765	7981			652
MEM	18.01	89	4268	5	2	1200
PI	19.-21.01	379	5931			652
MEM	22.-23.01	53	4507	5	2	1200
PI	24.-25.01	119	2833			652
PI	26.-29.01	101	1859			326
MEM	30.-31.01	53	3240	5	2	326
MEM	01.-02.02	53	2700	5	2	326
MEM	03.-04.02	46	2697	5	2	120
MEM	05.-06.02	40	1280	5	2	90
MEM	09.02	31	1538	7	1	326
MEM	12.-13.02	33	2082	3	1	326
MEM	16.-17.02	28	616	5	2	326
MEM	18.02	19	1248	15	2	326
MEM	19.02	19	1289	5	2	326
MEM	20.-21.02	25	1517	5	1	326
MEM	22.02	32	1163	5	2	326
MEM	24.-27.02	28	1130	5	2	326

A typical MEM step response of the phase difference between the signal of a crystal in a Loran C receiver and a Loran signal is presented in Fig. 5. The step was produced by changing the reference input value of R in the whole system of Fig. 1.

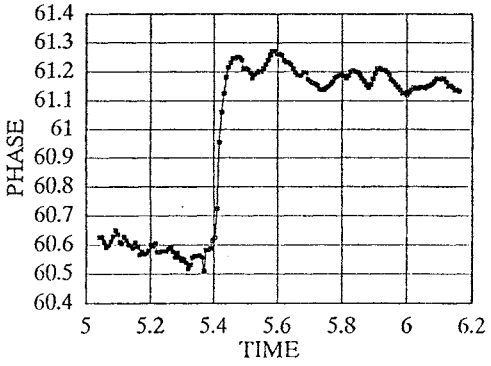


Fig. 5. A typical MEM step response for the system decribed in Fig. 1.

Discussion

The MEM controller has a relatively good frequency response for the whole system because there are no overshoots, whereas the PI controller has very high ones. Moreover, the purpose is to reduce the intersection between the curves for the low and high-pass filters as far as possible, since the control time can then be selected to achieve the most effective noise damping.

The MEM controller was assumed to be about four times as effective as the PII one after the theoretical tests, and this was proved in the practical tests in the case of phase stability. As regards frequency stability, its superiority depends on the conditions, but the MEM is frequently at least as good or better than the PI. Use of a PID -type controller with a D element in the feedback branch (as a tachometer) did not improve the performance in this kind of process.

There is one problem entailed in using the MEM controller, however, namely how to optimize its performance. For example, the shorter the length of the prediction error filter, the better is the filtering and the slower the controller is in reacting to changes in the state of the system. If the time series is long, the controller will be slow, and if M is too large the control system will suffer from vibration problems. In fact, a good performance is nearly always achieved by using N = 5 and M = 2.

The performance was improved markedly in the case of frequency stability if one tried to achieve only one half or a fourth part of the predicted error during one control cycle, for instance.

SESSION 7B

ATOMIC CLOCKS III

Chairman : A. BAUCH

Session 7B Number 74

Recent Progress in the Development of the EFOS-B Hydrogen Maser at Neuchâtel Observatory

Laurent-Guy BERNIER and Giovanni BUSCA

Neuchâtel Observatory, Switzerland.

Abstract

An advanced hydrogen maser with state of the art characteristics and a superior reliability is being developed at Neuchâtel Observatory for the European Space Agency.

The new features of the EFOS-B maser, called after the EFOS masers of the previous generation, include a new vacuum system, with an improved reliability and pumping speed, a microprocessor monitoring system with standard serial interface, a 100 MHz clean-up loop, an automatic cavity tuning system based on the auxiliary mode interrogation (AMI) principle and an automatic source tuning system.

Progress in the development of the above features as well as an improved wall coating and thermal coefficient are reported.

Introduction

In many current and future ESA projects, time and frequency imperatives require various high stability reference oscillators. The choice between quartz, rubidium, cesium and hydrogen frequency standards depends upon the short, medium and long-term stability requirements as well as upon environmental and size, weight, power considerations. The EFOS-B hydrogen maser, which is being developed for ESTEC at Neuchâtel Observatory, is a full-size active hydrogen maser. The EFOS-B design follows closely the original EFOS design of the previous generation. The new developments were oriented toward state of the art performances, high reliability and ease of maintenance.

Overview of Mechanical Design

The maser has 2 main parts placed side by side on a common base frame: the physics package and the electronics rack. The ion pumps and the valves of the vacuum system are placed on the bottom of the electronics rack. This arrangement makes unnecessary a magnetic shielding of the ion pumps, because they are far enough from the physics package magnetic shields, and, on the other hand, it leaves the bottom of the physics package almost empty which makes the access to the hydrogen dissociator very easy. The complete hydrogen distribution system is mounted on a base plate placed on the bottom of the physics package. In order to further facilitate the access to the hydrogen dissociator, the whole hydrogen distribution system can be taken out by removing 2 screws and disconnecting a single VCR connector.

There are two ion pumps in parallel for the pumping of hydrogen from the internal vacuum enclosure and a third ion pump for the pumping of the outgassing from the external vacuum enclosure. The function of the external vacuum is thermal isolation of the aluminium microwave cavity as in the original EFOS design. The use of 2 internal ion pumps in parallel double both the pumping speed and the ion pump lifetime. The reliability is also improved since a single pump is sufficient for the maser operation. The access to the ion pumps and to the vacuum valves is particularly easy since the front and side panels of the electronics rack are removable in a matter of seconds.

Wall Coating

An improved Teflon coating technique was experimented with the last few EFOS production masers.

	EFOS-5	EFOS-9
$Q_{line}(p)$	2.48×10^9	3.06×10^9
$Q_{line}(p=0)$	3.41×10^9	3.50×10^9
$Q_{line}(Teflon)$	7.50×10^9	7.96×10^9

Table 1
Atomic line Q on EFOS-5 and EFOS-9

The result is an atomic line quality factor that is both high and stable. Table 1 shows the atomic line Q of EFOS-5 and EFOS-9 masers. $Q_{line}(p)$ is the operating line Q measured at the operating pressure p, $Q_{line}(p=0)$ is the line Q extrapolated to zero hydrogen pressure, i.e. without spin exchange broadening, and $Q_{line}(Teflon)$ is the contribution of the wall coating to the extrapolated line Q, computed according to

$$\frac{1}{Q_{line}(p=0)} = \frac{1}{Q_{line}(Teflon)} + \frac{1}{Q_{line}(bulb)} \qquad (1)$$

and

$$Q_{line}(bulb) = \nu_0 \, \pi \, T_b \qquad (2)$$

where ν_0 is the hydrogen frequency and T_b the storage time constant of the bulb as determined by the storage bulb volume and the collimator dimensions [1] (equation 3B.48). The storage time constant determined by our standard EFOS collimator is 1.4 s, i.e. $Q_{line}(bulb) = 6.25 \times 10^9$, which is not optimal for the very high quality factor of the wall coating. Therefore the operating line Q could still be improved for the same power output of the maser by changing the collimator.

EFOS-5 has been operating with the new coating since May 1990 and EFOS-9 since September 1990 without any noticeable degradation of the quality factor. Between the first oscillation

and delivery of both masers the line Q was carefully monitored and there was no measurable initial degradation of the quality factor. After delivery to the radiotelescopes where the masers are used, a precise monitoring of the line Q was no longer possible. However the amplitude of the maser signal was monitored ever since and does not show any measurable decay.

Hydrogen Dissociator

The new hydrogen dissociator design completely eliminates the use of glue or Viton o-rings. The dissociator bulb is directly sealed to the metal mounting flask and the capillary hydrogen input of the bulb is connected to a Cajon VCR connection through a glass-metal transition. The reason of this is to eliminate the possibility of pollution of the discharge by organic molecules that can limit the lifetime of the dissociator through a process of carbon deposition on the inner bulb surface.

Vacuum System

Hydrogen is pumped out of the internal vacuum enclosure using two 45 l/s ion pumps in parallel. The flask of the 45 l/s ion pump has a CF63 flask as compared to the CF35 flask of the single 30 l/s ion pump used in recent EFOS masers (older EFOS masers have 20 l/s pumps). Through the use of the bigger CF63 tubing wherever possible, combined with the 2 ion pumps in parallel, the EFOS-B design achieves a 47 l/s pumping speed as compared to the 15 l/s of the original EFOS design. This is almost a 3 fold improvement which yields a lower background pressure inside the maser and reduces the risk of long-term ageing of the storage bulb or dissociator by interaction with molecules other than hydrogen.

The single 20 l/s pump of old EFOS masers, which is not specified for hydrogen, has a mean lifetime of about 2 years. The single 30 l/s of recent EFOS masers and the 45 l/s pumps of EFOS-B are both specified for hydrogen pumping and have the same capacity. The 30 l/s single pump that has been running for the longest time is the one now in EFOS-6 and has accumulated 34 months of continuous operation. No failure has been recorded yet for this type of pump.

Hydrogen Distribution System

The new hydrogen distribution system is mounted on a removable base plate as described above. The hydrogen bottle is connected to a new pressure regulator which is more reliable than the original EFOS model. All connections from the output of the pressure regulator to the pirani gauge, the palladium valve and the hydrogen dissociator capillary input are made using stainless steel rigid and flexible tubing and Cajon VCR vacuum connectors with copper joints.

Both the pirani gauge and the palladium valve are of a new simpler and more reliable design.

Thermal Coefficient

The thermal coefficient of the recently built EFOS masers is better than $1x10^{-14}/C°$. As a matter of fact, these recent masers are so superior to our old reference maser than a precise measurement was not possible. Figure 1 shows the frequency transient produced by a 10 C° ambient temperature step.

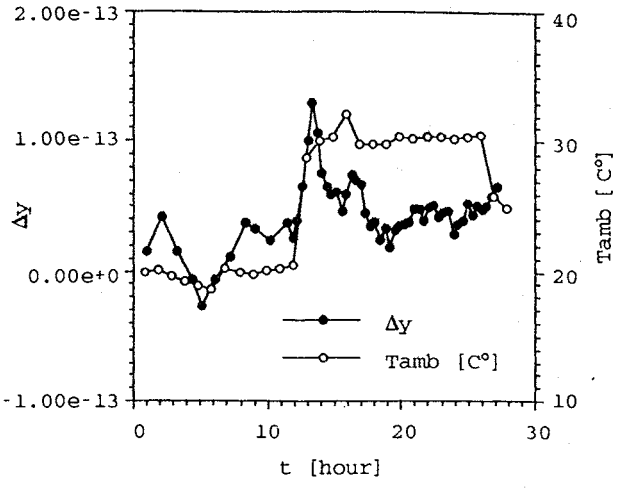


Figure 1
Recording of EFOS-9 Thermal Transient

Figure 1 shows the temperature in EFOS-9's room and the normalized frequency difference between EFOS-9 and our reference maser H1. There is a transient frequency step of about $1x10^{-13}$ during the 10 C° temperature step and then the frequency goes back to its original trend which is dominated by the thermal fluctuations of H1. H1 temperature coefficient is about $1x10^{-13}$ and the room where it stands is temperature stabilized to a few tenth of 1 C°. From this measurement we conclude that the dynamic temperature coefficient of EFOS-9, i.e. the transient frequency step during a fast temperature step, is about $1x10^{-14}/C°$ and its static thermal coefficient is sensibly less than $1x10^{-14}/C°$.

Monitoring System

The new monitoring system has 34 analog channels that can be read through a standard RS232 serial interface. In addition, the synthesizer can be both read and set through the serial interface. The external monitoring terminal is a portable compatible PC that can both display the readings on the screen and record them on a floppy disc. The monitoring can be performed remotely through a telephone line if a modem is installed inside the portable PC.

Automatic Cavity Tuning

The automatic cavity tuning is of the auxiliary mode interrogation type (AMI) [2]. Its principle is as follows. Instead of using the TE011 mode for the cavity tuning, at the risk of perturbating the atomic signal frequency, an auxiliary mode is used. The frequency stabilization of the auxiliary mode has the same effect as the stabilization of the TE011 mode if the ratio between the modes resonant frequencies is constant.

The TE013 mode was selected as the auxiliary mode because its temperature coefficient is close to the temperature coefficient of the TE011 mode.

We have installed a prototype AMI-ACT on our reference maser H1. In closed loop, the error signal of the AMI-ACT system is added at the output of the thermistor bridge that controls the cavity temperature in the unmodified H1 electronics system. In this way the maser can run normally with the ACT loop open. When the ACT loop is closed, the ACT error signal adds a fine correction signal to the thermistor bridge normal output.

A 0,0447 C° temperature step in the thermistor bridge, i.e. a 1400 Hz step of the TE011 mode frequency, produces an actual frequency step on the TE013 mode of 2 Hz with the ACT in closed loop. This corresponds to a gain of 700 in closed loop.

The true performance of the AMI-ACT system, i.e. its action on the long-term stability of the cavity temperature, depends upon the sensibility to ambient temperature of the ACT electronics. In the H1 prototype, the ACT electronics are placed outside the maser and are subject to the room temperature fluctuations. The next step is to develop a miniaturized ACT electronics set that can be fitted inside the temperature stabilized maser head. This work is now in progress at ON.

Automatic Source Tuning

An automatic tuning system was developed for the EFOS-B hydrogen dissociator. The need of a variable frequency generator for driving the dissociator plasma discharge comes from 2 contradictory frequency requirements. On the one hand the frequency at which the discharge can self-start is determined by the resonant frequency of the dissociator RF electrodes without a discharge. On the other hand the resonant frequency of the dissociator with the discharge on is a function of the dissociator electrodes, of the RF power and of the hydrogen pressure in the dissociator.

In our design the RF generator is initially frequency swept until the plasma discharge lights on. Then an automatic control loop is switched on that continuously tunes the generator to the match frequency, i.e. to the frequency that yields the maximum power transfer to the dissociator discharge. In addition there is an AGC control loop that stabilizes the power transmitted to the discharge.

100 MHz Clean-up Loop

A 100 MHz VCXO clean-up loop was developed in order to provide a 100 MHz output with a better equivalent spectral purity than the standard 5 MHz output controlled by the BVA 8600 master OCXO.

The white phase noise level of the 100 MHz output is -160 dB REF 1 [rad^2/Hz] as compared to the -121 dB of the 5 MHz BVA multiplied up to 100 MHz.

Conclusion

The developments of the new EFOS-B features described in this progress report are still in the course of realization. The first 2 EFOS-B masers will be ready this year. The main results reported are the high quality factor of the new Teflon coating and the improved temperature coefficient.

References

[1] Vanier J., Audoin C., *The Quantum Physics of Atomic Frequency Standards*, Adam Hilger, 1989.

[2] Busca G., Johnson L., "A New Automatic Cavity Tuning for Active H Masers," *Proc. 1st EFTF*, Besançon, March 1987, pp.339-341.

Acknowledgements

The development and the production of 2 EFOS-B masers is under way at ON under a contract with ESTEC.

We would like to thank M. Kari Kuittinen, from the Metsähovintie Radio Research Station, Finland, and M. Roberto Ambrosini from the Medicina Radio Observatory, Italy, for providing the data on the field operation of EFOS-9 and EFOS-5.

Session 7B Number 75

FAST CAVITY AUTO-TUNING SYSTEM FOR H-MASERS : EXPERIMENTAL RESULTS.

R. BARILLET, J. VIENNET, P. PETIT, C. AUDOIN

Laboratoire de l'Horloge Atomique
Unité Propre du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud - 91405 ORSAY Cedex - France

ABSTRACT

Fast cavity auto-tuners have been implemented on two H-masers at L.H.A. They are transmission and reflection operated, respectively. The time constants are of the order of 10 s.

Experimental results show that, when the auto-tuners are operated, the short term frequency stability is similar to that obtained with the free running cavities. In the medium term, the frequency stability is improved $\sigma_y(\tau = 10^4\ s) \sim 1 \times 10^{-15}$. In the long term, the comparison with a high performance caesium clock -HP 5061A option 04- shows that the frequency drift is extremely low, of the order of 1×10^{-15}/day or less, instead of 1×10^{-14}/day when the auto-tuner is not operated. After the thermal regulation of the maser has been turned off, then on, the retrace due to the cavity tuning -in the auto-tuning mode- is about 1×10^{-14}. The frequency stability is better than $3.10^{-15}(\tau \sim 10^3\ s)$ 2.5 days after the turn on, instead of 3 weeks when the cavity is free running. The final value of the frequency is acquired, within 10^{-13}, 3 days after the turn-on.

INTRODUCTION

In the H-masers, the well known cavity pulling effect limits the long term frequency stability. Various auto-tuning methods have been proposed and developed (1, 2,3,4,...) to overcome this limitation. The first one is founded on the observation of the oscillation frequency changes when the atomic line Q is spin-exchange modulated. In the second one, it is the cavity Q which is modulated. The third method is based on the observation of the amplitude of oscillation when the resonant frequency of the cavity is modulated. The fourth one is the fast cavity auto-tuning method, which is similar to the method used on passively operated H-masers. It seems that this fast auto-tuning method provides significant advantages : for instance, it does not require a second H-maser, it does not reduce the level of oscillation, it can be implemented on any kind of H-maser and it enables a short time-constant of the cavity control. In a preliminary experiment, we had successfully checked the efficiency of the fast cavity auto-tuning method (5)(6). Recently, two fast cavity auto-tuners have been implemented at L.H.A. They are respectively transmission and reflection operated.

1. PRINCIPLE OF THE METHOD

It is described in figure 1. A frequency modulated microwave signal, called the interrogation signal, feeds the cavity.

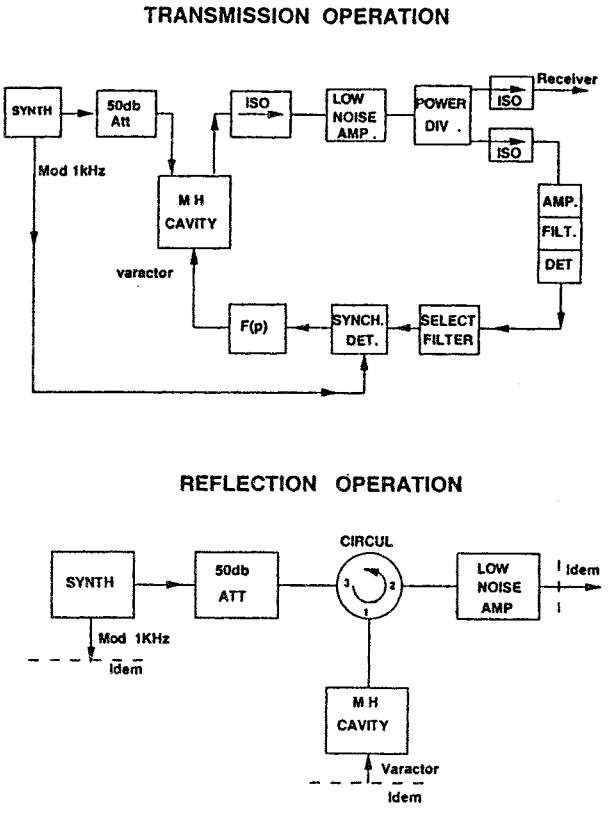


FIG 1 : PRINCIPLE OF THE METHOD

It is transmitted through (reflected by) the cavity. The frequency is square wave modulated. The condition of carrier suppression is fulfilled (4) i.e. the modulation depth is an even multiple of the modulation frequency. The amplitude of the transmitted (reflected) signal is modulated, and an error signal is provided by a synchronous detector when the resonant frequency of the cavity differs from the mean frequency of the interrogation signal. The error signal is processed by an operational filter -including an integrator- and the output of this filter is applied to the varactor diode which controls the resonant frequency of the cavity. In those conditions, the resonant frequency of the cavity is locked to the mean frequency of the interrogation signal.

1.1. General properties

This auto-tuning method shows several interesting properties. Firstly, it is possible to adjust the mean frequency of the interrogation signal in order to compensate the spin exchange frequency shift [4]. Therefore, the oscillation frequency will not depend on the atomic density in the storage bulb. Secondly, spurious frequency shifts induced by the method are quite negligible. Their relative value is smaller than 1×10^{-15}. Thirdly, the method shows practical advantages. After the initial calibration of the mean interrogation frequency, it does not require another H-maser. The oscillating level of the maser is not reduced. The method can be used on any kind of H-maser having a varactor tuned cavity. The time constant of the cavity servo can be short ($\sim$ 1 s), which is interesting in the case of disturbed environmental conditions.

1.2. Behavior in the reflection mode of operation

In the first papers [4],[5],[6], it has been shown theoretically and checked experimentally that the short term frequency stability is not altered in the transmission mode of operation provided that the power of the interrogation signal injected into the cavity is much larger than 10 P_o, where P_o is the power delivered by the atoms to the cavity.

In the reflection mode of operation, the behavior is a little bit different in this respect.

a) optimum depth of modulation

Let a_r be the amplitude of the signal reflected by the cavity and ρ the cavity reflection coefficient. The optimum depth of modulation is obtained at the maximum slope of $\left|\frac{1}{a_r}\frac{\partial a_r}{\partial \omega}\right|$, i.e. at the maximum slope of $\left|\frac{1}{\rho}\cdot\frac{\partial \rho}{\partial \omega}\right|$. We use the classical RLC radio-electrical model of a cavity, with inductive coupling. The impedance of the output loop of the cavity may then be written as :

$$Z = R(1 + jQ_oX)$$

where Q_o is the unloaded quality factor of the cavity, with $X = \frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}$ where ω_o is the resonant angular frequency of the cavity, and ω the angular frequency of the signal. Let R_c be the characteristic impedance of the transmission line. Thus we have :

$$|\rho| = \left|\frac{R(1 + jQ_oX) - R_c}{R(1 + jQ_oX) + R_c}\right|$$

By using the classical parameters β, the coupling coefficient and, Q_c, the Q of the loaded cavity, one has :

$$|\rho| = \left|\frac{\alpha + jQ_cX}{1 + jQ_cX}\right| \quad , \quad \text{with } \alpha = \frac{1 - \beta}{1 + \beta}$$

The frequency of the considered signal is very close to the cavity resonant frequency. X may then be written as :

$$X = \frac{\omega}{\omega_o} - \frac{\omega_o}{\omega} \cong \frac{2\Delta\omega}{\omega_o} \quad , \quad \text{with } \Delta\omega = \omega - \omega_o$$

In those conditions, the modulus of the reflection coefficient is :

$$|\rho| = \left|\frac{\alpha + 2jQ_c\frac{\Delta\omega}{\omega_o}}{1 + 2jQ_c\frac{\Delta\omega}{\omega_o}}\right| = \left|\frac{\alpha + ju}{1 + ju}\right|$$

where $u = 2Q_c\frac{\Delta\omega}{\omega_o}$, and we have :

$$\left|\frac{1}{\rho}\frac{\partial \rho}{\partial u}\right| = \frac{u}{\alpha^2 + u^2} - \frac{u}{1 + u^2} .$$

The maximum value of $\left|\frac{1}{\rho}\frac{\partial \rho}{\partial u}\right|$ is obtained for :

$$u^2_{opt} = \frac{-(1 + \alpha^2) + \sqrt{(1 + \alpha^2)^2 + 12\alpha^2}}{6}$$

In our cavities, β is of the order of 0.2 so that we have $\alpha \cong 2/3$ ans $u^2_{opt} = 0.213$, which gives the following optimum depth of modulation :

$$\Delta f_{opt} = \frac{f_o\, u_{opt}}{2Q_c} \cong 0{,}46\,\frac{B}{2}$$

where B is the linewidth of the loaded cavity. This leads to $\Delta f_{opt} \cong 8$ kHz. This depth of modulation is about half the optimum value in the transmission mode.

b) Effect of amplitude noise

It is related to the maximum value of $\left|\frac{1}{a_r}\frac{\partial a_r}{\partial \omega}\right|$. The comparison with the transmission mode is very simple , via the comparison of $\left|\frac{1}{a}\frac{\partial a}{\partial u}\right|$ in both configurations (a is the amplitude of either the transmitted or the reflected signal).

For the transmission mode, the optimum value of $\left|\frac{1}{a}\frac{\partial a}{\partial u}\right|$ is 0.5. For the reflection mode, it depends on β, as shown in the next table.

β	0.10	0.15	0.20	0.25	0.30
u^2_{opt}	0.270	0.241	0.213	0.187	0.163
$\left\|\frac{1}{a}\frac{\partial a}{\partial u}\right\|$	0.144	0.228	0.322	0.426	0.543

In our cavities, β is of the order of 0.2-0.25 and $|\partial a/a\,\partial u|$ is about 0.35. Thus, the effect of the amplitude noise will be about 3 dB larger than in the transmission mode.

For our operating conditions, the effect of the additive noise increases the frequency instability by 5 % in the transmission mode of operation. Thus, the corresponding effect will be larger, amounting to 10 %, in the reflection mode.

2. EXPERIMENTAL SET-UP

Our H-masers M3 and MH3 have been completed with auto-tuners. M3 is reflection operated and MH3 is transmission operated, which enables to test the properties of each operating mode.

2.1. Synthesis of the interrogation signal

The interrogation signal is generated in two steps, as shown in Fig. 2

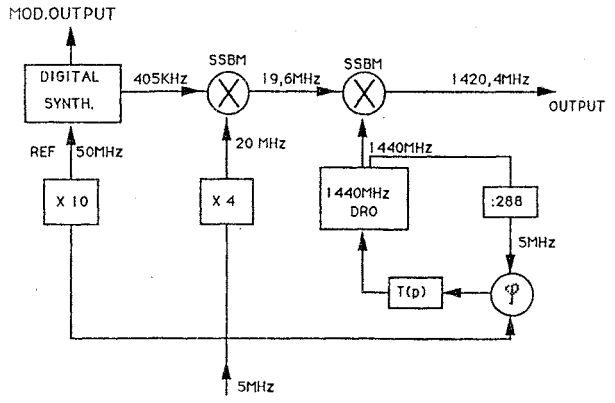


Fig 2: SYNTHESIS OF THE INTERROGATION SIGNAL

A 50 MHz signal, delivered by a frequency multiplier, is the frequency reference of the digital synthesizer. The principle of the synthesis [7] is phase accumulation. The synthesizer generates a low frequency signal at 405 kHz, with the desired square wave frequency modulation [4]. The depth of modulation is about 9 kHz in the reflection mode and 18 kHz in the transmission mode. The modulation frequencies are close to 1 kHz.

The frequency modulated 405 kHz signal is transferred at 1420.405 MHz by two local oscillators (20 MHz and 1440 MHz) and single side band mixers. The interrogation spectrum, observed at 1420.405 MHz, is identical to the theoretical one. The measurement of the carrier rejection, which is limited by the analyzer noise, shows that it is larger than 65 dB referred to the most powerful lines of the interrogation spectrum.

2.2. Processing of the cavity response signal

The interrogation signal is injected towards the cavity by way of several SMA attenuators which enable to adjust the interrogation signal at the desired level, such as $P_c \cong 100\ P_o$. Then the signal transmitted (or reflected) by the cavity is amplified at 1.42 GHz. At the output of the low noise amplifier, a part of the power is directed towards the receiver which phase locks the 5 MHz crystal oscillator to the maser. The operation of the receiver is not disturbed by the interrogation signal thanks to i) a dynamics margin of 20 dB and ii) a narrow bandwidth at 5.75 kHz, which rejects the spectral components of the cavity response signal.

Let us note, too, that the relatively narrow receiver bandwidths at 19.6 MHz and 405 kHz do not enable to process the cavity response signal inside the receiver, and this is the reason why we amplify and detect the cavity response at 1.42 GHz.

At the output of the low noise amplifier, the second part of the power is directed towards a high gain 1.42 GHz amplifier. The global gain of both amplifiers is about 80 dB. A 20 dB isolator is included between the 1.42 GHz amplifiers, to prevent the system from oscillating. The output of the second amplifier is loaded by an isolator whose output is connected to the 1.42 GHz detector. The applied RF power is about - 10 dBm, and the resulting DC output voltage is about 150 mV. The detector output shows a residual level of - 40 dBm at 1.42 GHz. We have suppressed it by a HF supply filter and an additive HF filter at the input of the low frequency processing set-up. So, we prevent the system from injecting a RF feedback signal inside the cavity via the varactor loop, which could induce spurious effects.

The low frequency signal processing includes i) a band pass filter centered at the modulation frequency close to 1 kHz, showing a 30 dB gain, ii) a synchronous demodulator which generates the servo loop error signal and iii) an operational filter including an integrator, a DC voltage shifter and two outputs, one follower and one inverter, that are suitable either for the transmission or the reflection modes of operation.

The output of the operational filter is then connected to the cavity varactor diode.

For optimum behavior, the mean frequency of the interrogation signal is adjusted to provide the same maser oscillation frequency as with the spin-exchange frequency tuning method. We have observed that the mean frequency of the interrogation signal differs from the maser frequency by as much as 160 Hz and 600 Hz, for the reflection and the transmission modes of operation respectively. This is larger than the theoretical value of the cavity mistuning, about 50 Hz, which is necessary to compensate the spin exchange frequency shift. The probable origin of this effect is an asymmetry of the cavity response pattern.

3. EXPERIMENTAL RESULTS

3.1. Effect of impedance fluctuations

Any change in the electrical characteristics of the elements connected to the resonant cavity may induce a change of the cavity resonant frequency, and thus of the maser oscillation frequency. For instance, we have observed that a temperature fluctuation of the output isolator yields a relative frequency variation of several $10^{-14}\ K^{-1}$. To test the efficiency of the auto-tuners towards this kind of effect, we have set a phase shifter between the input coupling loop and the injection attenuator on MH3.

We have observed that a length variation equal to $\lambda/2$ for the line induces a frequency fluctuation of

2.5 10^{-12} p.p. with the free running cavity.

Then the auto-tuner has been turned on. For the same length variation of the line, the frequency variation is smaller than 4. 10^{-14}. The initial spurious effect has been reduced by a factor of about 50, which shows that the effect of impedance variations is very well suppressed by auto-tuners.

3.2. Effect of a very large temperature variation of the MH3 cavity

Our masers have cavities which are very well thermally isolated, but with the consequence of a long warm-up time. For instance, a 3 weeks delay is necessary to reach a frequency stability of 3 x 10^{-15} for 1000 s, after turn-on [8]. We have tested the auto-tuner efficiency with respect to this characteristics.

In all the experiments reported, the auto-tuners were operated. After having set the oscillation frequency at the spin-exchange tuned value, we have turned the thermal regulation of maser MH3 off for three days. Then, the cavity temperature has decreased from 45°C to 22°C. The corresponding relative frequency change was + 4.4 x 10^{-11}, which is much larger than the expected change, of - 2.3 x 10^{-12}, due to the second order Doppler effect and to the wall shift temperature sensitivity. It is very likely that this discrepancy is related to a change in the asymmetry of the cavity resonance pattern.

Then, we have turned the thermal regulation on, and we have observed the MH3 oscillation frequency, which is shown in figure 3.

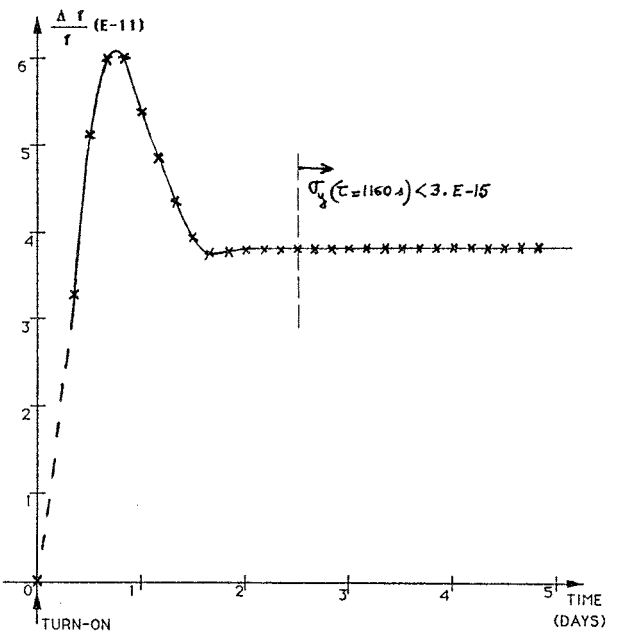


FIG 3: FREQUENCY BEHAVIOR AFTER THE TURN-ON (MH3).

The behavior is much disturbed for 2 days, because of probable changes in the asymmetry of the cavity resonance pattern. After 2.5 days, the frequency stability is better than 3 x 10^{-15} for τ = 1100 s. This represents a very significant improvement with respect to the case of a free running cavity. The final value of the maser frequency is reached, within 1 x 10^{-13}, three days after the turn-on. One week later, we have verified that the cavity resonant frequency was identical to its initial value within 0.5 Hz. Thus, the retrace due to the cavity auto-tuning is 1 x 10^{-14}.

3.3. Frequency stability in the auto-tuning mode

A beat note at approximately one hertz has been obtained from the 5 MHz outputs of the two masers. The measurement bandwidth was 10 Hz. The frequency stability results are shown in figure 4.

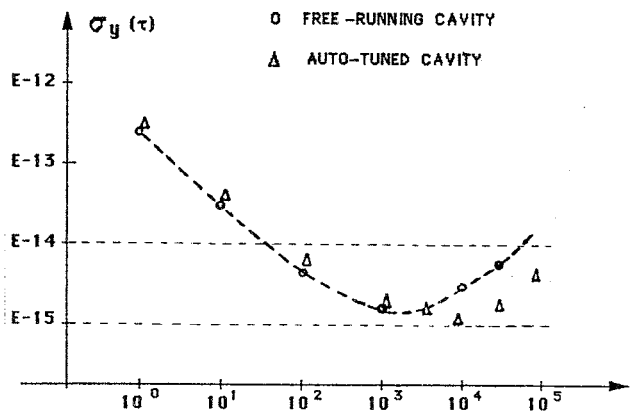


FIG 4 : FREQUENCY STABILITY

In the short term, the frequency stability is very close to that obtained when the auto-tuners are not operated. In the medium term, for τ > 3500 s, the improvement is obvious. For τ = 30 000 s and 1 day we observe the residual effect of daily temperature fluctuations, much reduced with respect to the free running cavity mode. Further improvements may be achieved because i) the auto-tuning electronics were not thermally regulated and ii) the M3 auto-tuner, which was operated in the reflection mode, showed an excessive thermal sensitivity of 2 - 3 10^{-14} K^{-1}.

3.4. Comparison M3/Cesium clock

With a free running cavity, the M3 frequency drift had been previously measured, and was equal to 1x10^{-14}/ day. We have compared M3, in the auto-tuning mode, with a high performance cesium clock (Hewlett-Packard 5061A option 04) to attempt to determine the residual drift of the maser frequency. To reduce the effect of the temperature on the measured value of the drift, we

have taken long samples (several weeks) where the initial and final room temperatures were identical within ± 0.1°C. In that condition, a 6 weeks run shows that the daily drift between M3 and the cesium clock, which is determined by a linear regression, is equal to 9×10^{-16} with an uncertainty of the same order of magnitude. A more precise measurement of the maser frequency drift would need a comparison link with TAI and a larger measurement time. Nevertheless, that measurement shows that the initial free-running cavity frequency drift has been strongly reduced, which is very interesting for time-keeping applications.

CONCLUSION

The fast cavity auto-tuners implemented on our masers M3 and MH3 have significantly improved the warm-up behavior and the long term frequency stability. A frequency stability better than $3. 10^{-15}$ is now acquired 2.5 days after the turn-on, the frequency drift of the free running cavity is much reduced, and the medium term frequency stability is about 10^{-15} for 10^4 s. Using thermally regulated auto-tuners, and the transmission mode of cavity interrogation, one can estimate that the frequency stability will stay of the order of 1×10^{-15} for sampling times larger than one day.

Acknowledgements

The authors gratefully acknowledge the Direction des Recherches, Etudes et Techniques (DRET) who has supported this work.

REFERENCES

(1) L.J. Rueger,J. Inst. Electron. Telecomm. Eng. 27 p. 493 (1981)

(2) T.K. Tucker and G.J. Dick, "Performance of a Hydrogen Maser with Auto-Tuning utilizing Cavity Q modulation". Proceedings 41st A.F.C.S., pp. 87-90 (1987)

(3) H.E. Peters, "Design and Performance of New Hydrogen Masers Using Cavity Frequency Switching Servos" 38th A.F.C.S. (June 1984)

(4) C. Audoin, "Fast Cavity Auto-tuning Systems for Hydrogen Masers", Rev. Phys. Appl., Vol. 16, pp. 125-130 (1981) and
C. Audoin, Rev. Phys. Appl. Vol. 17, p. 273 (1982)

(5) N. Yahyabey, Thèse, "Accord automatique rapide de la cavité résonnante dans une horloge atomique à hydrogène". Université Paris-Sud - Orsay - 21 Octobre 1986

(6) N. Yahyabey, R. Barillet, J. Viennet et C. Audoin, "Accord automatique de la cavité résonnante d'horloges à hydrogène : résultats préliminaires". Proc. 1er Forum Européen Temps-Fréquence, Besançon (France), pp. 334-338 - 18-20 Mars 1987

(7) J. Viennet, "Synthétiseur digital de fréquence pour standards de fréquence". Proc. 3e Forum Européen Temps-Fréquene - Besançon (France), 21-23 Mars 1989

(8) R. Barillet, P. Petit, J. Viennet et C. Audoin, "Masers à hydrogène actifs et compacts". Bulletin du BNM, Spécial Temps Fréquence, n° 63-64, Vol. 2, pp. 54-59 - Janv. Avril 1986

Session 7B Number 76

PRELIMINARY INVESTIGATIONS WITH THE NIST OPTICALLY PUMPED PRIMARY FREQUENCY STANDARD

R. E. Drullinger, J. H. Shirley, D. J. Glaze, John Lowe, HoSeong Lee* and A. S. Zibrov†
National Institute of Standards and Technology,
Boulder Colorado, 80303 USA

Abstract

An optically pumped, cesium beam, primary frequency standard has been constructed at NIST. The atomic beam tube is essentially complete but the lasers and electronics remain areas of active research. We have used the system to demonstrate a solution to the problem of coherence trapping of population that could otherwise limit clock performance. We also report spectra taken on the system and draw conclusions about the beam tube construction and operation.

The Apparatus

An optically pumped, cesium beam, primary frequency standard has been constructed at NIST and is undergoing preliminary testing. The design goal is to have a nearly self evaluating, continuously operating clock with a short term stability characterized by $\sigma_y(\tau) \approx 4 \times 10^{-13} / \sqrt{\tau}$ and an overall accuracy of about one part in 10^{14}. The analysis of systematic errors and the resulting logic that led to the design of the atomic beam tube have been published previously[1]. The geometric features of the design are shown in Fig. 1 where only half of the tube is shown because of symmetry. The microwave cavity is about 155 cm long and is terminated with ends designed to minimize distributed-cavity, phase shift[2]. There are separate regions of optical state preparation and detection at each end of the system to allow for simultaneous counter-propagating atomic beams[3]. The C-field is axial and is generated by a solenoid which extends over the entire length of the standard. The atomic beam cross section is defined by the windows in the microwave cavity to be 3 mm in diameter.

Newly developed electronics have been built and are being tested. A new modulator scheme is described in these proceedings[4].

Two laser systems were assembled for these experiments. They were external cavity, grating feedback systems made after the style of reference 5 but using commercially available diodes. They were individually referenced to saturated absorption features generated in room temperature cells using a Doppler compensated probe scheme[6]. For the experiments reported here, only one laser was used to prepare the atoms by pumping on the F=4 → F'=4 transition. The light beam for optical detection was synthesized from the pump laser with an acousto-optic modulator to drive the F=4 → F'=5 transition.

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

* permanent address: Korea Standards Research Institute, Time and Frequency Laboratory, P.O.Box 3, Taedok Science Town, Taejon 305-606, Korea

† permanent address: P.N.Lebedev Physics Institute, Academy of Sciences of the USSR, Leninsky prospect 53, Moscow 117924, USSR

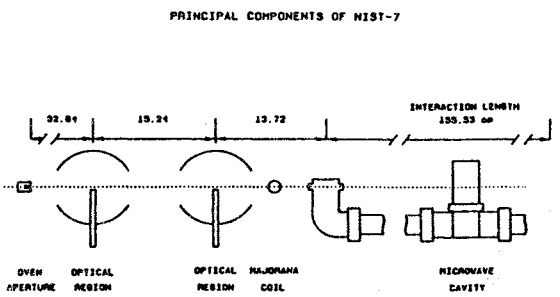


Fig. 1. Schematic representation of the atomic beam tube for NIST-7

The linewidth and frequency stability of the lasers were demonstrated with the following experiments. Both lasers were aligned in the state preparation region as for two-laser optical pumping[7]. One laser was locked to the F=4 → F'=4 transition and the other scanned across the F=3 → F'=4 transition. The dark resonances[8] were easily observed with about 25% contrast. Correspondingly, with one laser locked to the F=4 → F'=4 transition and the other locked to the F=3 → F'=4, the lasers were heterodyned together on a fast photodiode and the beat note recorded on a spectrum analyzer. The average of 500 sweeps is shown in Fig. 2. The frequency modulation necessary for the lock to the atoms was suppress in the beat note by synchronizing the modulation on the two lasers. The 3 dB width of the beat is about 500 kHz.

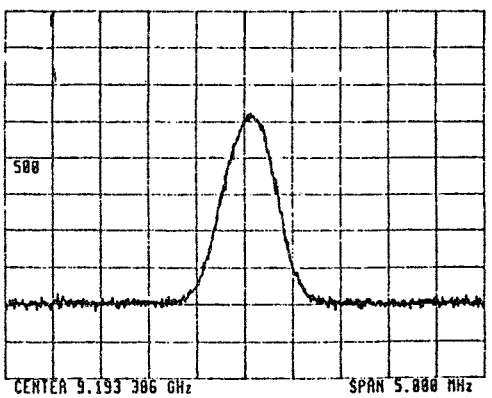


Fig. 2. Beat note between two lasers individually locked to saturated absorbtion features in cesium cell. One laser is locked to the F=4 → F'=4 transition, while the other is locked to the F=3 → F'=4 transition.

Coherence Trapping

When optical state preparation is attempted at low magnetic fields, coherent trapping of the population (nonlinear Hanle effect) may occur[9], leading to incomplete pumping. When linearly σ polarized light is used to excite a transition for which $F' \leq F$, there is a linear combination of Zeeman sublevels which is not coupled to the optical field. Such a state cannot be pumped and the population remains trapped. In a magnetic field, Larmor precession causes the trapped state to evolve into a different linear combination of Zeeman sublevels which is coupled to the optical field and hence can be pumped. By using a sufficiently high magnetic field in the pumping region, trapping can be avoided[10]. It is, however, still possible to achieve complete optical pumping in a low magnetic field in the following way:

If a linear polarization intermediate between σ and π is used, there will again be some linear combination state that is not coupled to the field. However, the exact combination depends on the direction of the polarization. Hence, by changing the polarization direction back and forth, atoms that are at first trapped in one polarization will be pumped by the second polarization while atoms trapped by the second polarization will be pumped by the first. For complete pumping the polarization must be changed several times during the transit time of the atoms across the pumping region. One can change the polarization either in time, as with an electro-optic modulator, or in space, as in polarization gradient cooling[11]. This latter scheme involves retro-reflecting the optical pumping beam back through the pumping region with a quarter-wave plate in front of the mirror.

Figure 3 shows a typical scan of the fluorescence yield from the optical pumping region as a function of magnetic field. Curve a shows the effect of coherence trapping where some of the population is not optically pumped at low magnetic field. Curve b shows the much greater degree of optical pumping achieved when the polarization gradient scheme is used.

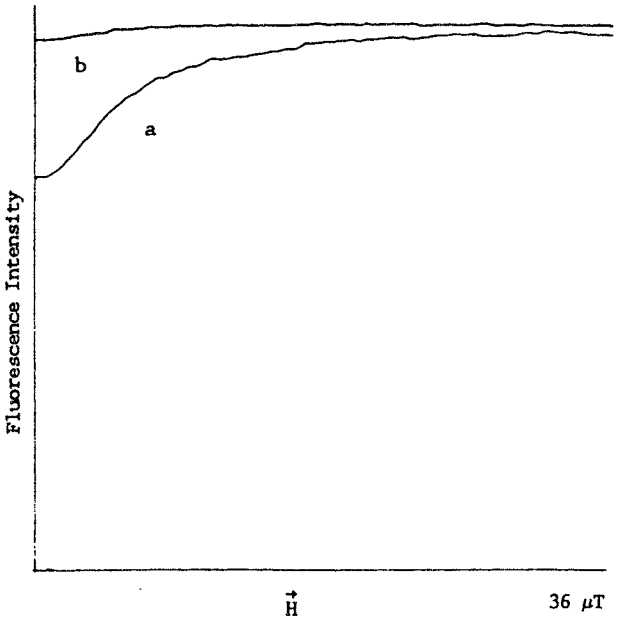


Fig. 3. Fluorescence from the optical pumping region as a function of magnetic field. The transition is the $F=3 \rightarrow F'=3$. Trace a is with linear σ polarization while trace b is made with counter-propagating, orthogonally polarized beams.

Spectra

To begin characterizing the performance of the atomic beam tube, we have recorded a number of spectra of the cesium hyperfine transition. Some of these are reproduced here as figures 4-6. The spectra were taken under the following conditions: The temperature of the oven was 90°C. The detection laser beam was about 4 mm in diameter, with a power of 35 μW and tuned to the $F=4 \rightarrow F'=5$ transition. This resulted in an average of about 20 scattered photons per atom in the $F=4$ state and essentially unit detection probability. The pump laser was tuned to the $F=4 \rightarrow F'=4$ level and had about 300 μW in a 4 mm diameter beam. Its polarization was established by the use of the polarization gradient pumping scheme outlined above.

Figure 4 shows the hyperfine spectrum excited by a microwave power a few decibels below optimum. It consists of the 7 expected $\Delta m=0$ Zeeman components, each of which has a Ramsey interference pattern superimposed on a Rabi pedestal. The symmetry resulting from the optical state preparation is apparent. The symmetry of the Ramsey peaks on the high-field Rabi peaks demonstrates the C-field uniformity expected from the long solenoid. The axial-field geometry chosen for this standard gives rise to a microwave field distribution that varies as a half sine wave in the direction of the atomic beam. The resulting Rabi line shape with narrow, smooth tails can be seen at each end of the spectrum. The spectral symmetry and line shapes should result in no Rabi[12] or Ramsey[13] pulling at the design accuracy.

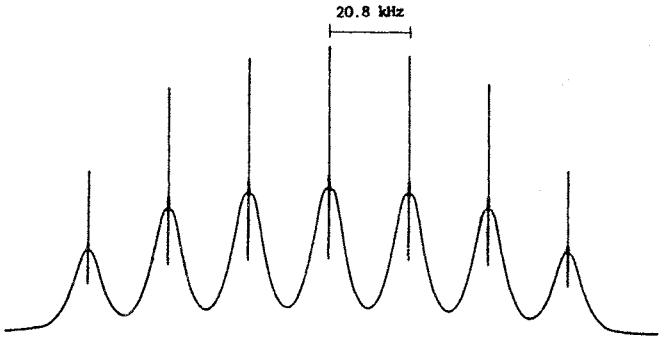


Fig. 4. The entire hyperfine spectrum showing all 7 $\Delta m=0$ Zeeman components. The microwave power is a few decibels below optimum.

With the microwave radiation tuned well off resonance or turned off, blocking the atomic beam gives a measure of the degree of optical pumping. Using the central peak as a calibrator, we can show that the amount of population left in the $F=4$ state after optical state preparation is less than 0.1% of the total population. Correspondingly, blocking the detector laser demonstrates that the scattered light background is on the order of 5% of the signal level in the main peak.

Figure 5 shows an expansion of the central Ramsey peak. The linewidth is 65 Hz ($Q = 1.4 \times 10^8$), in agreement with theoretical predictions based on the geometry of the standard, the oven temperature, and the chosen preparation and detection schemes. The signal-to-noise ratio at line center is 5000 in a 1 Hz bandwidth centered at 49 Hz. This is a factor of 3 smaller than the theoretical limit for these conditions and is a result of the beam signal at the detector being below its design value. The reason for this loss of signal is not known at this time.

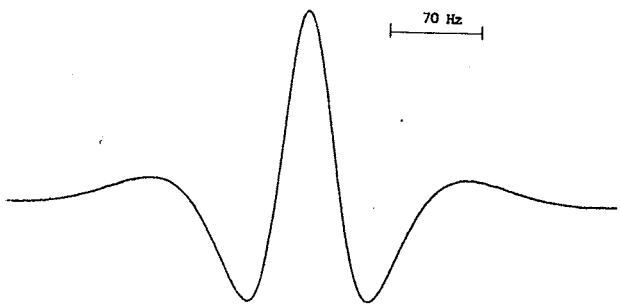


Fig. 5. Central Ramsey feature, "clock transition." The linewidth is 65 Hz and the S/N is 5000 in a 1 Hz bandwidth at 49 Hz.

Figure 6 shows the $m_F=-3 \rightarrow m_F=-3$ transition at a higher C-field value and a microwave power about 4 dB above optimum. It shows in greater detail the quality of the C-field by the perfect symmetry of the Ramsey peak on the Rabi peak. It also shows the presence of $\Delta m = \pm 1$ transitions. If we scale the size of these peaks for the change in power to optimum power and for the change in transition probability for those peaks next to the "clock" transition, the ratio of the main peak to the $\Delta m = \pm 1$ peaks adjacent to the main peak becomes 400:1. In this standard, these transitions should be excited only by curvature in the microwave field at the edges of the atomic beam. Calculating the effect of this field curvature over the atomic beam window results in a mean square value for the RF field perpendicular to the C-field which is 1/100th of the mean square field parallel to the C-field. The field curvature is thus adequate to explain the observed level of $\Delta m = \pm 1$ transitions. A quantitative theory for the $\Delta m = \pm 1$ lineshapes, which are excited by a "half cosine wave" is required to get quantitative agreement.

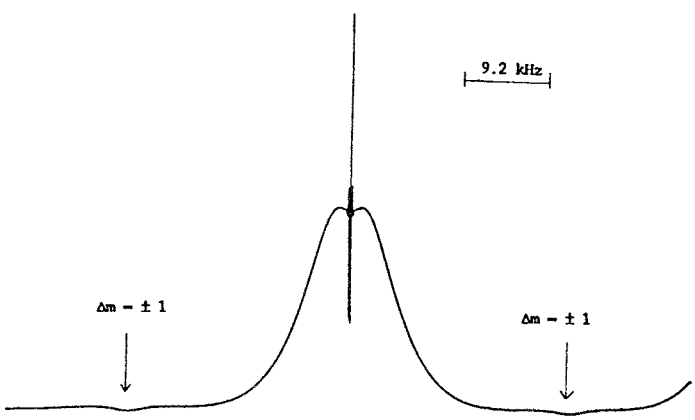


Fig. 6. The $m_F=-3 \rightarrow m_F=-3$ transition at 4 dB above optimum power showing the $\Delta m = \pm 1$ transitions.

Conclusions

Preliminary experiments have been performed with the NIST optically pumped, primary frequency standard. A polarization gradient scheme of optical pumping has been demonstrated to result in complete optical pumping even at low magnetic fields where coherence trapping could have been a problem. Microwave spectra have been used to evaluate some aspects of the standard. The "clock" signal line Q is the expected value of 1.4×10^8. The signal-to-noise ratio on the "clock" transition is essentially atomic shot-noise limited at about 5000 (in a 1 Hz bandwidth at 49 Hz) in the present experiments. The spectral symmetry indicates that magnetic field homogeneity and spectral pulling will not be important at the design accuracy. Finally, the scattered light does not degrade the signal-to-noise ratio.

Acknowledgements One of the authors, RED, would like to express his gratitude to SungHoon Yang for enjoyable and valuable interactions during these experiments.

References

1) Robert E. Drullinger, David J. Glaze, J. L. Lowe and Jon H. Shirley, IEEE Trans. Instrum. Meas. IM-40, (1991).

2) Andrea DeMarchi, Jon Shirley, David J. Glaze and Robert Drullinger, IEEE Trans. Instrum. Meas. IM-37, 185 (1988).

3) R. E. Drullinger, in Proc. 44th Annual Symposium on Frequency Control, 76 (1990).

4) John Lowe and F. L. Walls, "Ultralinear Small-Angle Phase Modulator," these proceedings.

5) A. M. Akul'shin, V. L. Velichanskii, A. S. Zibrov, V. V. Nikitin, V. A. Sautenkov, and G. G. Kharisov, Sov. J. Qunatum. Electron. 18, 1214 (1988).

6) T. W. Hänsch, I. S. Shahin and A. L. Shawlow, Phys. Rev. Lett. 27, 707 (1971).

7) L. S. Culter, U.S. Patent No. 4,425,653.

8) G. Orriols, Nuovo Cimento 55b, 1 (1979).

9) G. Théobald, N. Dimarcq, V. Giordano and P. Cérez, Opt. Commun. 71, 256 (1989).

10) V. Giordano, A. Hamel, P. Petit, G. Théobald, N. Dimarcq, P. Cérez and C. Audoin, "New Design for a High Performance Optically Pumped Cesium Beam Tube," to be published.

11) J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Amer. B, 11 2023 (1989).

12) A. DeMarchi, G. D. Rovera and A. Premoli, Metrologia 20, 37 (1984).

13) 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. March 1, (1991).

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 7B Number 77

H-masers in USSR

Nickolay B. Koshelyaevsky

NPO "VNIIFTRI", Mendeleevo, Moscow region, 141570, USSR

The first H-masers in USSR[1], fig.1, were put into operation at the quantum radiophysics laboratory of Lebedev Physical Institute at December 31, 1963.

After that at least three institutions in USSR were involved in this activity, they were VNIIFTRI, GNIPI and LNIRTI[2]. Little by little all these installations began to specialize at different kinds of H-masers. The VNIIFTRI H-masers were designated mainly for equipping of the State Primary Time and Frequency Standard (SPTFS) and the network of secondary standards of the State Time and Frequency Service (STFS). GNIPI and LNIRTI began to manufacture the commercially available H-masers.

The first H-maser at VNIIFTRI, fig.2, was put into operations by G.Yolkin[3] at March 1964, fig.3. It was pure investigation work but next experimental setup which was put into operations at 1966 has been included into SPTFS. It was used like frequency standard along with a set of ammonia $^{15}NH_3$ masers. The both above mentioned H-masers paying to tradition of molecular masers had a horizontal hydrogen beam and use for evacuation purpose the titanium sublimation vacuum pumps. The passively thermostabilized cavity was made usually of copper or brass and was not evacuated. The both H-masers had no AFC loop and cavity times to times was tuned manually.

The three next designs of VNIIFTRI H-masers leads to the instrument which looks quite similar to the modern one: the vertical cavity and the complex titanium getter and ion vacuum pump. They were equipped with the cavity AFC loop, the automatic cavity temperature control unit and magnetic field stabilization system. Since beginning of 70-th we started to use these H-masers like a clock at time laboratories, fig.5. There were produced about 15 such an H-masers, a part of them were used at VNIIFTRI for SPTFS equipping and the rest at time laboratories of STFS.

The last model of the full sized traditional designed H-maser was finished in VNIIFTRI at 1975. It gave us 6 precise instruments, fig.6, in which for the first time cavity AFC was based at beam intensity modulation. The frequency stability of these instruments were about $\sim 1\times10^{-13}$ for one day averaging time and reproducibility about $\sim 1\times10^{-13}$ per year.

The further VNIIFTRI work were directed to creation of an undersized H-masers. For example at 1980 there were produced the first two portable H-masers, fig.7, which employed the cylinder cavity of the standard size and undersizing was achieved due to better magnetic, temperature shielding, vacuum part design and miniaturization of electronics.

The two latest VNIIFTRI H-maser designs are reflected in separate papers, so not going into detail of it I briefly mention the main results of the last ten years. At January 1979 was obtained a generation of the H-maser with small sized cylinder cavity (diameter 196 mm)[4], for cavity undersizing there were some copper rods installed in it. Five years later, in 1984 for the first time was obtained the active maser operation with even less cavity[5] which was partially filled with a material of high permeability and low losses - sapphire (diameter 150 mm). The investigations in this direction leads to creation of portable H-clock "SAPPHIRE", fig.8. At present it is the main clock at SPTFS for remote time scale comparison and for testing radio time transfer links including GLONASS.

At the same time were created one more portable clock based on H-maser with full size spherical cavity - "SPHERE", fig.9.

The above mentioned thirty years hydrogen program of VNIIFTRI is connected with the names first of all Boris Gaigerov and George Yolkin.

At the beginning of 1970-th we have seen the first commercial H-masers of CH1-44 type produced by GNIPI. We used theirs for equipping SPTFS along with our own

instruments. Slightly later at the middle of 70-th CH1-46 was manufactured, fig.10, and at the beginning of the 80-th - CH1-70, fig.11. The availability of commercial instruments leads to equipping of all laboratories including SPTFS with it. It worth notice that technical assignments at these and even future models of commercial H-masers were adopted by VNIIFTRI. Nowadays H-masers are the main precise time keeping instruments at USSR time laboratories. The overall number of commercial H-masers manufactured up to now are about 200 instruments. The life expectancy of the commercial maser according to specifications is about 2 years, as a rule maser continuously operates about 5 years after that one need change storage bulb, regenerate titanium vacuum pump, change hydrogen source etc.

At present GNIPI produces the new state of the art H-masers CH-75 and CH1-80 types, fig.12, 13 with full size cavity and portable clock type CH1-76, based on passive H-maser with small sized cavity with delaying structure in it. The detailed information on these instruments, and instrumentation complexes based at these masers and intended to time laboratory equipping are reflected at Appendix 1.

The authors of commercial H-masers are Nickolay Demidov, Boris Sakharov and Adolf U'lyanov.

At the latest time appeared some information on LNIIRTI H-masers, Appendix 2.

The further part of the paper is devoted to such a long term characteristics oF H-masers which are not reflected at manufacturer specification and may be determined only during exploitation over long period. As a rule all first instruments of each commercial type H-masers are installed at SPTFS laboratories and then another users get it. In other words the time laboratory of SPTFS plays a role of testbed for determination the long term characteristics of the all commercial H-masers.

Fig. 15, 16 presents the result of 10 day averaged frequency difference relative to UTC(SU) time unit for the commercial H-masers CH1-70 and CH1-80 types which is used at SPTFS for time keeping needs. One can see new generation of H-masers to be obviously stable compare to the old one. But one have to take into account CH1-80 masers are running only one year, compare to CH1-70 masers of No 13xx more than 3 years and No 11xx more than 7 years.

Along with it the behavior of No 151 is strikingly poor - this H-maser has a significant frequency drift and no we nor manufacturer can't explain it.

If one would like to discuss presented at fig. 15, 16 dependencies he need some information on frequency stability of reference time scale, relative which measurements were made. This time scale is UTC(SU), it is free time scale based on an ensemble of about 10 H-masers[6]. Fig. 17 presents a part of results of frequency comparison between UTC and UTC(SU) based on preliminary data of GLONASS time link between VNIIFTRI and P.Daly laboratory from University of Leeds, UK.

Looking at the fig. 15,16 and 17 it is too difficult to estimate H-maser stability at short interval, concerning the annual interval possible frequency changes of the best masers No 147 & 149 are comparable to that of UTC(SU), which is estimated by the value about $\sim 3\times10^{-14}$per year.

Some words about long term characteristics of portable clocks based on undersized H-maser. We have no continuous frequency comparison of "SAPPHIRE" clock because of some updating and investigations were continuously made under it. At the usual laboratory environmental relative frequency instability for one day averaging time is about $\sim(1\text{-}2)\times10^{-14}$, transportation of it during 6 hours gain uncertainty about 3-5 nS.

The "SPHERE" H-maser without cavity AFC was installed at April 1990 at the main Moscow TV station Ostankino where it plays role of master clock for time dissemination needs. Up to now more than half year it runs free which is obvious from fig.18. The designer of this standard G.Yolkin is convinced that an obvious continuous frequency drift of a free running maser is due to tension decreasing in the outer silver coating of the fused quartz sphere constituting maser cavity. Removing this continuous drift one can extract the random part of frequency instability for 10 day averaging time to be about 4×10^{-14}. Apart from maser instability this value include the phase instability of transmitting and receiving part of time link between TV station and UTC(SU) time scale. For the usual laboratory environmental this maser has frequency instability $(1\text{-}2)\times10^{-14}$ for one day average time.

Acknowledgements. The author wish to thanks B.Gaigerov, N.Demidov, G.Yolkin, B.Sakharov and A.Ul'yanov who devoted the

main part of their life to H-masers and created some generations of an excellent instruments and that way they do his task easy and pleasant. More over he is indebted to all of them and to V.Tatarenkov also for discussion and facilitation of preparing this paper.

References.

[1]Н.Г.Басов, Г.М.Страховский, А.И.Никитин, Т.Ф.Никитина, В.М.Татаренков, А.В.Успенский, Радиотехника и электроника, X, 1809 - 1813, 1965.

Н.Г.Басов, Г.М.Страховский, А.И.Никитин, Т.Ф.Никитина, В.М.Татаренков, А.В.Успенский, Труды Физического института им.П.Н.Лебедева, XXXI, 139 - 177, 1965.

[2]VNIIFTRI, GNIPI and LNIRTI are russian abbreviations of All Union Scientific Research Institute for Physical-Technical & Radiotechnical Measurements, Nizhniy Novgorod (formerly Gorky) Scientific Research Institute for Instrumentation Manufacturing and Leningrad Scientific Research Institute for Radiotechnique, respectively.

[3]Г.А.Елкин, Исследование водородного генератора с целью его использования в составе эталона времени и частоты, диссертация на соискание ученой степени кандидата технических наук, Москва, 1968.

[4]В.А.Бочков, Б.А.Гайгеров, В.П.Королев, Ф.С.Русин, Радиотехника и электроника, No 4, 881, 1981.

[5]Ф.С.Русин, Б.А.Гайгеров, С.Ф.Пономаренко, Измерительная техника, No 5, 15, 1985.

Б.А.Гайгеров, Измерительная техника, No 6, 8, 1986.

[6]N.B.Koshelyaevsky, S.B.Pushkin, Proceedings of the Twenty-Second Annual Precise Time and Time Interval Meeting, December 4 - 6, 1990, to be published.

Figure captions.

Figure 1.
The first H-masers in USSR at Lebedev institute laboratory (magnetic shielding is removed).

Figure 2.
The first H-maser in VNIIFTRI.

Figure 3.
The IF signals of the first VNIIFTRI H-maser (a. far from threshold, b. near threshold, c. the maser signal)

Figure 4.
The first VNIIFTRI H-maser which take part at State Time and Frequency Standard.

Figure 5.
The overall view of typical VNIIFTRI H-maser.

Figure 6.
The last VNIIFTRI full size design of H-maser.

Figure 7.
The first VNIIFTRI portable H-clock.

Figure 8.
"SAPPHIRE" portable H-clock.

Figure 9.
"SPHERE" portable H-clock.

Figure 10.
The overall view of commercial H-maser CH1-46.

Figure 11.
The overall view of commercial H-maser CH1-70.

Figure 12
The overall view of commercial H-maser CH1-75.

Figure 13.
The overall view of commercial H-maser CH1-80.

Figure 14.
The overall view of the first commercial portable H-clock CH1-76.

Figure 15.
Ten day relative frequency differences $[F_{CH1-70(k)} - F_{UTC(SU)}]/F_{UTC(SU)}$

Figure 16.
Ten day relative frequency differences $[F_{CH1-80(k)} - F_{UTC(SU)}]/F_{UTC(SU)}$

Figure 17.
Ten day relative frequency difference $[F_{UTC(SU)} - F_{UTC}]/F_{UTC(SU)}$

Figure 18.
Ten day relative frequency difference $[F_{UTC(SU)} - F_{"SPHERE"}]/F_{UTC(SU)}$

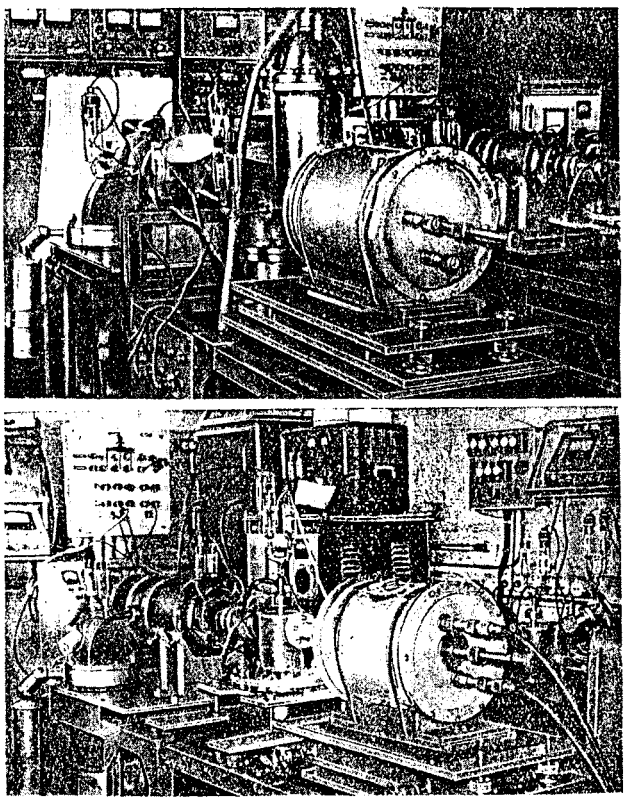


Fig. 1

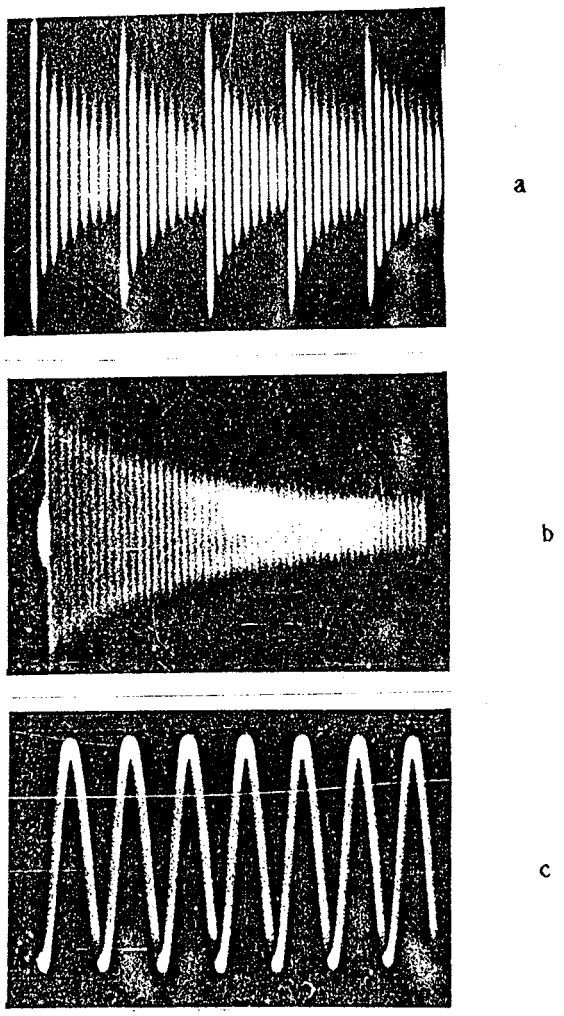


Fig. 3

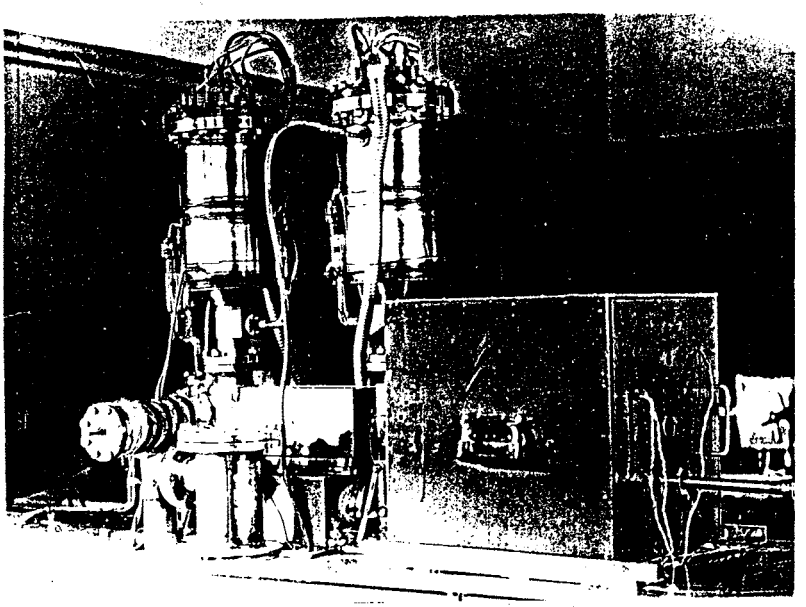


Fig. 2

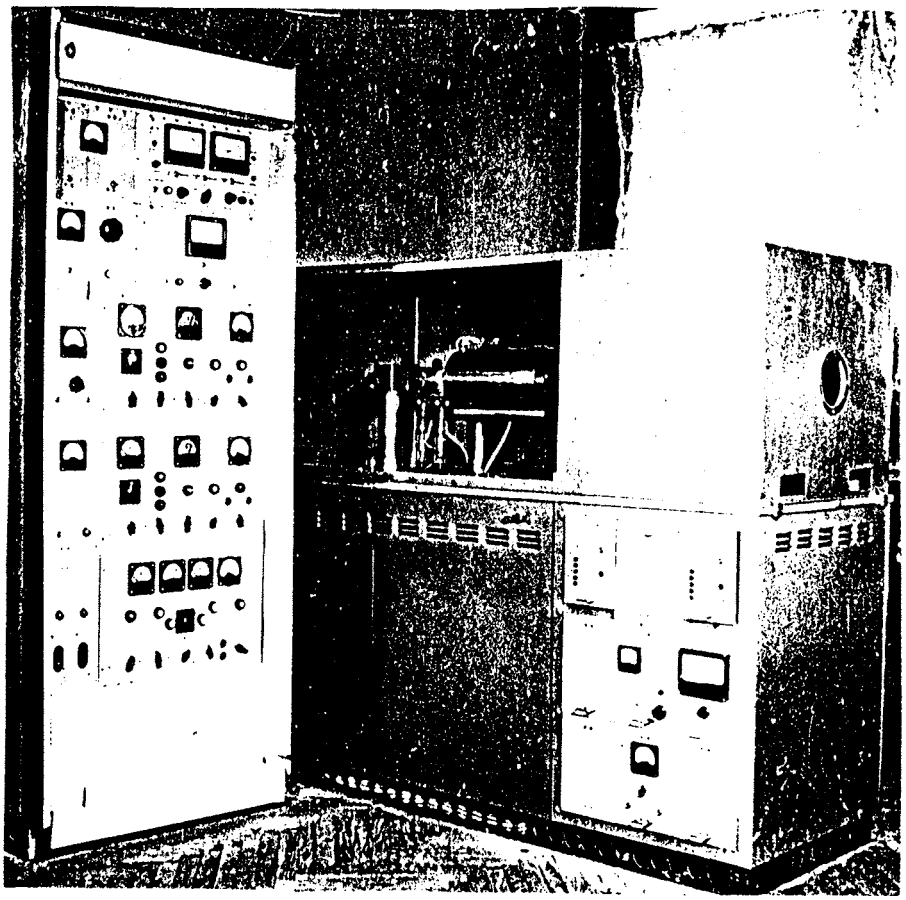


Fig.4

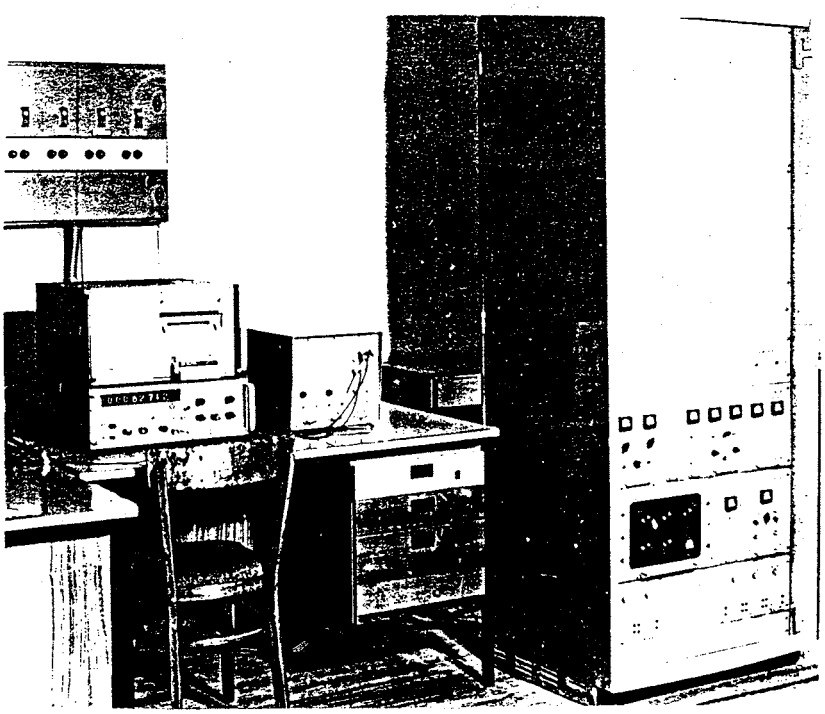


Fig. 5

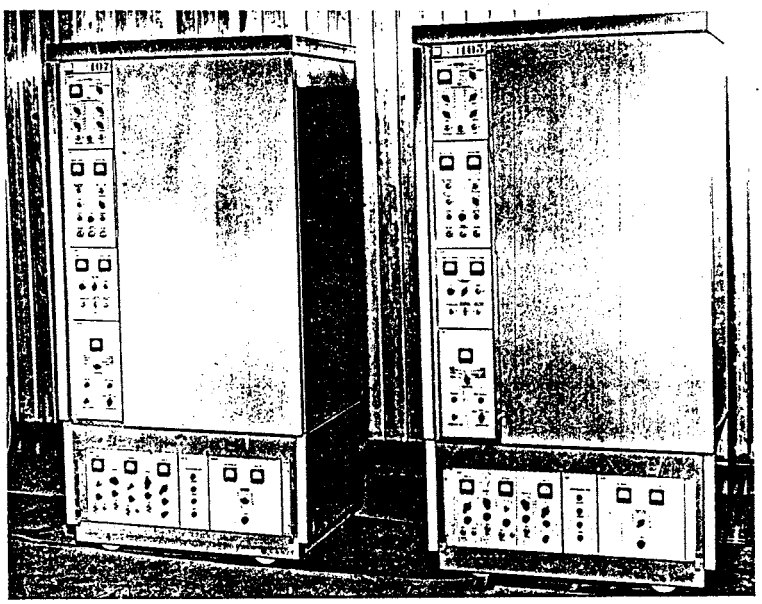


Fig. 6



Fig. 7

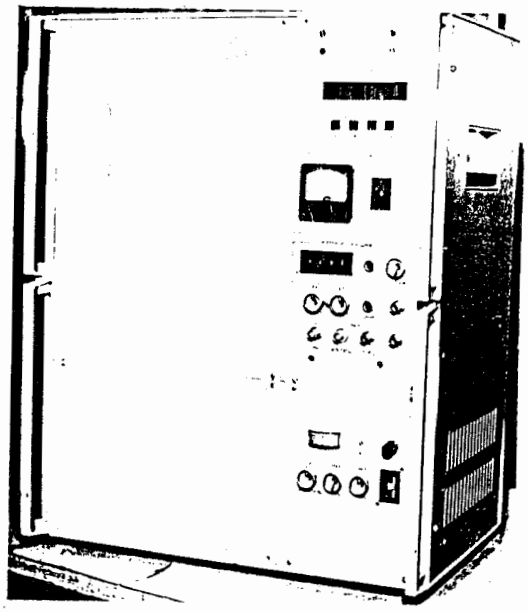


Fig.8

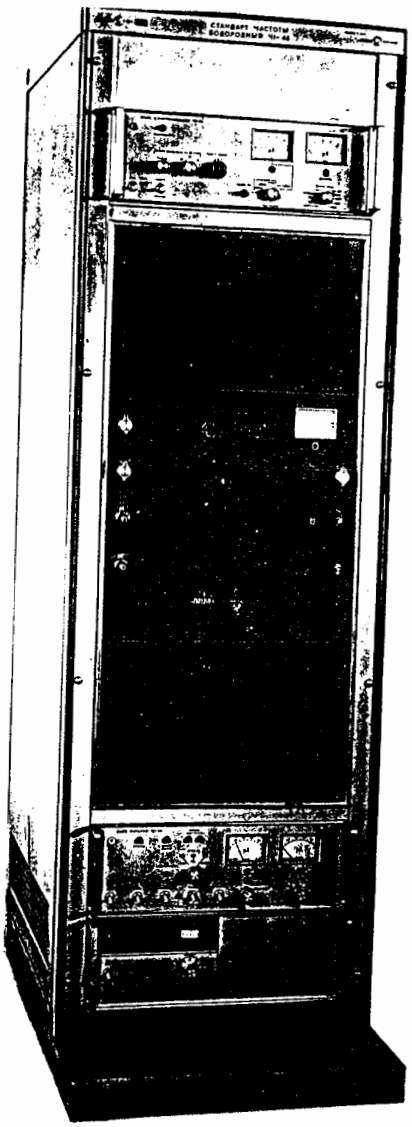


Fig. 10



Fig. 9

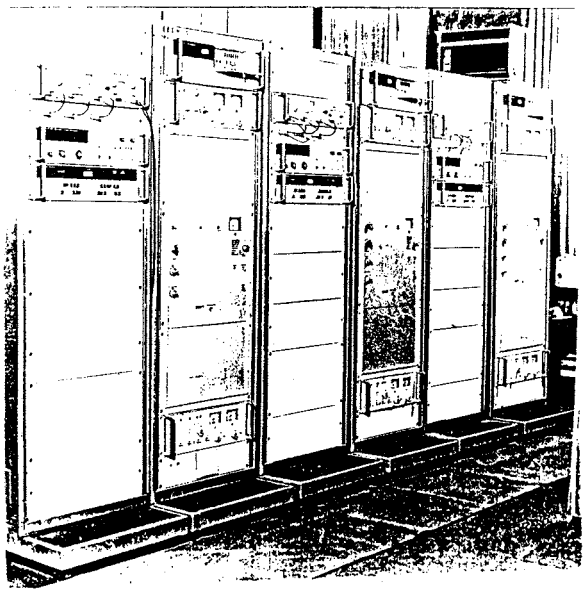


Fig. 11

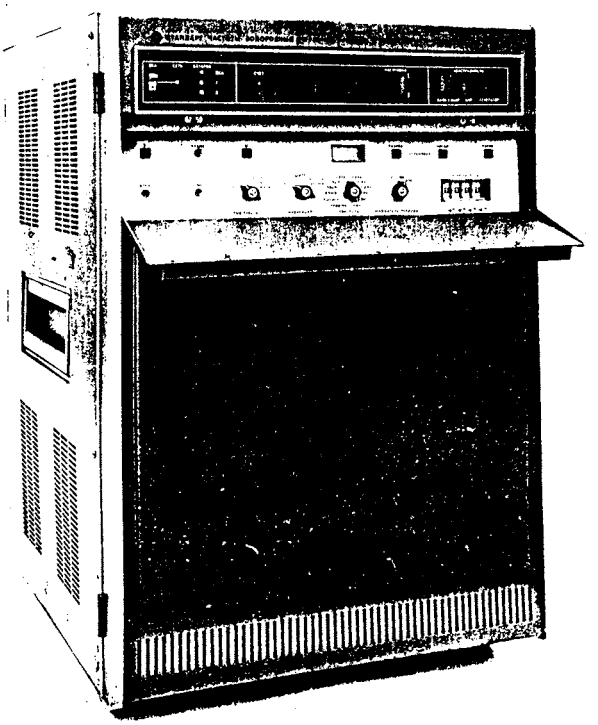


Fig. 12

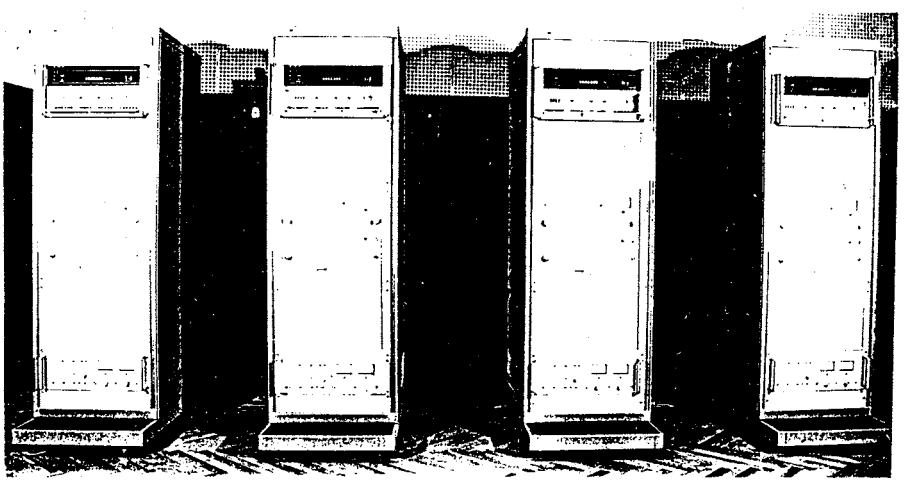


Fig. 13

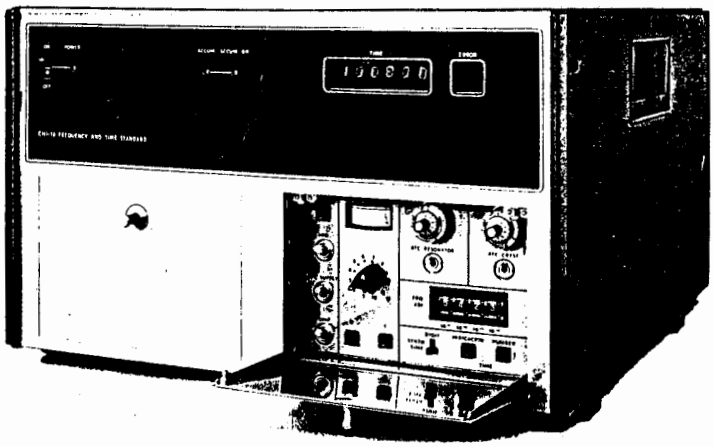


Fig. 14

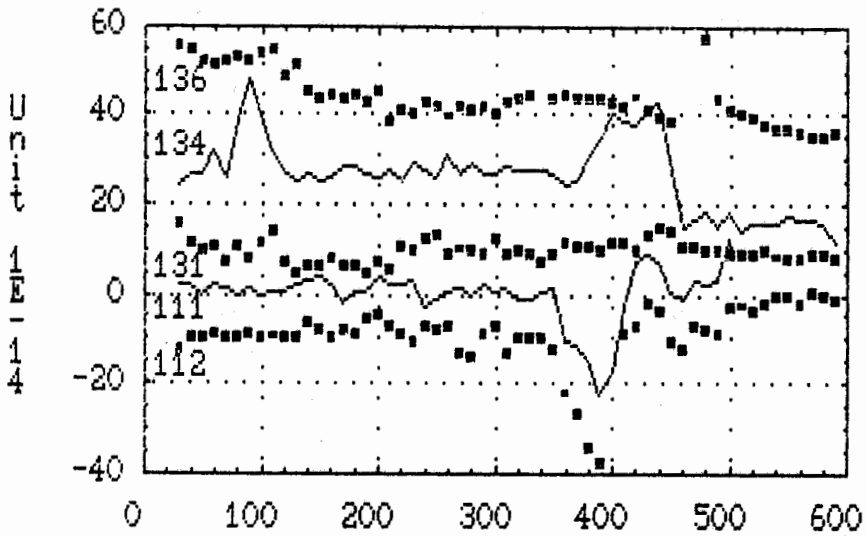


Fig. 15

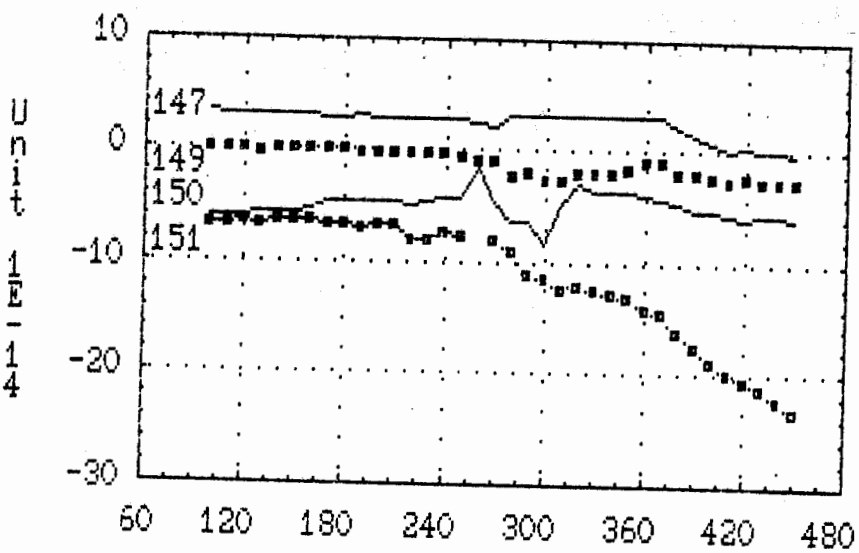


Fig. 16

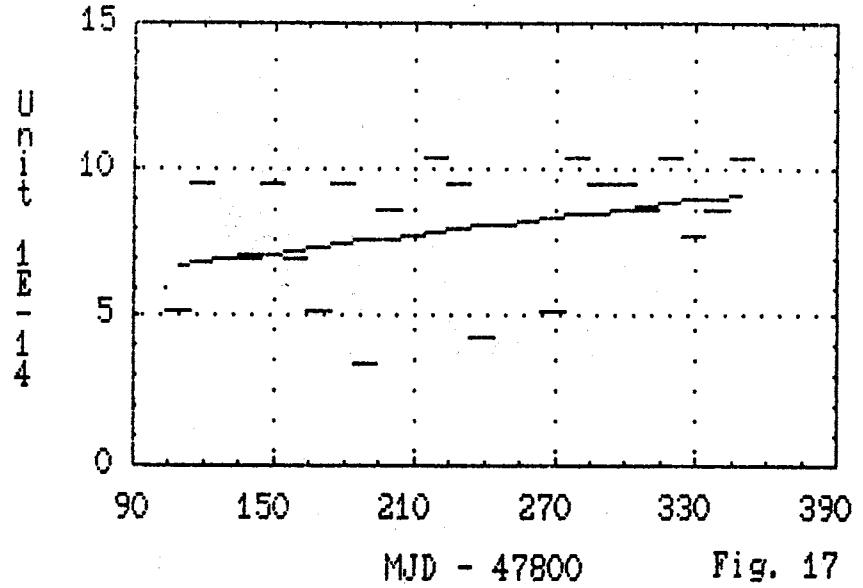

15
10
5
0
Unit 1E-14
90
150
210
270
330
390
MJD - 47800

Fig. 17

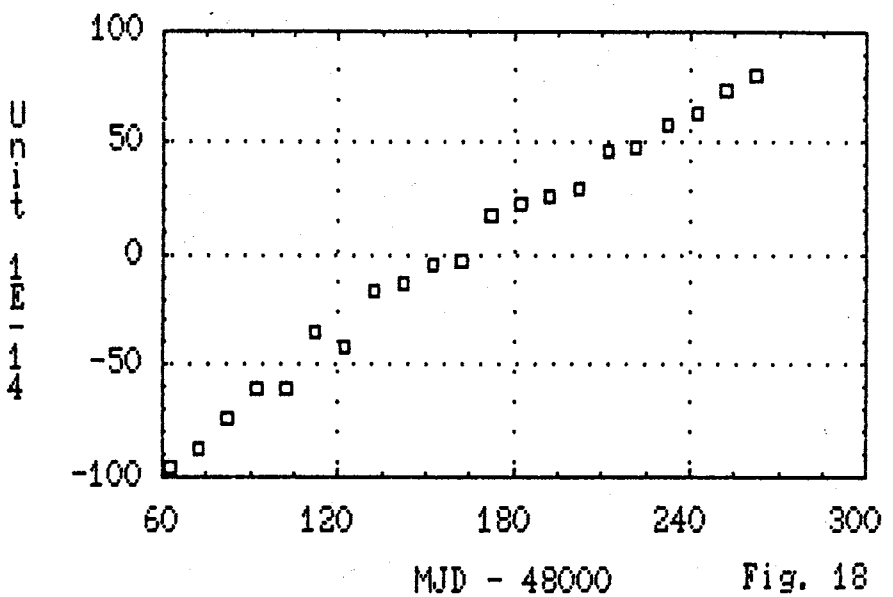
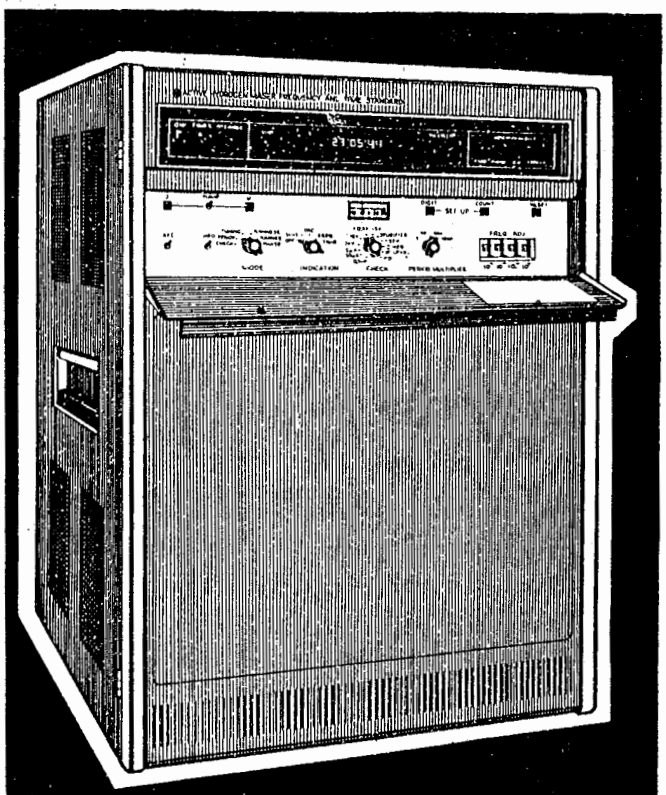

100
50
0
-50
-100
Unit 1E-14
60
120
180
240
300
MJD - 48000

Fig. 18

Active Hydrogen Maser Frequency and Time Standard

CH1-75



OUTPUT SIGNAL FREQUENCY STABILITY:

NOT MORE THAN $4 \cdot 10^{-13}/s$
NOT MORE THAN $3 \cdot 10^{-15}/h$
NOT MORE THAN $5 \cdot 10^{-15}/24\,h$

радиоизмерительная техника

BASED ON HYDROGEN MASER

APPLICATION:

NATIONAL TIME AND FREQUENCY CALIBRATION SERVICES

RADIOASTRONOMY

NAVIGATION

GEODESY

RESEARCH

SPECIFICATION

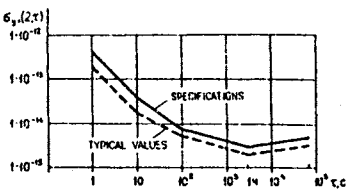
Output frequency: 5 MHz, 100 MHz.

Output signal voltage RMS value: (1 ± 0.2) V at load 50 Ohm.

Frequency relative accuracy: not more than $\pm 1 \cdot 10^{-12}$ (during one year period).

5 and 100 MHz output signal frequency stability (RMS relative two-selective frequency deviation or Allan variance) not more than: $4 \cdot 10^{-13}$ (1s); $3 \cdot 10^{-15}$ (1 h);
$4 \cdot 10^{-14}$ (10s); $5 \cdot 10^{-15}$ (24 h).
$8 \cdot 10^{-15}$ (100s);

Frequency stability per 100 s, 1 hour and 24 hours is provided at ±1°C ambient temperature change in operating temperature range 5-40°C



Frequency stability $5 \cdot 10^{-15}$/24 h is achived, when reference signal is applied from the second CH1-75 active hydrogen maser frequency standard.

Additional frequency variation at 1°C ambient temperature change: not more than $5 \cdot 10^{-15}$.

Frequency reproducibility relative error: not more than $5 \cdot 10^{-14}$.

Frequency compensator resolution $1 \cdot 10^{-14}$ in retuning range is not less than $1 \cdot 10^{-10}$.

SSB noise power spectral density is not more than: - 125 dB/Hz (20 Hz);
- 130 dB/Hz (330 Hz);
- 135 dB/Hz (1 kHz).

Frequency measurement error due to the built-in comparator is not more than:
$1 \cdot 10^{-13}$ (per 1s); $3 \cdot 10^{-15}$ (per 100 s);
$1{,}5 \cdot 10^{-14}$ (per 10 s); $5 \cdot 10^{-16}$ (per 1 hour).

1 Hz pulse signal parameters: amplitude - 2.5 V in 50 Ohm;
pulse width - 10-20 μs.

Pulse signal synchronization is performed by 1Hz positive polarity pulses.
Synchronization error is not more than ± 50 ns.

Power: (220 ±22) V,(50±0.5)Hz; (220 ±11) V, (115±5.75)V,(400 $^{+28}_{-12}$)Hz.
In case of emergency the hydrogen standard switches automatically to DC external power supply 22-30 V (is not included in the kit).

Power consumption: AC - 150 VA; DC - 100 W.

Dimensions: 480 x550 x 680 mm.

Weight: 90 kg.

Operating conditions:
environmental temperature: 5-40°C;
relative humidity: up to 80% at 25°C;
atmospheric pressure: 84-106 kPa (630-795 mmHg).

Magnetic sensitivity: not more than $1 \cdot 10^{-14}$ 1/Oersted at magnetic "C" field strength 0.5 mOersted.

Горьковское научно-производственное объединение "Кварц"
СССР 603009 Горький, пр. Гагарина, 176
Тел. 66-70-93. Телекс: 151115 LUCH SU
Факс в Горьком: 64-06-15

Nauchno-Proizvodstvennoje Objedinenije "KVARZ"
176 Prospect Gagarina, Gorky,603009, USSR
Tel. in Gorky 66-70-93. Telex 151115 LUCH SU
Fax in Gorky 64-06-15

Passive Hydrogen Maser Frequency Standard CH1-76

радиоизмерительная техника

BASED ON A QUANTUM HYDROGEN DISCRIMINATOR

APPLICATION:

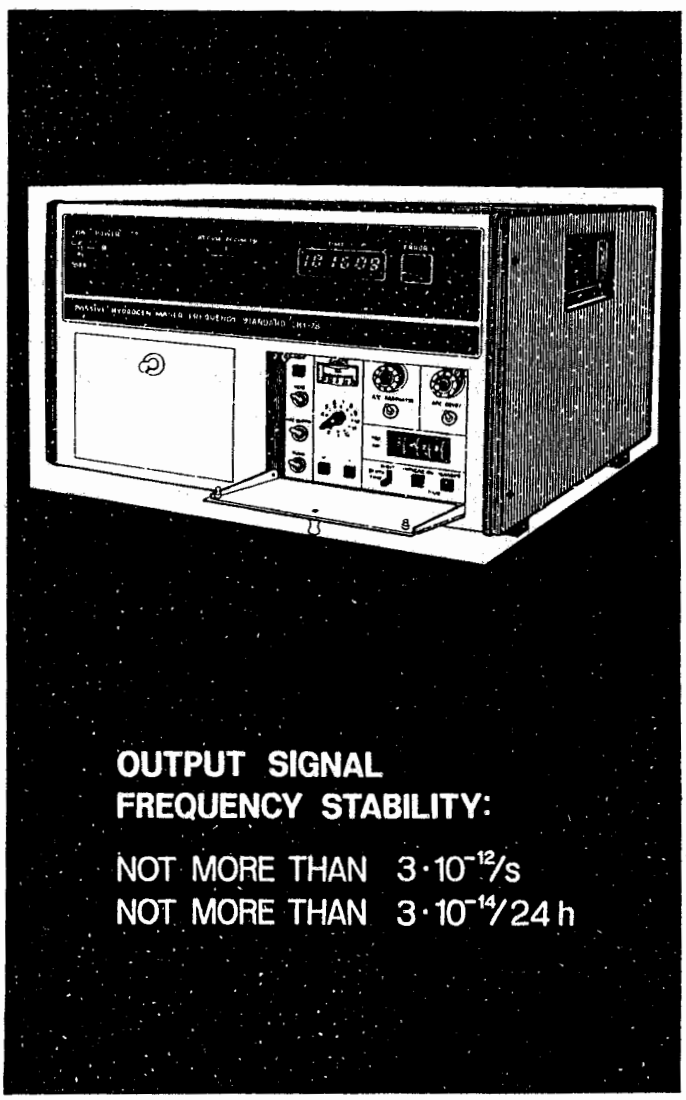
NATIONAL TIME AND FREQUENCY CALIBRATION SERVICES

RADIOASTRONOMY

NAVIGATION

GEODESY

RESEARCH



OUTPUT SIGNAL FREQUENCY STABILITY:

NOT MORE THAN $3 \cdot 10^{-12}/s$
NOT MORE THAN $3 \cdot 10^{-14}/24\,h$

SPECIFICATIONS

Output frequency: 1 Hz, 5 MHz

5 MHz output signal voltage RMS value: (1 ± 0.2) V in 50 Ohm

Frequency accuracy: not more than $\pm1.5\cdot10^{-12}$

1 Hz pulse signal parameters: amplitude - not less than 2.5 V in 50 Ohm
pulse width - 10-20 μs
rise time - not more than 30 ns

5 MHz output signal frequency stability (at ±1°C ambient temperature change in operating temperature range 5-40°C) not more than: $3\cdot10^{-12}$(1s)
$7\cdot10^{-13}$(10s)
$2\cdot10^{-13}$(100s)
$7\cdot10^{-14}$(1000s)
$5\cdot10^{-14}$(1h)
$3\cdot10^{-14}$(24)

Additional frequency variation at 1°C ambient temperature change within operating temperature range: not more than $\pm2\cdot10^{-14}$

Frequency reproducibility: not more than $3\cdot10^{-13}$

Compensator frequency retuning range: not less than $1\cdot10^{-10}$ at min step $1\cdot10^{-14}$

1 Hz pulse train is synchronized by positive polarity pulses with repetition rate 1e, pulse width not more than 2 μs, amplitude not less than 2.5 V in 50 Ohm; synchronization accuracy - not more than ±200ns.

5 MHz output signal harmonic component suppression: not less than 26dB

SSB phase noise power spectral density is not more than: -125dB/Hz (20Hz)
-195dB/Hz (330Hz)
-140dB/Hz (1kHz)

Power: (220 ± 22) V, (50 ± 0.5)Hz
(220 ± 11) V, (115 ± 5.75) V, (400^{+28}_{-12})Hz

In case of emergency the hydrogen standard automatically switches to DC external power supply 22-30 V, preserving specifications

Power consumption: AC - not more than 140 VA
DC -not more than 90 W

Weight: 55 kg

Dimensions: 280 x 480 x 555 mm

Operating conditions: environmental temperature: 5-40°C
relative humidity: up to 80% at 25°C
atmospheric pressure: 84-106 kPa
(630-795 mm Hg)

Горьковское научно-производственное объединение "Кварц"
СССР 603009 Горький, пр. Гагарина, 176
Тел. 66-70-93. Телекс: 151115 LUCH SU
Факс в Горьком: 64-06-15

Nauchno-Proizvodstvennoje Objedinenije "KVARZ"
176 Prospect Gagarina, Gorky, 603009, USSR
Tel. in Gorky 66-70-93. Telex 151115 LUCH SU
Fax in Gorky 64-06-15

**ОСНОВА –
КВАНТОВЫЙ
ВОДОРОДНЫЙ
ГЕНЕРАТОР**

**НЕСТАБИЛЬНОСТЬ
ЧАСТОТЫ
ВЫХОДНЫХ
СИГНАЛОВ:**

**ЗА СЕКУНДУ
НЕ БОЛЕЕ $5 \cdot 10^{-13}$,**

**ЗА СУТКИ
НЕ БОЛЕЕ $1 \cdot 10^{-14}$**

ПРИМЕНЕНИЕ:

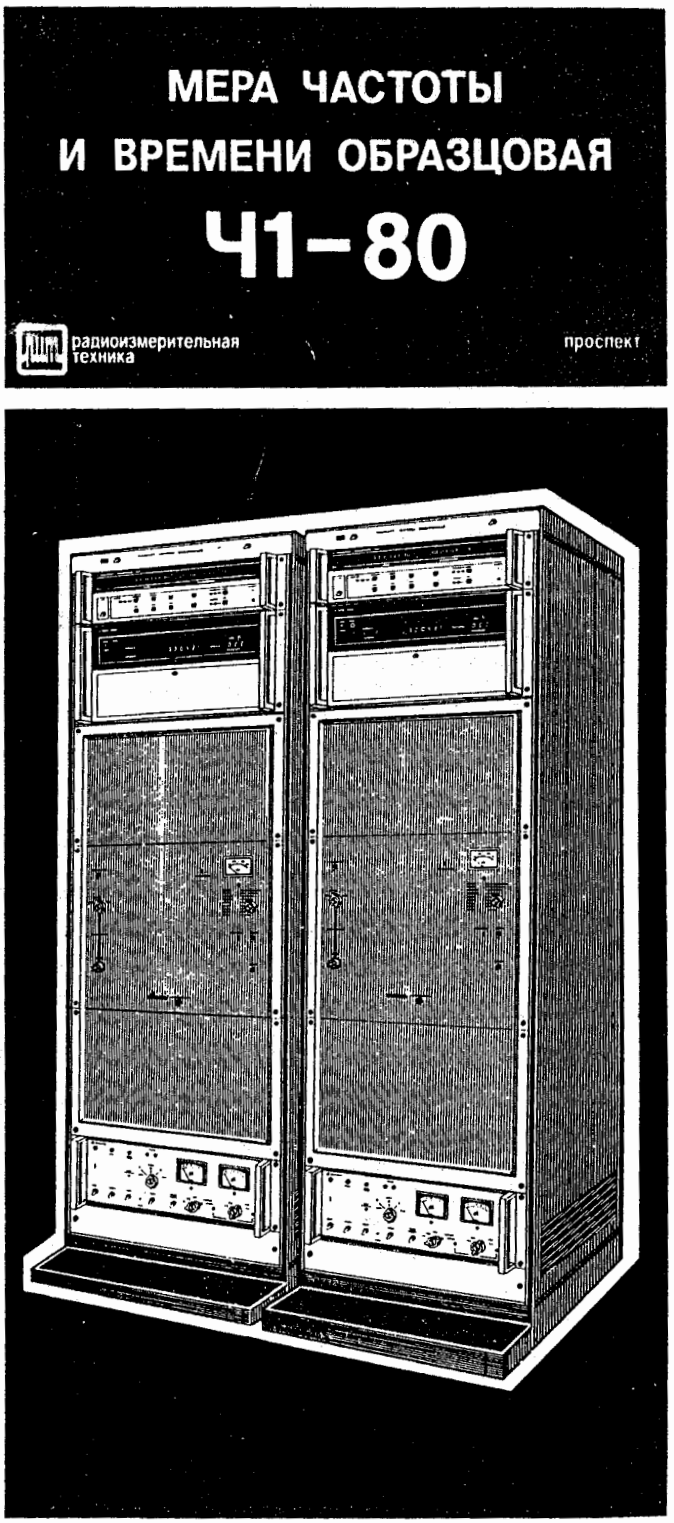
ГОСУДАРСТВЕННАЯ
СЛУЖБА
ВРЕМЕНИ И ЧАСТОТЫ

РАДИОАСТРОНОМИЯ

НАВИГАЦИЯ

ГЕОДЕЗИЯ

НАУЧНЫЕ
ИССЛЕДОВАНИЯ



ОСНОВНЫЕ ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

Выходные сигналы частотой 5 и 100 МГц.

Среднеквадратическое значение напряжения выходных сигналов (1±0,2) В на нагрузке 50 Ом.

Относительная погрешность по частоте (на интервале времени 1 год) не более $\pm 1 \cdot 10^{-12}$.

Нестабильность частоты выходных сигналов 5 и 100 МГц (среднеквадратическое относительное двухвыборочное отклонение частоты или дисперсия Аллана) не более:

$5 \cdot 10^{-13}$ (за 1 с),
$5 \cdot 10^{-14}$ (за 10 с),
$1 \cdot 10^{-14}$ (за 100 с),
$1 \cdot 10^{-14}$ (за 1 ч),
$1 \cdot 10^{-14}$ (за 1 сут).

Нестабильность частоты за 100 с, 1 ч, 1 сут обеспечивается при изменении температуры в пределах ±1°С в любой точке интервала рабочих температур 10–35°С (рис. 5).

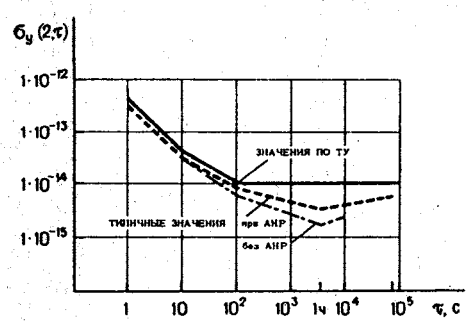


Рис. 5. Зависимость нестабильности частоты от времени измерения

Дополнительное относительное изменение частоты при изменении окружающей температуры на 1°С в интервале температур 10–35°С не превышает $1 \cdot 10^{-14}$.

Относительная погрешность воспроизведения частоты не более $3 \cdot 10^{-14}$.

Разрешающая способность корректора частоты $1 \cdot 10^{-14}$ в диапазоне перестройки не менее $1 \cdot 10^{-10}$.

Погрешность измерения частоты, вносимая встроенным компаратором, не более.

$1 \cdot 10^{-13}$ (за 1 с),
$2 \cdot 10^{-14}$ (за 10 с),
$3 \cdot 10^{-15}$ (за 100 с),
$5 \cdot 10^{-16}$ (за 1 ч).

Импульсные сигналы прямоугольной формы положительной полярности, амплитудой не менее 5 В на нагрузке 50 Ом, частотой 100, 10, 1 кГц; 100, 10, 1, 1/10, 1/60 Гц и длительностью (1±0,5), (10±1), (100±10) мкс; (1±0,1), (10±1), (100±10), (1000±100) мс соответственно.

Сигналы времени:

шкала времени – импульсы положительной полярности с периодом повторения 1 с, длительностью 10–30 мкс, амплитудой не менее 2,5 В на активной нагрузке 50 Ом;

задержанная шкала времени – импульсы положительной полярности с параметрами, аналогичными указанным выше; задержка импульсов в пределах 0–999999,9 мкс дискретно через 0,1 мкс и в пределах 0–0,1 мкс плавно относительно импульсов основной шкалы времени.

Питание: сеть напряжением (220±22) В частотой 50 Гц. При отключении сети образцовая мера может автоматически переключаться на работу от внешнего источника напряжением 25–30 В (в комплект поставки не входит) с сохранением технических характеристик.

Мощность, потребляемая каждым водородным стандартом частоты от сети переменного тока, не более 500 В·А при номинальном напряжении сети и в прогретом состоянии, а от источника 25–30 В не более 250 Вт.

Масса каждого водородного стандарта не более 275 кг.

Габариты 600х1985х900 мм.

ДОПОЛНИТЕЛЬНЫЕ ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

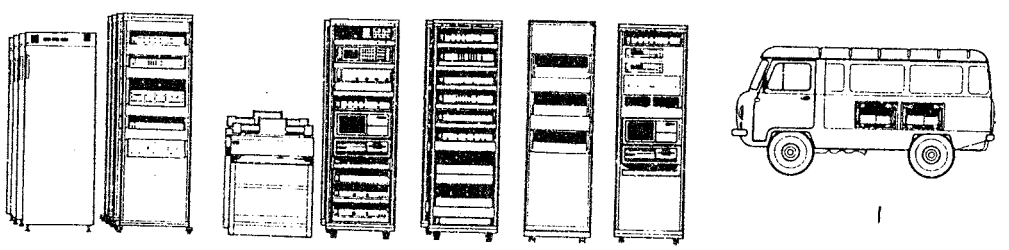
Чувствительность к магнитному полю не более $2 \cdot 10^{-13}$ 1/Э при величине поля „С" 10^{-3} Э.

Чувствительность к внешнему давлению не обнаружена.

© Центральный отраслевой орган научно-технической информации „ЭКОС". Подписано к печати 23.03.89. 01143. Зак. 10941

Эталон частоты и времени Ч0-110

Time and frequency Standard Set CH0-110



ПРИМЕНЕНИЕ

- ГОСУДАРСТВЕННАЯ СЛУЖБА ЧАСТОТЫ И ВРЕМЕНИ
- РАДИОАСТРОНОМИЯ
- НАВИГАЦИЯ
- НАУЧНЫЕ ИССЛЕДОВАНИЯ

APPLICATION

- NATIONAL TIME AND FREQUENCY CALIBRATION SERVICES
- RADIOASTRONOMY
- NAVIGATION
- RESEARCH

радиоизмерительная техника

ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

- Частота выходных сигналов 5, 100 МГц и 1 Гц
- Погрешность воспроизведения частоты $1 \cdot 10^{-14}$
- Нестабильность частоты выходных сигналов эталона определяется хранителем частоты и времени Ч0-202:
  $4 \cdot 10^{-13}$ (за 1 с),
  $4 \cdot 10^{-14}$ (за 10 с),
  $1 \cdot 10^{-14}$ (за 100 с),
  $3 \cdot 10^{-15}$ (за 1 ч)
  $3 \cdot 10^{-15}$ (за 1 сут)
- Погрешность измерения нестабильности частоты:
  $1 \cdot 10^{-13}$ (за 1 с),
  $2 \cdot 10^{-14}$ (за 10 с),
  $2 \cdot 10^{-15}$ (за 100 с),
  $1 \cdot 10^{-15}$ (за 1 ч)
- Погрешность внешних сличений транспортируемыми часами за 2 сут не более 10 нс
- Разрешающая способность и погрешность измерения интервалов времени 0,1 нс
- Аппаратура эталона питается от сети переменного тока напряжением 220 ± 22 В, частотой $50 \pm 0{,}5$ Гц
  При отсутствии напряжения в сети вся аппаратура эталона (кроме автоматизированной измерительной системы и системы автоматизированного контроля) переходит на питание от внешних аккумуляторов напряжением 25-30 В
- Потребляемая мощность, не более:
  от сети переменного тока 7500 В·А
  от источника постоянного тока 2100 Вт
- Габариты: 10 стоек 596x1985x900 мм (каждая)
- Масса не более 2500 кг

SPECIFICATIONS

- Output frequency: 5, 100 MHz, 1 Hz
- Frequency reproducibility error: $1 \cdot 10^{-14}$
- The standard output signal frequency instability due to the CHO-202 time and frequency standard:
  $4 \cdot 10^{-13}$ (1 s)
  $4 \cdot 10^{-14}$ (10 s)
  $1 \cdot 10^{-14}$ (100 s)
  $3 \cdot 10^{-15}$ (1 h)
  $3 \cdot 10^{-15}$ (24 h)
- Frequency stability measurement error:
  $1 \cdot 10^{-13}$ (1 s)
  $2 \cdot 10^{-14}$ (10 s)
  $2 \cdot 10^{-15}$ (100 s)
  $1 \cdot 10^{-15}$ (1 h)
- Transportable clock time external comparison error: not more than 10 ns/48h
- Resolution and time interval measurement error: 0.1 ns
- Power: 220 ± 22 V, 50 ± 0.5 Hz
  In case of emergency the standard instrumentation (except automated measurement system and automated control system) switches automatically to DC external power supply 25-30 V
- Power consumption: not more than
  AC - 7500 VA
  DC - 2100 W
- Dimensions: 10 racks, 596x1985x900 mm each
- Weight: not more than 2500 kg

Горьковское научно-производственное объединение "Кварц"
СССР 603009 Горький, пр. Гагарина, 176
Тел. 66-70-93. Телекс: 151115 LUCH SU
Факс в Горьком: 64-06-15

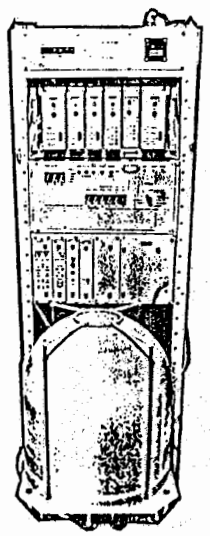
Nauchno-Proizvodstvennoje Objedinenije "KVARZ"
176 Prospect Gagarina, Gorky, 603009, USSR
Tel. in Gorky 66-70-93. Telex 151115 LUCH SU
Fax in Gorky 64-06-15

Appendix 2

„СФЕРА"

ВОДОРОДНЫЙ СТАНДАРТ ЧАСТОТЫ И ВРЕМЕНИ ДЛЯ НАЗЕМНОЙ СТАЦИОНАРНОЙ И ПОДВИЖНОЙ АППАРАТУРЫ

ПРЕДНАЗНАЧЕН ДЛЯ ФОРМИРОВАНИЯ ВЫСОКОСТАБИЛЬНОГО ПО ЧАСТОТЕ СИГНАЛА И ШКАЛЫ ВРЕМЕНИ



- ВЫСОКАЯ СТАБИЛЬНОСТЬ
- БОЛЬШОЙ РЕСУРС РАБОТЫ
- РЕМОНТОПРИГОДНОСТЬ
- ЭКСПЛУАТАЦИЯ В ШИРОКОМ ДИАПАЗОНЕ ТЕМПЕРАТУР И ВИБРАЦИЙ

- НАВИГАЦИЯ
- СВЯЗЬ
- МЕТРОЛОГИЯ
- РАДИОЛОКАЦИЯ
- НАУЧНЫЕ ИССЛЕДОВАНИЯ

ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

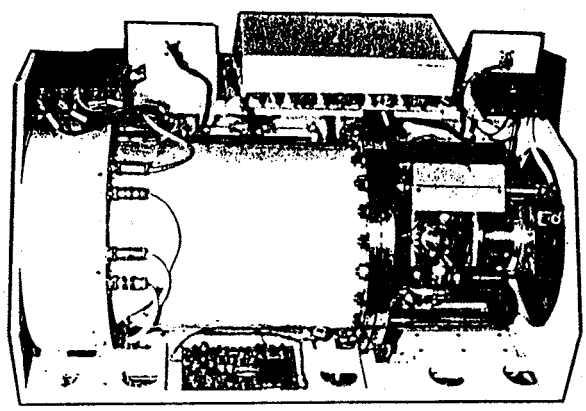
ЧАСТОТА ВЫХОДНОГО СИГНАЛА		5МГц, 1 Гц
НЕСТАБИЛЬНОСТЬ ЧАСТОТЫ	за 1с	5.10^{-13}
	за 100с	3.10^{-14}
	за 1 час	1.10^{-14}
	за 1 сутки	1.10^{-14}
ДИСКРЕТНОСТЬ УСТАНОВКИ ВРЕМЕННОЙ ШКАЛЫ		100нс
РЕСУРС		20.000 часов
ДИАПАЗОН РАБОЧИХ ТЕМПЕРАТУР		+5°С – +40°С
НАПРЯЖЕНИЕ ПИТАНИЯ		+27В
ПОТРЕБЛЯЕМАЯ МОЩНОСТЬ		120Вт
ГАБАРИТЫ		1500x540x450 мм
МАССА		120 Кг

193124 Ленинград
Телетайп 121001
Телеграф Ленинград ОМ
Телефакс 2731033
Телефон 274-18-94

„ЛАЗУРИТ" КОМПАКТНЫЙ ВОДОРОДНЫЙ СТАНДАРТ ЧАСТОТЫ ДЛЯ КОСМИЧЕСКИХ АППАРАТОВ

НАВИГАЦИЯ
СВЯЗЬ
МЕТРОЛОГИЯ
РАДИОЛОКАЦИЯ
НАУЧНЫЕ ИССЛЕДОВАНИЯ

ПРЕДНАЗНАЧЕН ДЛЯ ФОРМИРОВАНИЯ ВЫСОКОСТАБИЛЬНОГО ПО ЧАСТОТЕ СИГНАЛА



ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ

ЧАСТОТА ВЫХОДНОГО СИГНАЛА	5МГц
НЕСТАБИЛЬНОСТЬ ЧАСТОТЫ	
для τ = 1-1000с	$1{,}5 \cdot 10^{-12}/\sqrt{\tau}$
для τ = 1 сутки	$2 \cdot 10^{-14}$
ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТЬ УСТАНОВКИ ЧАСТОТЫ	$1 \cdot 10^{-14}$
РЕСУРС	15000 часов
НАПРЯЖЕНИЕ ПИТАНИЯ	+ 27В
ПОТРЕБЛЯЕМАЯ МОЩНОСТЬ	90Вт
ДИАПАЗОН РАБОЧИХ ТЕМПЕРАТУР	0 - +40°С
ОБЪЕМ	75 $дм^3$
МАССА	40 Кг

ВЫСОКАЯ СТАБИЛЬНОСТЬ
УЛУЧШЕННЫЕ ГАБАРИТНО-ВЕСОВЫЕ ХАРАКТЕРИСТИКИ
БОЛЬШОЙ РЕСУРС РАБОТЫ

193124 Ленинград
Телетайп 121001
Телеграф Ленинград ОМ
Телефакс 2731033
Телефон 274-18-94

SESSION 8B

TIME AND FREQUENCY IV

Chairman : S. LESCHIUTTA

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 8B Number 78

STUDY OF THE CLOCK-ENSEMBLE CORRELATION IN ALGOS USING REAL DATA

Patrizia Tavella
Istituto Elettrotecnico Nazionale Galileo Ferraris
Strada delle Cacce 91, 10135 Torino, Italy

Jacques Azoubib and Claudine Thomas
Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex, France

ABSTRACT

For a time scale computed as the weighted average of the readings of an ensemble of clocks to be stable, each clock weight can be chosen as inversely proportional to the variance of its frequency.

If the clock frequencies are estimated relative to the ensemble time the variances, and hence the clock weights, are biased. In this paper we derive an analytic correction factor to the biased variance, representing this clock-ensemble correlation. We also study how the implementation of an upper limit of weight can help control the correlation effect.

As a numerical experiment, we introduce such a correction factor in the algorithm ALGOS used to compute the international reference time scale TAI. The experiment consists in reprocessing the real clock data, collected for TAI over the last few years, using either the total clock set or a selection. The calculations show that the correlation correction makes no significant change in the clock weights or in the stability of the resulting time scale even when few clocks are selected. This agrees with theoretical expectations for a large number of clocks. Particular features of the design of the ALGOS weighting procedure also serve to reduce the impact of the clock-ensemble correlation effect.

INTRODUCTION

To keep time, national timing centers have at their disposal commercial caesium clocks which can be carefully compared among them. They can be treated together as an ensemble through an algorithm designed to raise the stability, accuracy and reliability of the obtained time scale above the level of performance of any individual clock in the ensemble.

To provide stability, the time scale can be computed as a weighted average of the readings of individual clocks and the weight attributed to a given clock is chosen as inversely proportional to its frequency variance. The choice of the variance type (classical, filtered or Allan variance) and of the observation period of the clock used to estimate it, depends on the needs of the user and on the available time measurements [1].

In most cases, the only frequencies available for an individual clock are those obtained relative to ensemble time. Since ensemble time necessarily contains the contribution of this particular clock, the frequencies, the clock frequency variance and the weight are inherently biased. This clock-ensemble correlation effect is well known and is taken into account, with slightly different analytic forms, in most current time scale algorithms [2] [3] [4].

In the first section of this article, we develop the theoretical approach of Yoshimura [5], on the variance bias due to clock-ensemble correlation. The effect of this correlation and its partial control through the implementation of an upper limit of weight, is then illustrated using simulated clocks.

Our second section is devoted to a numerical experiment performed on real clock data collected, over the last three years, for the computation of the International Atomic Time TAI. The weighting procedure for TAI does not take into account the clock-correlation effect. Here the time scale is recomputed after implementation of the correction factor to the weights, using either the total clock set (about 180 clocks) or a selection of clocks. The resulting weight distributions and time scale stabilities are then analysed.

Note: The symbols used in the mathematical development of this work are:

t: date of the time scale update,

H_i: clock identification,

$h_i(t)$: reading of clock H_i at date t (This quantity is not directly accessible by experiment.),

$p_i(t)$: absolute weight assigned to clock H_i for the time scale computation at date t,

$w_i(t)$: relative weight assigned to clock H_i for the time scale computation at date t (w_i is obtained from p_i by normalization.),

N: number of clocks in the ensemble,

$x_i(t) = TA(t)-h_i(t)$: time offset of clock H_i, at date t, from the time scale TA under computation (This quantity gives user access to the time scale.),

τ: time interval between two consecutive updates of the time scale,

$y_i(t) = [x_i(t)-x_i(t-\tau)]/\tau$: frequency departure of clock H_i, at date t, from the time scale TA under computation,

$f_i(t)$: frequency departure of clock H_i, at date t, from an ideal time scale.

1. THEORETICAL DERIVATION OF THE CLOCK-ENSEMBLE CORRELATION

For a time scale computed as a weighted average of readings of individual clocks to be stable, the weight attributed to a given clock is chosen as inversely proportional to its frequency variance. Its relative weight is then written as:

$$w_i = \frac{1}{\sigma_i^2} \left[\Sigma_k \frac{1}{\sigma_k^2}\right]^{-1}. \qquad (1)$$

However, often the only available frequency values for a given clock are obtained relative to the ensemble time. This prevents access to the real performance of the clock dissociated from the ensemble and biases its weight.

The general analytic expression for the biased variance is obtained with writing the frequency departure of clock H_i, at date t, from the time scale TA under computation, as:

$$y_i(t) = f_{TA}(t) - f_i(t). \qquad (2)$$

The variance of the quantity $y_i(t)$ is:

$$var_i = E[(y_i - \bar{y}_i)^2], \qquad (3)$$

where the date identification t has been suppressed and where E and the superscript bar indicate the expectation value.

Inserting (2) in (3), we obtain:

$$var_i = \sigma_i^2 + \sigma_{TA}^2 - 2\sigma_{i,TA}, \qquad (4)$$

where σ_i^2 is the true frequency variance of clock H_i (already involved in (1)), σ_{TA}^2 is the true frequency variance of TA and $\sigma_{i,TA}$ is the covariance between clock H_i and TA.

Equation (4), derived by Yoshimura [5], relates the biased frequency variance of clock H_i to the unbiased variance. An analytic definition of TA is necessary for further steps in the derivation.

1.1 EXPRESSION WITH CONSTANT WEIGHTS

The biased frequency variance of clock H_i is determined from a finite number of frequency samples. Suppose that this estimation is performed over the same period for all the clocks of the ensemble and that the clock weights are kept constant for that period. The implementation in the time scale definition of a mode of prediction for the clock frequencies is, in this case, unnecessary [6]. The time scale is then simply the weighted average of clock readings, written, at date t, as:

$$TA(t) = \Sigma_i w_i h_i(t), \qquad (5)$$

where the relative weights w_i are numerical coefficients subject only to the condition $\Sigma_i w_i = 1$.

Equation (5) leads, together with the assumption that the contributing clocks are uncorrelated among them, to:

$$\bar{y}_{TA}(t) = \Sigma_i w_i \bar{y}_i(t) \qquad (6)$$

and to the expression for the true variance, σ_{TA}^2, of the time scale TA:

$$\sigma_{TA}^2 = \Sigma_k w_k^2 \sigma_k^2. \qquad (7)$$

With the same hypothesis, the covariance between clock H_i and TA is:

$$\sigma_{i,TA} = w_i \sigma_i^2. \qquad (8)$$

The relationship (4) between the biased and the true variances of clock H_i may then be expressed as:

$$var_i = \sigma_i^2(1-2w_i) + \Sigma_k w_k^2 \sigma_k^2, \qquad (9)$$

an equation which is valid for the classical variance, for a two-sample variance, or for any other filtered variance.

The biased variance of clock H_i thus contains the true variance of clock H_i coupled with its relative weight w_i and another term which includes the true variances and relative weights of all other clocks contributing to the ensemble time. When (9) is applied to the N clocks contributing to TA and combined with N computed values of var_i, we obtain a solvable system which gives access to the N true variances σ_i^2.

In the limiting case of a very large ensemble of similar clocks, all the weights are equal and close to 1/N. Equation (9) then reduces to:

$$\sigma_i^2 = var_i, \qquad (10)$$

which means that the biased variances are very close to the true ones.

So far the relative weights w_i have been treated merely as numerical coefficients satisfying $\Sigma_i w_i = 1$. Suppose now that TA is computed at date t using weights w_i which are representative of the true stability of the clocks. Weights w_i then approximate those of (1). Including (1) in the expression (7), which represents the true variance of TA, gives:

$$\Sigma_k w_k^2 \sigma_k^2 = w_j \sigma_j^2 \text{ for any } j. \qquad (11)$$

One then obtains:

$$var_i = \sigma_i^2(1-w_i), \qquad (12)$$

for the biased variance, that is, the effect of the clock-ensemble correlation is proportional to the relative contribution of the clock inside the ensemble. This correction factor to the weight appears in [3] and [4] and, with a multiplicative factor close to 1, in [2]. We also use the expression in the numerical experiment of Sect. 2.

1.2 ILLUSTRATION OF THE CORRELATION EFFECT, ROLE OF AN UPPER LIMIT OF WEIGHT

To illustrate Sect. 1.1, suppose we have at our disposal six clocks, the true variances σ_i^2 of which are known and remain constant for the whole period under observation. These six clocks should enter the TA computation with constant relative weights representative of their individual quality as in (1). But this is not quite correct because of the clock-correlation effect. We therefore use (9) to show the effect of using biased variances rather than true variances in the weighting of these six clocks.

The experiment consists in studying the evolution in time of the biased weights of the clocks. We proceed as follows: we imagine that we run TA for n intervals of duration τ, from date t+mT to date t+(m+1)T where t is the date of initialisation of the procedure and where $T=n\tau$. The clock weights are supposed constant from date t+mT to date t+(m+1)T and are denoted $w_i(m)$, to characterize sample m. The variance of the n frequencies of clock H_i relative to TA, for sample m, is biased and is calculated from:

$$var_i(m) = \sigma_i^2[1-2w_i(m)] + \Sigma_k w_k(m)^2\sigma_k^2. \quad (13)$$

The absolute weight $p_i(m+1)$ attributed to clock H_i is kept constant for sample m+1 and includes the biased variance $var_i(m)$ of sample m. It is written as:

$$p_i(m+1) = [var_i(m)]^{-1} \quad (14)$$

and corresponds to the relative weight:

$$w_i(m+1) = \frac{1}{var_i(m)} \left[\Sigma_k \frac{1}{var_k(m)}\right]^{-1}. \quad (15)$$

For our illustration, (13) (14) and (15) are iterated from m=0 to m=11 with and without the implementation of an upper limit of weight. The procedure was initialized with:

$$w_i(0) = [1/\sigma_i^2][\Sigma_k 1/\sigma_k^2]^{-1}, \quad (15.a)$$

which corresponds to clock H_i true quality.

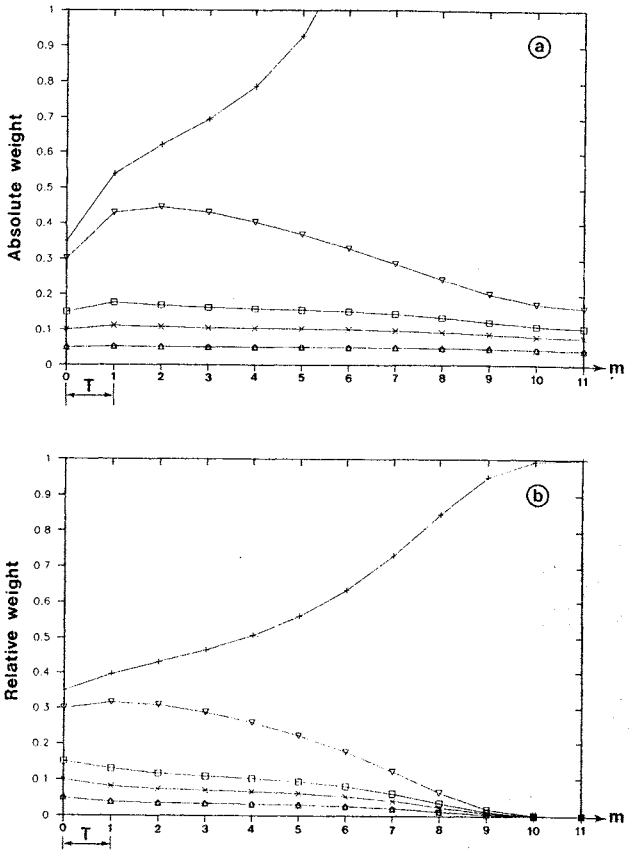


FIGURE 1: *Evolution of biased weights of six clocks simulated with known and constant true variances.*
a - absolute weight
b - relative weight

The results are shown in Figs. 1, 2 and 3 with:

$$\sigma_1^2 = \frac{1}{0.05}, \quad \sigma_2^2 = \frac{1}{0.05}, \quad \sigma_3^2 = \frac{1}{0.1},$$

$$\sigma_4^2 = \frac{1}{0.15}, \quad \sigma_5^2 = \frac{1}{0.3} \text{ and } \sigma_6^2 = \frac{1}{0.35}.$$

Clock H_6 is the most stable and the relation:

$$\Sigma_i \frac{1}{\sigma_i^2} = 1$$

is satisfied.

In Figs. 1a and 1b, no upper limit of weight is implemented. The absolute biased weight of the best clock (here clock H_6) grows indefinitely with time. In terms of relative weight, the best clock gains influence until the others cease effectively to contribute. This situation does not correspond to the optimal clock weights and is undesirable both for stability and for reliability. Thus, if no account is taken of the clock-ensemble correlation, it is necessary to limit the contribution of individual clocks to the ensemble time. This limit can be absolute or relative.

In Figs. 2a and 2b, an absolute cut-off, equal to 0.7, is implemented. The first iterations m=0, 1, 2 and 3 show growth of the absolute weight of the best clock (clock H_6) till the limit is reached (m=4). This is in agreement with Fig. 1a. One might think that, as the sum of the absolute weights is not fixed, the second best clock (clock H_5) would also increase its absolute weight to the cut-off value. This is not observed in our example which shows that the absolute weight of clock H_5 is already decreasing (m=3 and 4) when H_6 reaches the upper limit. This means that the time scale is already controlled by clock H_6 when the absolute cut-off comes into action. The upper limit is thus too high to fulfil its role in providing reliability. In succeeding iterations, the absolute weight of clock H_6 remains at the cut-off level and a stationary state is set up. The constant weight assigned to each clock differs from that computed from its true variance and so degrades the time scale stability. It also depends on the chosen value for the upper limit.
Because the stationary state fixes absolute weights, the absolute cut-off also induces a relative cut-off. In consequence our conclusions are identical when relative contributions, Fig. 1b, are examined.

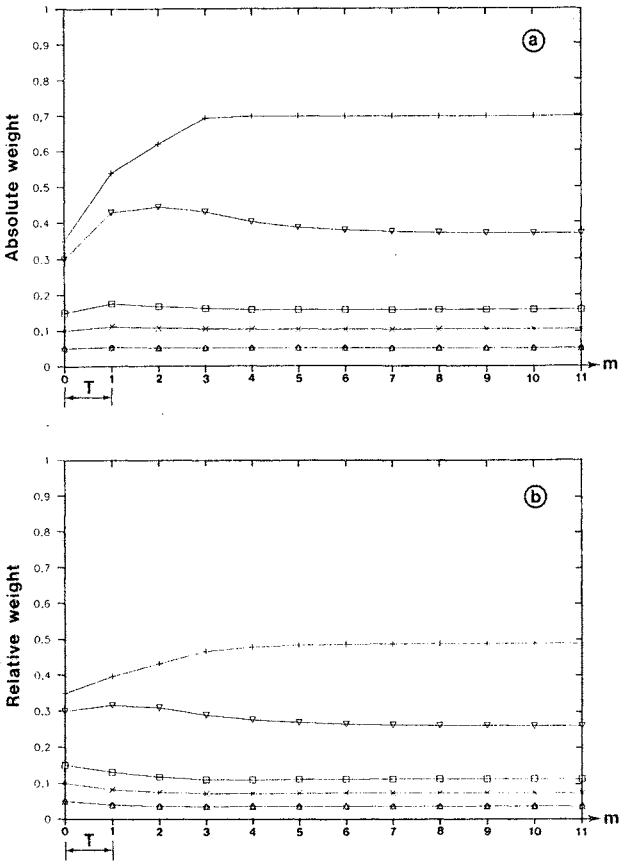
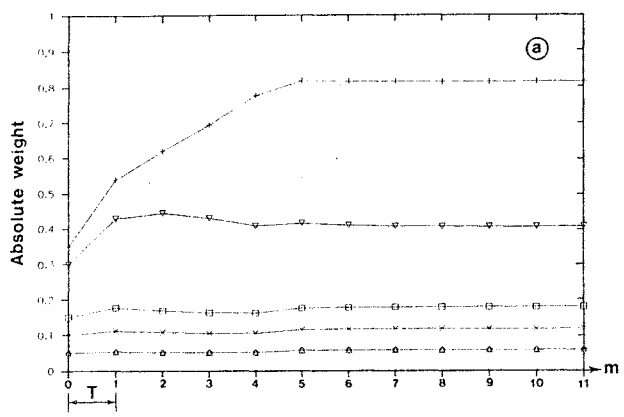


FIGURE 2: *Evolution of biased weights of six clocks simulated with known and constant true variances, with the implementation of an upper limit of absolute weight equal to 0,7.*
a - absolute weight
b - relative weight

In Figs. 3a and 3b, the upper limit operates on the relative weights and is set to a maximum contribution of 50%. The best clock reaches the limit at m=4 according to Fig. 1b. As this clock has already controlled the time scale, it retains the maximum contribution for succeeding iterations. The sum of the contributions of the other clocks is necessarily limited to 50% but a redistribution of relative weights occurs at the next iteration (m=5). As no other clock reaches then the cut-off, a stationary state is set up with constant weights which differ from those expected from the true variances (Fig. 3b). The same features can be observed on the graph of absolute weights (Fig. 3a).



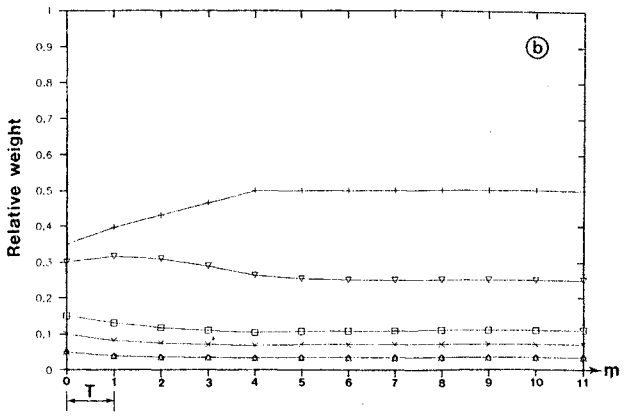


FIGURE 3: *Evolution of biased weights of six clocks simulated with known and constant true variances, with the implementation of an upper limit of relative weight equal to 0,5.*
a - absolute weight
b - relative weight

To conclude, if not corrected, the clock-ensemble correlation biases the weights so that the requirements of stability and reliability are not fulfilled. The implementation of an upper limit of absolute weight or of relative contribution helps to control this effect and hence to improve the time scale stability. However it is not sufficient to ensure that the clocks are represented at their right levels in the weighted average.

2. NUMERICAL EXPERIMENT

The numerical experiment described in this section acts on real data collected from about 180 clocks distributed worldwide, and uses the algorithm ALGOS from which the International Atomic Time TAI is computed. The description of the TAI algorithm can be found in Refs. [1] and [7]. It is an iterative and post-processed algorithm designed to ensure the best long-term stability of the resulting time scale EAL (Echelle atomique libre) from which TAI is deduced by frequency steering. The ALGOS weighting procedure is briefly reviewed in the following paragraph.

2.1 ALGOS WEIGHTING PROCEDURE

ALGOS post-processes two-month blocks of clock data. The frequency $y_i(t)$ relative to EAL of a given clock H_i is computed on the τ=2 month interval ending at date t. The absolute weight $p_i(t)$ attributed to clock H_i, is inversely proportional to the classical variance of its six last two-month frequencies y_i. As the clock-ensemble correlation effect is not taken into account in the ALGOS weighting procedure, this variance is biased and p_i is written as:

$$p_i = \frac{1000}{var_i}, \qquad (16)$$

where the date identification t has been dropped. The use of a classical variance over six two-month intervals allows us to observe clocks over a one-year period and then to deweight clocks with high seasonal variations and high drifts. The iterative procedure of ALGOS allows us to include the current frequency relative to EAL of a clock in the computation of its current weight [1].

An upper limit of absolute weight is implemented in ALGOS. This upper limit is set at p_{imax}=100 which corresponds to a square root of the clock frequency variance equal to 3.16ns/day ($3.7\ 10^{-14}$). With this choice, the maximum relative contribution of a given clock to EAL was limited to about 2% for the years 1988-89-90 and about 12% of the clocks reached this upper limit.

New clocks which enter the EAL computation are kept under observation, with a nul weight, during their two first two-month intervals of continuous operation. For the three following two-month intervals, they are given a weight obtained from an extrapolated six-sample classical variance, assuming random walk frequency noise.

A procedure for abnormal clock behaviour is implemented in ALGOS. It turns to zero the weight of a clock for which a non-intentional frequency step occurs during the two-month interval under computation [7].

2.2 Implementation in ALGOS of a correction factor for the clock-ensemble correlation

The correction factor, to be implemented in ALGOS to take account of the clock-ensemble correlation is deduced from (12).

The corrected absolute weight of clock H_i is then written as:

$$p_{icor} = \frac{1000}{var_i}\left(1 - \frac{1/var_i}{\Sigma_k 1/var_k}\right), \qquad (17)$$

which includes the relative contribution of clock H_i for the two-month interval under computation.

All other characteristics of the ALGOS weight determination are retained without change.

Here we implement (17) in ALGOS and run it with the total clock set and with a small subset for the years 1988-89-90. The frequencies and stabilities of the resulting time scales are analysed by comparison with the primary frequency standard PTB CS1. The clock weight distribution is also examined for each case.

2.3 Numerical experiment with the total clock set

Because 180 clocks are used and because each clock contribution to the time scale is limited to at most 0.02, the true variance of the time scale σ_{EAL}^2, given in (7) and (11), and the covariance between any clock H_i and the ensemble time EAL, $\sigma_{i,EAL}$ given in (8), are small when compared with the frequency variance σ_i^2 of clock H_i. The effect of the clock-ensemble correlation should thus be negligible and (12) reduces to:

$$var_i = \sigma_i^2. \qquad (18)$$

The frequency of TAI relative to the primary frequency standard PTB CS1 is shown in Fig. 4 for the years 1988-89-90. The two curves obtained with ALGOS either corrected or not corrected for clock-ensemble correlation, are nearly superimposed which clearly demonstrates, as expected, that the addition of the correction factor for the weights has negligible effect. The computation of the Allan deviation of TAI relative to PTB CS1, $\sigma_y(\tau)$ with τ=2 months, remains unchanged with the correction and is equal to $1.2\ 10^{-14}$.

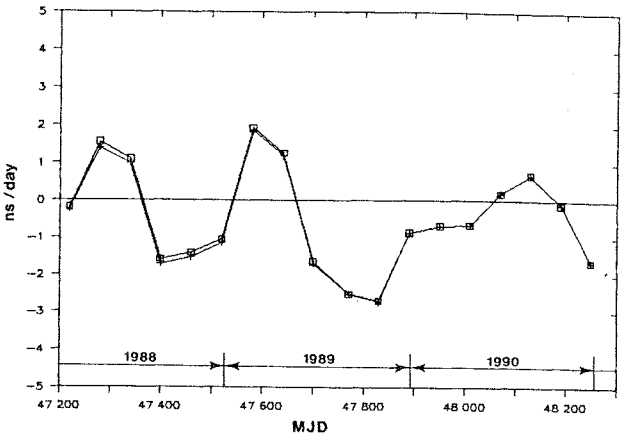


FIGURE 4: *Frequency of TAI relative to PTB-CS1 for years 1988-89-90.*

+ : values as published in Circular T (BIPM) and corresponding to σ_y(60 days) = 1,19 × 10^{-14}

□ : values obtained with the implementation of the correction for the clock-ensemble correlation and corresponding to σ_y(60 days) = 1,20 × 10^{-14}

When the complete clock set is involved, the clock-ensemble correlation plays a negligible role in TAI computation. This correction was introduced at the beginning of TAI computation but was abandoned after some years as the number of clocks increased [3] [8]. This is fully justified.

2.4 Numerical experiment with a small number of clocks

The same numerical experiment was repeated with sets of about 10 and about 20 clocks, chosen from those contributing to TAI. The two sets do not overlap and were chosen so that the particular situations, recognized in ALGOS weighting procedure, effectively occur during the period under study: abnormal clock behaviours, clock exits and entries, intervention of the upper limit of weight, seasonal fluctuations and drifts.

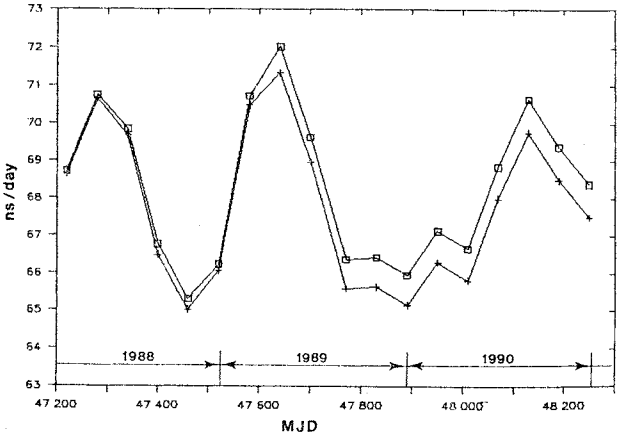


FIGURE 5: *Frequency relative to PTB-CS1, for years 1988-89-90, of an ensemble time scale TA computed from about 10 clocks*

+ : values obtained with TAI algorithm and corresponding to σ_y(60 days) = 1,64 × 10^{-14}

□ : values obtained with TAI algorithm with the implementation of the correction for the clock-ensemble correlation and corresponding to σ_y(60 days) = 1,64 × 10^{-14}

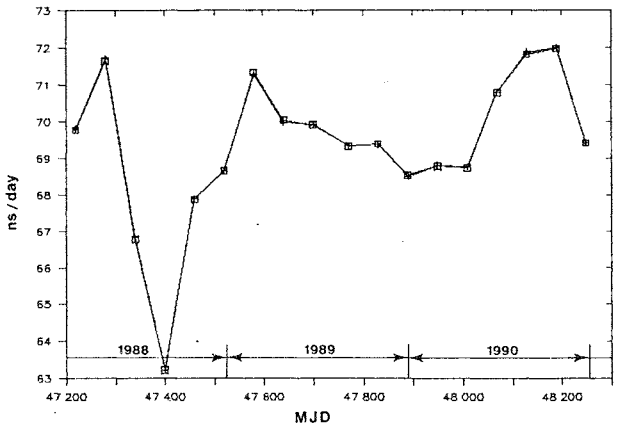


FIGURE 6: *Frequency relative to PTB-CS1, for years 1988-89-90, of an ensemble time scale TA computed from about 20 clocks.*

+ : values obtained with TAI algorithm and corresponding to $\sigma_y(60\ days) = 1{,}83 \times 10^{-14}$

□ : values obtained with TAI algorithm with the implementation of the correction for the clock-ensemble correlation and corresponding to $\sigma_y(60\ days) = 1{,}82 \times 10^{-14}$

The frequencies relative to PTB CS1 of the two resulting time scales, TA_{10} and TA_{20}, are shown in Figs. 5 and 6, with and without implementation of the correction factor to the weights. It appears that a small difference occurs only with the smaller number of clocks. This difference, which can be seen on the frequency graph of TA_{10}, is not significant as far as the stability of the time scale is concerned. The computation of the Allan deviation gives the value 1.64 10^{-14} for TA_{10}, with or without taking into account the clock-ensemble correlation.

To analyze this quite surprising result in more detail, we show in Tables 1 and 2 the absolute weights attributed by ALGOS to the clocks contributing to TA_{10}. For each clock, the weight sequence contains at most 18 values, corresponding to the 18 two-month intervals of years 1988-89-90. Tables 1 and 2 exhibit different clock behaviour:

(i) Clocks 1, 3, 6 and 8 remain at the upper limit of weight for a long period. The correction factor for clock-ensemble correlation decreases their absolute weight, obtained from (16), before the intervention of the upper limit, but not enough to deweight them to a value lower than the upper limit. This is because the value chosen for the upper limit is set low to ensure that TAI relies on a sufficient number of clocks. These clocks control the time scale and make inefficient the correction for the clock-ensemble correlation.

(ii) With or without the correction, new clocks entering the ensemble have very low weight. This occurs to many clocks during the three years under study.

(iii) Abnormal behaviour detection operated for clocks 4, 7 and 9. Correction for clock-ensemble correlation has no effect in this case.

(iiii) Some clocks belong to the middle classes of weight. Clearly, they are not the most stable clocks and have probably been deweighted because of seasonal frequency fluctuations or frequency drifts. Here, the correction has some impact as can be observed for instance on clock 5. Nevertheless, the absolute weight changes correspond to an amount of a few units and are not sufficient to significantly change the time scale stability.

To conclude, the weighting procedure in ALGOS is adapted to real situations where systematic frequency variations must be recognized, where frequency steps must be detected, where new clocks must be entered with some care and where an upper limit of weight must be reasonably chosen. The consequences of such constraints appear clearly in the typical ALGOS weight distribution shown, for instance, in Fig. 7: a rather large number of clocks remain at the upper limit or close to zero while a small fraction of clocks have middle weights. The choices made for the ALGOS weighting procedure operate so that the impact of the clock-ensemble correlation is not as crucial as might be expected from a purely theoretical point of view, even when only a small amount of clocks is involved.

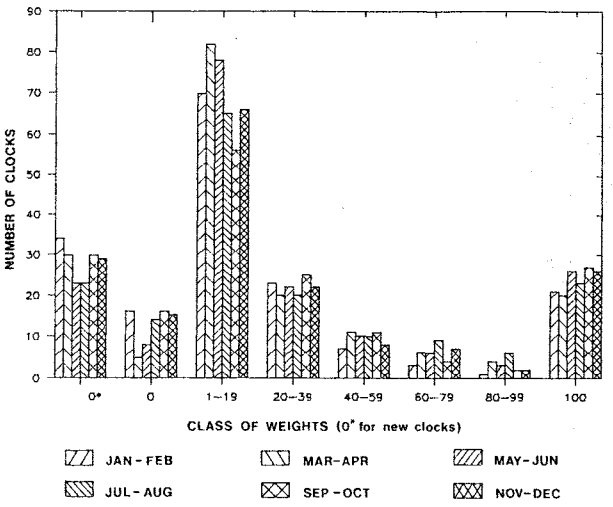


FIGURE 7: *Histograms of TAI absolute weight distributions for the six TAI computations of year 1989. (Class 0* corresponds to clock entries in the ensemble).*

Note: Equations (12) and (17) are obtained with constant relative clock weights for the whole period used to compute frequency variances. This is not the case for most time scales, where weights change at each interval. Equation (17) is then approximate. An analytic expression for the clock-ensemble correlation factor in the case of varying weights should involve average weights, and corresponding standard deviations, as well as clock frequency prediction which is essential to ensure time scale continuity. The generalisation of (17) to this case is not trivial but work is in progress.

Table 1: *Absolute weights assigned to clocks of a small ensemble from TAI algorithm*

Clock	1988						1989						1990					
1	100	100	100	100	100	100	85	48	50	58	100	100	100	100	100	100	100	100
2	6	7	11	9	9	9												
3	93	100	100	100	100	100				0	0	3						0
4	8	7	6	8	6	4	5	5	5	0	2	1			0	0	1	1
5	0	0	26	41	65	77												
6		0	0	55	51	75	98	71	100	100	100	100	100	100	100	100	100	100
7		0	0	29	16	18	24	17	0	3	3	3	5	23	22	32	100	100
8		0	0	23	35	31	43	39	91	91	98	100			0	0	25	49
9			0	0	43	31	0	5	6	7	14	0	3	2	1	1	1	2
10					0	0												
11					0	0												
12							0	0	1	1	1	2	4	13	12	69	79	63
13													0	0	11	18	27	40
14																	0	0

Table 2: *Absolute weights assigned to clocks of a small ensemble from TAI algorithm with the correction for the clock-ensemble correlation*

Clock	1988						1989						1990					
1	100	100	100	100	100	100	100	65	60	74	100	100	100	100	100	100	100	100
2	6	7	11	9	9	9												
3	100	100	100	100	100	100				0	0	3						0
4	9	8	6	8	6	4	5	5	5	0	2	1			0	0	1	1
5	0	0	27	45	72	89												
6		0	0	65	57	88	100	100	100	100	100	100	100	100	100	100	100	100
7		0	0	29	16	19	25	18	0	4	3	3	5	24	22	32	100	100
8		0	0	24	38	33	47	46	100	100	100	100			0	0	25	50
9			0	0	48	33	0	6	6	8	15	0	3	2	1	1	1	2
10					0	0												
11					0	0												
12							0	0	1	1	1	2	4	14	12	74	85	65
13													0	0	12	19	28	40
14																	0	0

CONCLUSIONS

In the computation of a time scale as a weighted average of clock readings, the clock-ensemble correlation biases the contribution of each clock, optimally defined as inversely proportional to the clock frequency variance. This effect degrades the stability and reliability of the time scale. It can be partially controlled, but not completely removed, with the implementation of an upper limit of weight.

An analytic expression for the correction to biased weights can be developed when the clock weights remain constant over the period of estimation of the frequency variances. This expression can also be used as an approximation in ensemble time computation when weights change at each interval.

The implementation of this approximate correction factor in TAI algorithm does not significantly change the weight distribution and the time scale stability, even if a small number of clocks is involved. This is so because the weighting procedure of the TAI algorithm is adapted to observable systematic effects. In the general case of an ensemble time generated from a large number of clocks, the clock-ensemble correlation effect should be negligible.

REFERENCES

[1] P. Tavella and C. Thomas, 'Comparative study of time scale algorithms', 1991, *Metrologia*, in press

[2] D. W. Allan, J. E. Gray and H. E. Machlan, 'The NBS atomic time scale: generation, stability, accuracy and accessibility', *NBS Monograph 140*, p. 218, 1974

[3] M. Granveaud, 'Echelles de temps atomique', *Monographie du Bureau National de Métrologie*, 1986, Editions Chiron, pp. 47-50

[4] S. R. Stein, G. A. Gifford and L. A. Breakiron, 'Report on the time scale algorithm test bed at USNO', *Proc. of the 21th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 1989, pp. 269-288

[5] K. Yoshimura, 'Calculation of unbiased clock-variances in uncalibrated atomic time scale algorithms', *Metrologia* 16, 1980, pp. 133-139

[6] B. Guinot, 'Some properties of algorithms for atomic time scales', *Metrologia* 24, 1987, pp. 195-198

[7] B. Guinot and C. Thomas, 'Establishment of the International Atomic Time', *Annual Report of the Time Section of the BIPM*, 1988, pp. D3-D22

[8] B. Guinot, personal communication

5th EUROPEAN FREQUENCY AND TIME FORUM BESANÇON - 12-13-14 MARCH 1991

Session 8B Number 79

Promise into Practice: Implementing TA2 on Real Clocks at NIST

M. A. Weiss and T. Weissert

Time and Frequency Division, 847.5
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303, U.S.A.

Abstract
TA2 is a time scale algorithm designed to run as a post-processor by smoothing, i.e. combining estimates from runs in both forward and backward temporal directions. It has been tested using the time-scale data at NIST from 1989 through the first half of 1990. We review the characterization parameters for clocks with white FM, σ_α, and random-walk FM, σ_β, and the dependence of these estimates on integration time, τ. We also discuss the transformation of these parameters from their use as time deviations to frequency variances. We compare the scale generated by TA2 over 541 d starting at the beginning of 1989, with the two other scales at NIST, AT1 and TA. We find that TA2 and AT1 agree well in frequency, though TA, the pure Kalman-filter algorithm, walks off more than 2 parts in 10^{13} over the interval. Using GPS data to link with other time scales around the world, we compute the stabilities of the three NIST scales, AT1, TA, and TA2 using N-cornered hat estimates. We find that they have similar short-term stabilities for integration times out to 16 d, within the confidence of our estimates. TA appears to be worse in stability than the other two scales by a factor of about 2. There is some indication that TA2 improves slightly over AT1 in the long term, matching previous results from simulation, though the confidence on these estimates is poor.

Introduction
As we reported previously [1], TA2 is a smoother, an algorithm that runs both forward and backward in time. First we estimate frequency using the AT2 algorithm [2] in both directions. Then we combine these frequency estimates in a third (forward) pass to estimate time, using only the time estimate portions of AT2. We are now ready to run TA2 at NIST in parallel with the other scales, AT1 [2] and TA [3]. One of our goals with this algorithm is to improve long term frequency stability [4], using the advantage of post-processing with a smoother.

Please note that in this paper we will use the same name for a time scale algorithm and the time scale it produces. Thus we refer to TA2, TA, and AT1 both as algorithms and as the time scale output from a run on a particular ensemble of clocks over a particular interval. We trust that the context will keep these things clear.

AT2, the single-pass algorithm underlying TA2, is designed to generate a time scale in real time that is more stable than the best clock in the ensemble, both in long and short term. For each clock in the ensemble, it adaptively estimates the clock's weight and a confidence on the frequency estimate. It detects frequency steps using the frequency estimate confidence [2].

We run TA2 over a period of 1 month plus 10 d and use only the data from the month as definitive results. Thus, all definitive results have the benefit of information from both the past and the future. The clock characterization parameters at the end of the month from the third pass are saved for initializing the forward pass for the next month. We run month-by-month over past data for two reasons. First, we plan to run TA2 operationally after the tenth day of each month for the previous month. Also, as was discussed previously [1], the random walk of the scale itself creates problems for combining the forward pass estimate of frequency with that of the backward pass, and these problems become more serious as the length of the pass grows. For the NIST clocks, we find that the amount the frequency of the scale walks off from the beginning of the month through the end and back again does not cause significant problems.

While installing TA2 in the NIST time-scale system, we encountered several problems concerning the parameters we use to characterize the clocks. The time interval between measurements is 2 h, as compared to the 1 d interval we had used in developing the algorithms for GPS time-transfer data. Also, in a real ensemble, clocks must enter and leave the scale at arbitrary times. Both of these factors made us aware of subtle considerations with our new algorithm, whose resolution we report in this paper.

We also present the results of running TA2 on the clocks at NIST over the period from January 1989 through June 1990. 1989 was chosen to provide a rigorous test of the scale because of a failure of the measurement system which occurred in July of that year. Both the time and frequency offsets of each clock in the ensemble had to be estimated to restart the time scales after this failure. Apparently we succeed in this test because there seems to be no evidence of the discontinuity in the measurements in the resultant scale.

To study the short- and long-term stability of TA2 from this run, we compare it with both the other local scales at NIST, AT1 and TA, and with scales at other international laboratories using GPS time transfer. AT1 is the real-time time scale at NIST used to calibrate various oscillators through the NIST two-frequency measurement system [6], and to generate frequency and time signals

electrically through a microphase stepper. Like all the scales at NIST, and other scales such as ALGOS [5], AT1 estimates time as a weighted average of the measured clock differences [2]. In AT1, the weights are adaptively estimated in an attempt to optimize short term stability. Frequency is estimated using exponential filters with time constants chosen differently for each clock to optimize long term stability. TA is a Kalman filter, run once a month as a post-processor to generate a scale over the previous month [6]. Recently, design problems have been discussed concerning the use of a pure Kalman filter as a time-scale algorithm [2]. We show results here comparing TA to AT1 and TA2 which suggest that TA has been walking off in frequency faster than AT1 and TA2. By comparing the scales with other international scales using an N-cornered hat analysis, we find that the stability of TA becomes significantly worse than AT1 and TA2 by a factor of 2 at an integration time of about 128 d. TA2 is slightly better than AT1 at 128 d and appears to improve even more at 256 d. The confidence on both of these points is poor enough that we perhaps should wait for more data to draw hard conclusions, but the improvement of TA2 over AT1 at 128 d matches closely results predicted from simulation [1], hence giving us more confidence in our results. We also note no degradation in short term stability even with good long term performance.

Anomalous Behavior Detection

The TA2 algorithm could be applied to generate an ensemble from any group of standards, such as voltage, time, or length, but the way it detects and responds to anomalous behavior is what makes it particularly adapted to clocks, and hence to generating a time scale. By characterizing anomalous behavior in another type of standard, and then designing and implementing an algorithm for detecting and responding to such behavior, we could customize this algorithm for application to a set of that type of standard.

The anomalous clock behavior with which we are concerned consists of time steps and frequency steps, which of course must be detected in the presence of the usual stochastic processes of the clocks. In addition, these stochastic processes can change. Defining these in a language independent of clocks, we would say that a time step is a change in the estimated value of the standard's offset from the ensemble, while a frequency step is a change in the estimated slope. In either case, the step is defined only when the magnitude significantly exceeds the changes due to the stochastic processes. This becomes complicated since we assume both white noise and random-walk deviations of the slope. Hence we must estimate the normal size of these stochastic deviations to detect and respond to steps. Finally, for clock steps, we assume the change in value is sudden and final. Though some mechanisms have been postulated for clocks for the gradual decay of a sudden change in slope (a frequency step) we assume no decay mechanism when presented with a particular frequency step since we use no information to correlate that step with a causing mechanism. It is very rare for a time step to decay. In other standards, such as standard chemical voltage cells, one might want to include decay mechanisms in the anomalous behavior algorithm..

Clock Parameters in TA2

Clocks in TA2 are characterized with the same deterministic and stochastic parameters as TA [7]. The deterministic parameters are time, $\underline{x}$, frequency, $\underline{y}$, and frequency drift offsets from the scale. The stochastic parameters are white and random-walk frequency modulation (FM), denoted σ_α and σ_β, respectively. These parameters represent the expected size of the noise impulses driving their respective noise process. Thus σ_β represents the expected size of the white noise impulses which are integrated to form the random walk process in $\underline{y}$. We estimate the values of these parameters as the expected amount of time deviation due to the particular noise process after a set time interval. In TA2, time and frequency offsets and white FM are estimated adaptively, whereas frequency drift and random-walk FM, both of which deal with long term phenomena, are set by the operator. In addition, the algorithm estimates a confidence, $\underline{P}$, on $\underline{y}$, based on σ_α and σ_β, which responds to the adaptability of σ_α.

While there is provision in TA2 for using frequency drift parameters, in practice we rarely use them. The benefit of using frequency drift parameters should be to improve the long-term stability of the scale, since it should allow us to see the smaller random walk of clocks underneath the drift. The problem is that the drift of very few systems remains constant long enough to be useful. Also, drift in a cesium clock is a sign of pathology, in which case the clock probably will not last to contribute long term stability.

Let us discuss aspects of the stochastic parameters. Giving σ_α the correct units requires some care. We enter σ_α initially as the expected time deviation due to the white FM process after a time interval of τ_0 = 1 d. We then use σ_α in the algorithm in different ways. σ_α is estimated adaptively in the algorithm from the predictability of the time offset, just as in AT1, assuming the measurement interval, τ, is short enough that the white FM process of the clock is dominant. The interval τ = 1/12 d is different than τ_0, the interval used in estimating σ_α for entry into the algorithm. Thus σ_α must be converted from the value as entered to initialize the adaptive, exponentially filtered estimate of σ_α.

Similarly, σ_β is enterred as a time deviation due to random-walk FM at τ_0. It is used internally as the expected frequency variation due to random-walk FM over the measurement interval. Thus it too must be converted. These conversions are shown in Table I, where, for the frequency variance, we have simply solved for $\sigma_y^2(\tau)$ with the proper power law dependence on τ.

For initial entry, σ_α and σ_β can be estimated [3] from the Allan variance $\sigma_y(\tau)$ as follows [8]:

$$\sigma_y^2(N\tau_0) = \frac{\sigma_\alpha^2}{N\tau_0^2} + \frac{\sigma_\beta^2(2N^2+1)}{6N\tau_0^2}. \qquad (1)$$

This equation gives σ_α and σ_β as the time dispersion due to the specific noise types after the interval τ_0. Please note that, in this paper, τ_0 is the time interval which applies to σ_α and σ_β when entered externally, while τ is the time interval between measurements. For TA2, σ_α and σ_β are entered estimated at τ_0 = 1 d, while the time between measurements at NIST is τ = 1/12 d.

We give an example of the use of equation 1 for estimating σ_α and σ_β from the Allan variance $\sigma_y^2(\tau)$ for a particular clock, the HP601 at NIST. Figure 1 gives the Allan deviation of the HP601 over the 541 d starting January 1989. Two lines are drawn on this plot, one for the white FM component of the variance with a slope of $-\frac{1}{2}$ and one for the random-walk FM component with a slope of $+\frac{1}{2}$. We have then marked that the intercepts of these lines at the one day point are $2.7*10^{-14}$ for the white FM and $0.61*10^{-14}$ for the random-walk FM. Applying equation (1) for each of σ_α with τ_0= $0.864*10^{14}$ ns we find σ_α= 1.1ns. For σ_β we must realize that the line we have drawn is the asymptotic form of equation 1 for large N. In this case we have

$$\sigma_\beta^2|_\tau = 3*\tau*\sigma_y^2(\tau). \qquad (2)$$

Substituting the above values in (2) we have σ_β = 1.6ns.

In steady-state TA2, we use our estimate σ_α in two different ways: for determining the relative weights of the clocks for the time update, and for filtering the white FM in our estimate of frequency. The value of σ_α is also involved in the anomalous behavior routines. Since the measurement interval is fixed for TA2 at NIST, the value of τ is not important in determining the relative weights of the clocks. However, in order to use σ_α as the estimate of the white FM level in filtering the frequency, we must transform it from being a time deviation to a frequency variance. σ_β must also be adjusted from its being time deviation at τ_0 as entered to its use as a frequency variance at τ. The relations we use for these are shown in Table 1.

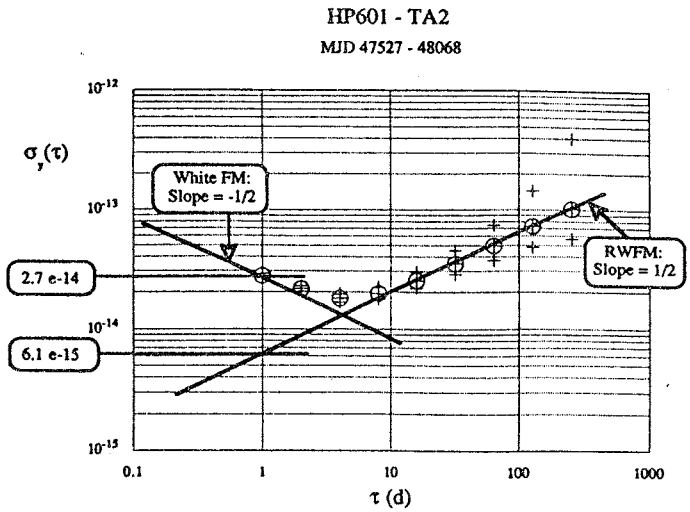


Figure 1. Allan deviation plot of the NIST clock HP601 over 541 d starting January 1989 showing white FM and random-walk FM components and levels.

A final note on the application of the stochastic clock parameters concerns sudden increases in σ_α initiated at certain times to avoid transients. There are four such times: 1) when starting a run for a new month based on estimates of parameters from the previous month, 2) when a clock is first entered as a new clock, 3) when a frequency step is found, and 4) when a clock goes off line and comes back on in one month. These increases in σ_α are necessary because the weights of the clocks in the time update are proportional to $1/\sigma_\alpha^2$, thus it is necessary to increase σ_α to de-weight a clock. The factors we use to increase σ_α in these cases are purely phenomenological; we basically picked values that worked.

When starting a run for a new month, the estimates of σ_α^2 for clocks from the previous month are multiplied by a factor of 1.5, because the frequency estimates from the end of the last month are smoothed estimates based on data including 10 days into the current month. Hence because there is more

Table I

	σ_α	σ_β
Entered as:	$\sigma_\alpha\|_{\tau 0}$	$\sigma_\beta\|_{\tau 0}$
Used as: time variance	$\sigma_\alpha^2\|_\tau = \frac{\tau}{\tau_0}\cdot\sigma_\alpha^2\|_{\tau_0}$	not used
Used as: frequency variance	$\left(\frac{K}{\tau}\right)^2\cdot\sigma_\alpha^2\|_\tau$	$\frac{K^2}{3\tau_0^2}\cdot\left(\frac{\tau}{\tau_0}\right)\cdot\sigma_\beta^2\|_{\tau_0}$

K in this table determines the units for frequency. For TA2 we use ns/d internally, hence K= $8.64*10^{13}$ ns/d.

information implicitly contained in them than is available for the new first pass, these initial frequencies may not match the raw data well for the forward pass of the current month. The 1.5 factor increase of σ_q^2 compensates for this mismatch by allowing the algorithm a little extra time to learn the forward frequency, yet still starting each clock with the smoothed frequency from the previous month. The algorithm remembers the old and learns the new frequency of a clock using an exponential filter. Thus the old frequency is gradually forgotten with a "half-life" equal to the exponential filter time constant. This time constant adjusts adaptively with σ_q, and should equal approximately the minimum of the Allan variance curve, the integration time when the dominant noise form changes from white FM to random-walk FM. The factor 1.5 adjustment at the beginning of a month additionally makes the definitive (third pass) estimates for the beginning of the month more influenced by the backward run with its longer foresight of information than the forward run.

When a clock is first entered as a new clock, even though one may have characterized its parameters well, we find that the scale still needs some time to fine tune its parameters. Toward this end, the initial value of σ_q^2 is multiplied by a factor of 3. Similarly, when a frequency step is detected, the scale is re-run with the stepped clock's weight set to 0 for a period long enough for the algorithm to learn its new frequency; we then re-admit the clock with non-zero weight and multiply the estimated σ_q^2 by a factor of 2.

When a clock goes off line and does not return before the end of the month, it then comes back on line in the reverse direction as a new clock. Alternatively, the clock can come back on line again in the forward direction before the end of the month. In either case, when a clock goes off line and comes back on two corrections are applied to σ_q^2. We multiply by a factor of 2 when it returns because, even though we assume the white FM level of the clock to be nominally the same, we don't want to weight the clock in the same way for the time update since the frequency may have changed in unpredictable ways. We also add to σ_q^2 the amount of dispersion of the scale frequency during the off-line interval. If we denote $\sigma_{\beta x}$ as the random-walk FM deviation of the scale frequency in ns/d at one day, and δ as the time interval for which the clock has been off line, then we add to σ_q^2 the value $\delta^3 * \sigma_{\beta x}^2$, the time domain dispersion of the scale itself over the interval δ due to random-walk FM.

TA2 from January 1989 through June 1990

One of the features of the AT2 algorithm, inherited by TA2, is automatic frequency-step detection and response [1]. When the AT2 algorithm detects a frequency step, it removes that clock from the scale until the scale can learn its new frequency. Although this works automatically in TA2, we can also manually tell the scale when a clock has a step and what its magnitude is. This latter action, though more labor intensive, allows TA2 to keep the clock on line, thus providing a small improvement to the scale's stability. In this action we are using our knowledge of clock behavior after the fact beyond what TA2 will do automatically. In the January 1989 - June 1990 run we estimated and inserted frequency steps by hand whenever the algorithm detected a step in a clock with a large amount of weight. Thus the results we obtained were through a lot of careful grooming of the ensemble.

Over the run, frequency steps were found in seven clocks: one step in two clocks, two steps in one clock, three steps in three clocks and six steps in one clock. Thus a total of nineteen frequency steps were found. There were four other clocks that had no steps, making a total of 11 clocks. All of the frequency steps were estimated and inserted manually except for two in one of the three step clocks and four in the six step clock.

In figure 2a we see the comparison TA2-TA run on the same clocks at NIST over the period MJD 47527-48068, all of 1989 and the first half of 1990. Figure 2b is the frequency plot of this same data. Clearly, there are some frequency excursions between the two scales. Comparing both of these

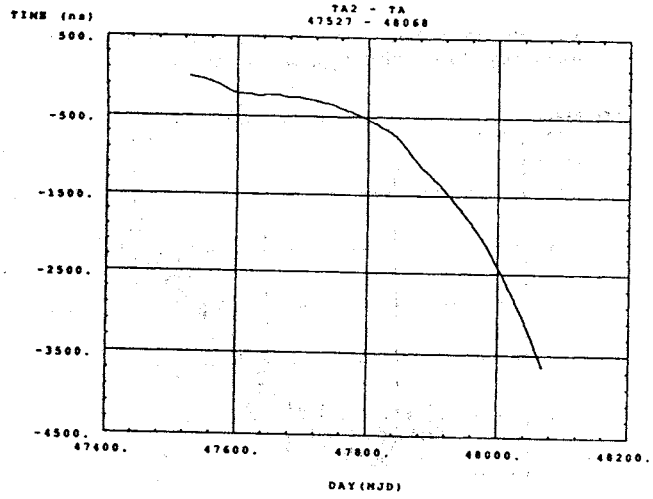


Figure 2a. The time deviation TA2 - TA from January 1989 through June 1990.

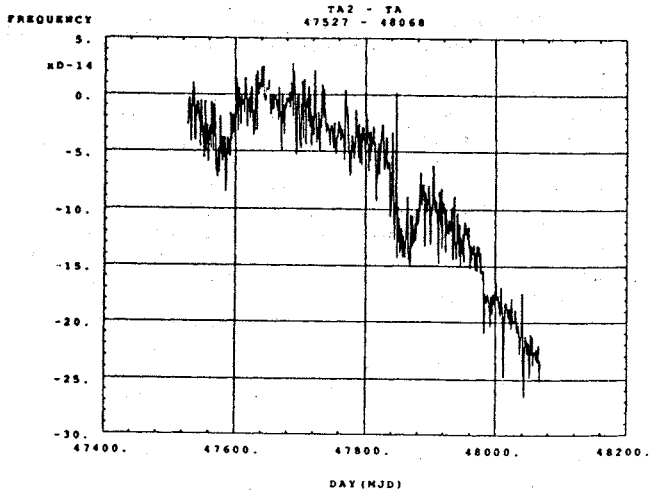


Figure 2b. The frequency deviation TA2 - TA from January 1989 through June 1990.

scales with AT1, TA-AT1 in figures 3a and 3b, and TA2-AT1 in figures 4a and 4b, we see that AT1 and TA2 agree quite well in frequency for the period. Thus we conclude with reasonable probability that the pure Kalman filter algorithm, TA, has walked off more than 2 parts in 10^{13} over this time period.

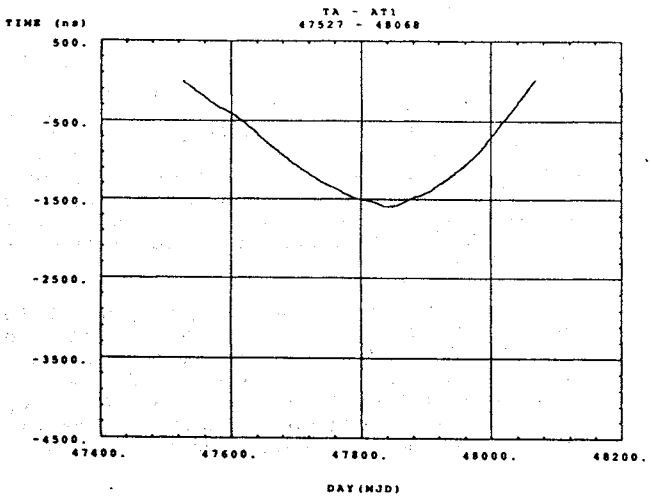


Figure 3a. The time deviation TA - AT1 from January 1989 through June 1990.

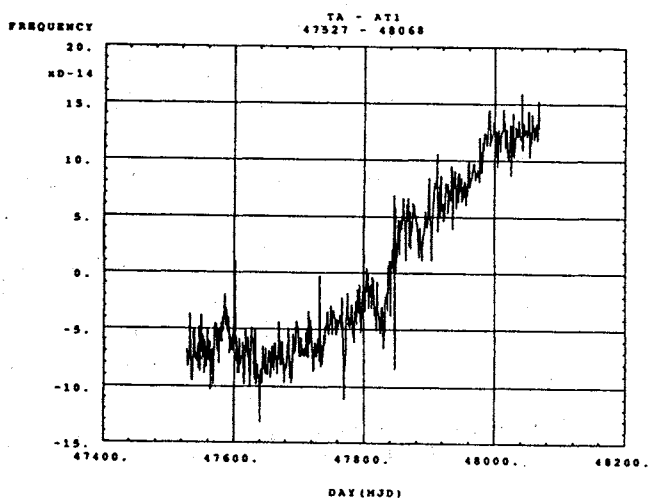


Figure 3b. The frequency deviation TA - AT1 from January 1989 through June 1990.

A second analysis of the results over the 541 d run was performed comparing the NIST time scales with UTC(lab) for several time standards laboratories around the world using GPS data. The laboratories used were the National Research Council (NRC), Ottawa, Canada, l'Observatoire de Paris (OP), Paris, France, the Tokyo Astronomical Observatory (TAO), Tokyo, Japan, the Technical University of Graz (TUG), Graz, Austria, and the United States Naval Observatory (USNO), Washington D.C., U.S.A. We did a type of separation of variance analysis called "N-cornered hat," [9] on all four labs plus each (independently) of AT1, TA and TA2.

The N-cornered hat technique requires that the input time series be independent, which is not true in our case for two reasons. First, since the labs were compared using

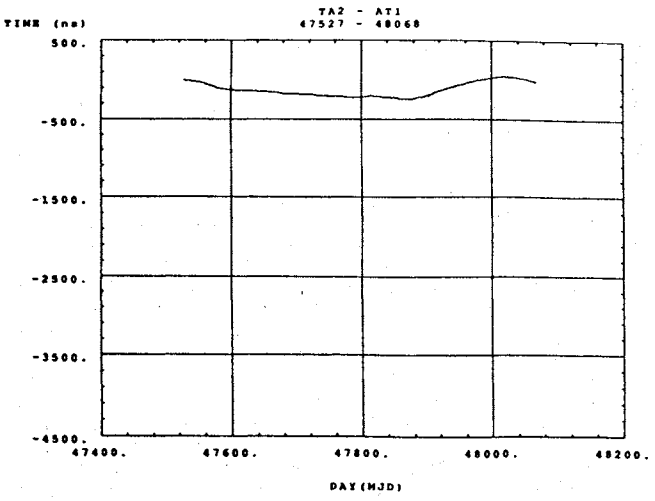


Figure 4a. The time deviation TA2 - AT1 from January 1989 through June 1990.

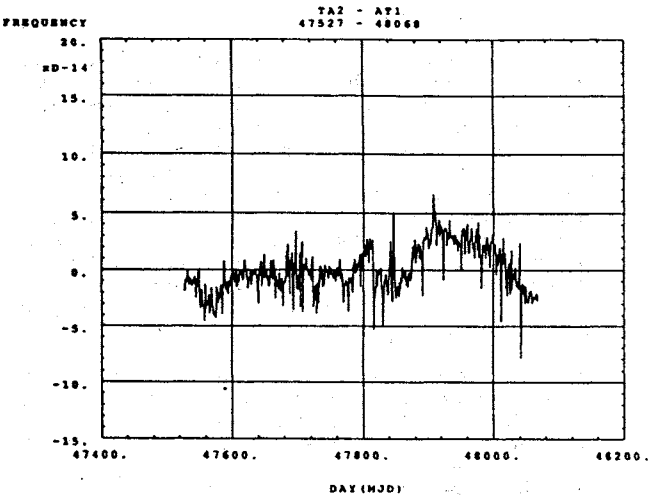


Figure 4b. The frequency deviation TA2 - AT1 from January 1989 through June 1990.

data from GPS satellites in common-view [8], there is the same satellite noise on several of the time transfer links. This would affect short term stability estimates out to perhaps a 4 d integration time. Further, each UTC(lab) is steered to match UTC as generated by the International Bureau of Weights and Measures. The steering, however, is usually slow and of small effect within a year. We know, however, that the USNO applied a lot of steering in 1989 and 1990; eight frequency steps were applied of magnitude 1 ns/d, the first six of which were applied in a series 20 d apart at the beginning of the year. We also note that there can be apparent correlations in the data due to the finite data length.

The results of our 5-cornered hat analysis on all four labs plus each of AT1, TA and TA2 are shown in figures 5, 6, and 7, respectively. We see little difference among the short term stabilities of the three different scales at NIST. We see a significant degradation in the long-term stability of TA over both AT1 and TA2, agreeing with our analysis above concerning TA. There is also a small improvement in TA2 over AT1 at 128 d, and an indication of

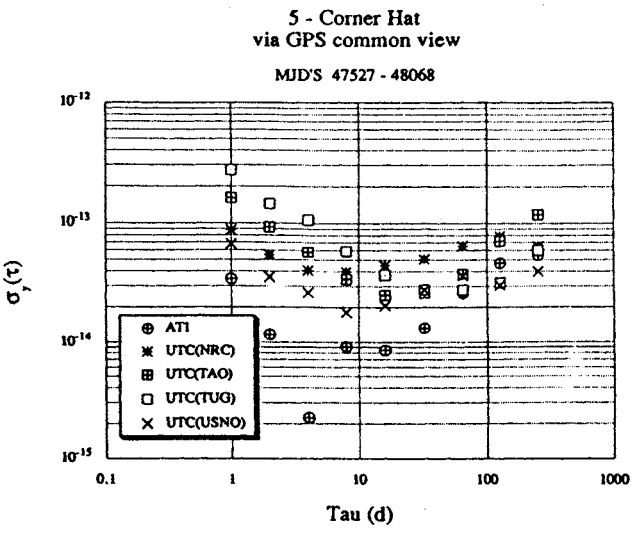


Figure 5. 5-cornered hat Allan deviation estimates for AT1 against UTC(NRC), UTC(TAO), UTC(TUG) and UTC(USNO) from January 1989 through June 1990.

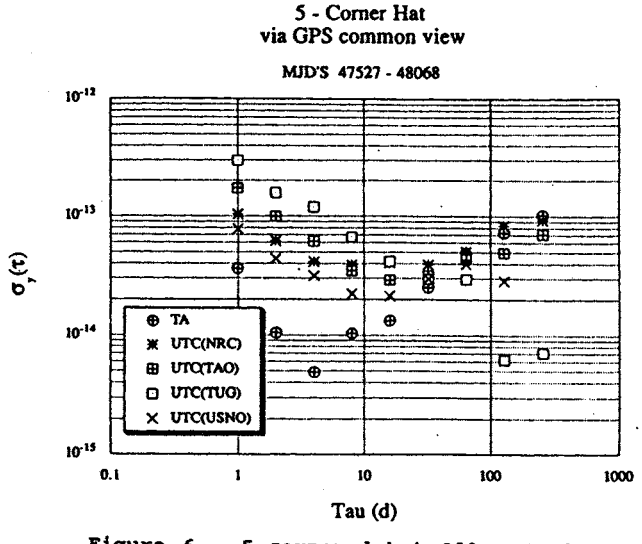


Figure 6. 5-cornered hat Allan deviation estimates for TA against UTC(NRC), UTC(TAO), UTC(TUG) and UTC(USNO) from January 1989 through June 1990.

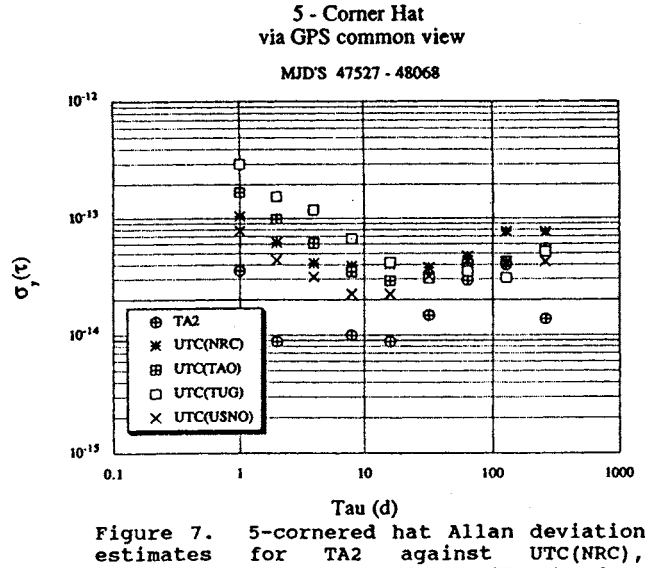


Figure 7. 5-cornered hat Allan deviation estimates for TA2 against UTC(NRC), UTC(TAO), UTC(TUG) and UTC(USNO) from January 1989 through June 1990.

improvement at 256 d, though we have little confidence in this last point. The improvement at 128 d is also probably not statistically significant by itself, but it agrees well with the improvement we expect to find from simulation [1]. It is not known how to estimate confidence bars on N-cornered hat estimates, but the confidence cannot be better than that of a direct Allan deviation computation. These error bars can be seen in figure 1. We conclude that we have some indication that we have realized here in practice the results we predicted in simulation.

Conclusions

We have presented the implementation of TA2 in the NIST time scale system. We discussed how to characterize clocks with white FM and random-walk FM levels, and various relations for these numbers as a function of integration time between treatment as a time deviation or frequency variance. We found little over-all frequency change between AT1 and TA2 over 1989 and the first half of 1990, but a walk in frequency in TA accumulating to more than 2 parts in 10^{13}. Finally, we studied the fractional frequency stability of the three scales in relation to UTC of various labs around the world using an N-cornered hat approach. We note that the stability of TA is worse than AT1 and TA2 by a factor of almost 2 at 128 d. We also conclude that there is indication of an improvement in the stability of TA2 over AT1 matching that predicted by former simulation, though the confidence is poorer than the apparent improvement.

Bibliography

[1] M. A. Weiss and T. Weissert, "TA2, A time scale algorithm for Post-Processing: AT1 plus frequency variance," IEEE Trans. Instrum. Meas., vol. 40, no. 2, April '91, to be published.

[2] M. A. Weiss and T. Weissert, "AT2, A new time scale algorithm: AT1 plus frequency variance," Metrologia, to be published.

[3] R. H. Jones and P. V. Tryon, "Estimating Time From Atomic Clocks," Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 17-24, Jan - Feb 1983.

[4] D. W. Allan, M. A. Weiss, T. K. Peppler, "In Search of the Best Clock," IEEE Trans. on Instrum. and Meas., vol. 38, no. 2, pp. 624-630.

[5] B. Guinot and C. Thomas, "The Establishment of International Atomic Time," Annual Report of the BIPM Time Section, Part D, pp. D1-D22, 1988.

[6] D. W. Allan, D. J. Glaze, J. E. Gray, R. H. Jones, J. Levine, and S. R. Stein, "Recent Improvements in the Atomic Time Scales of the National Bureau of Standards," Proc. of the 15th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 29-40, Naval Research Laboratory, Washington D. C., Dec. 6-8, 1983.

[7] P. V. Tryon and R. H. Jones, "Estimation of Parameters in Models for Cesium Beam Atomic Clocks," Journal of Research of the National Bureau of Standards, vol. 88, no. 1, pp. 3-16, Jan.-Feb. 1983.

[8] J. A. Barnes, R. H. Jones, P. V. Tryon, and D. W. Allan, "Stochastic Models for Atomic Clocks," NASA Conference Publication 2265-- Proc. of the 14th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 295-306, 1982.

[9] J. Barnes,"Time Scale Algorithms Using Kalman Filters -- Insights from Simulation," Proc. 2nd Symposium on Atomic Time Scale Algorithms, NBS, Boulder, CO, June 23-25, 1982.

Session 8B Number 80

PRACTICAL CALCULATION METHODS OF SPECTRAL DENSITY OF INSTANTANEOUS NORMALIZED FREQUENCY DEVIATION FROM INSTANTANEOUS TIME ERROR SAMPLES

F. Vernotte*, J. Groslambert**, J. J. Gagnepain**

*Observatoire de Besançon, 41 bis Avenue de l'Observatoire, BP 1615, 25010 BESANÇON Cedex, FRANCE
**Laboratoire de Physique et Métrologie des Oscillateurs, 32 Avenue de l'Observatoire, 25000 BESANÇON

ABSTRACT.

In order to measure the different types of noises present in a signal, it is generally necessary to obtain samples of the instantaneous normalized frequency deviation y(t), or, in the frequency domain, of the spectral density $S_y(f)$. However, these quantities are not observable, and one have to measure other quantities, which are observable, and which are linked to the others by well-known relationships.

Thus, with a time interval counter, it is possible to measure directly the quasi instaneous time error x(t) between two signals. Since y(t) is the time derivative of x(t), it is possible to calculate $\overline{y_k}$ samples with x(t) values. If the spectral density $S_y(f)$ is wanted thus just the Fourier transform of $\overline{y_k}$ samples are calculated. On the other hand, the spectral density of instantaneous normalized frequency deviation $S_y(f)$ must be directly obtained from the spectral density of instantaneous time error $S_x(f)$, which can be calculated by Fourier transform of x(t) data.

However, the results of these calculations may be altered if certain precautions are not taken into account. These alterations are due to the use of Fast Fourier Transform of mean-values which corresponds with the theoretical Fourier transform only if there is an infinite number of samples and if the time interval between two samples tends toward zero. Nevertheless, it is possible to solve this problem by defining non restrictive hypothesis, for example, assuming that the measure sequence is periodic.

The precautions inherent in these two calculation methods and their results will be discussed.

I. TWO CALCULATION METHODS OF THE SPECTRAL DENSITY $S_y(f)$.

I.1. Calculation of frequency deviation samples with time error samples.

Instantaneous time error x(t) and instantaneous normalized frequency deviation y(t) are linked by the relationship[1]:

$$y(t) = \frac{dx(t)}{dt} \qquad (1)$$

Each $\overline{y_k}$ sample is the time average of y(t) during an interval τ :

$$\overline{y_k} = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(\theta)\, d\theta \qquad (2)$$

Therefore, $\overline{y_k}$ samples may be calculated with values of time error function x(t), following the relationship :

$$\overline{y_k} = \frac{x(t_k+\tau) - x(t_k)}{\tau} \qquad (3)$$

Then, it is possible to obtain spectral density of frequency deviation samples by Fourier transform of $\overline{y_k}$ values. It may be noticed that N time error data allow to calculate N-1 samples $\overline{y_k}$. The first algorithm which has been used, performs this calculation, and then, computes FFT (Fast Fourier Transform) of these samples. To obtain spectral density $S_y(f)$ values, just the sum of the real and imaginary parts squared are calculated. Figure 1 shows the graph of the spectral density of a simulated[2] f^2 noise (white phase noise) computed with this algorithm.

This curve exhibits, for high frequencies, a typical f^2 slope, which is characteristic of white phase noise, but, for lower frequencies, it seems that there is a white frequency noise (f^0) in the signal. It must be noticed that time error values have been simulated by a white noise sequence. Therefore, spectral density of frequency deviation samples must only follow a f^2 slope (relationship 5). Thus, the white frequency noise has been generated by the calculation method and does not correspond to a noise present in the signal.

In order to verify this purpose, an other calculation algorithm has been used. The only difference with the first one is that for N values of time error function x(t), N samples $\overline{y_k}$ have been calculated (relationship 3).

The last sample is obtained by subtracting the first value x(t) of the last one and dividing the result by τ. It follows that the sequence x(t) is considered as periodic, with a period equal to $N\tau$. This hypothesis, which may appear without meaning from the physics point of view, is yet implicit each time that a FFT algorithm is used : the spectrum which is calculated is discrete, therefore the data sequence in the time domain is considered as periodic[3] (the period is the time duration of the all sequence). The spectral density shown in figure 2 has been calculated by this algorithm with the same data sequence as figure 1.

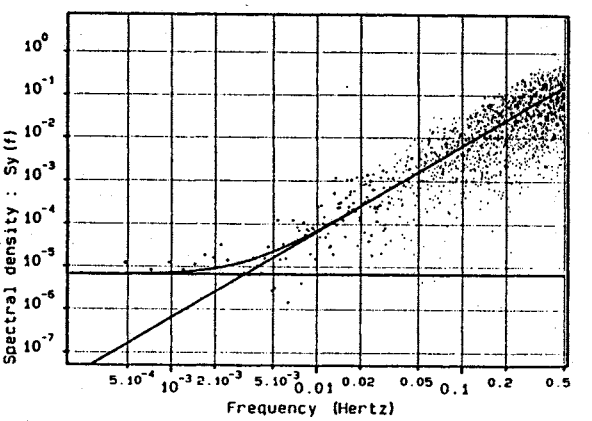


Figure 1 : Spectral density $S_y(f)$ of 4096 samples of f^2 noise. $\overline{y_k}$ samples have been calculated by the first algorithm described in chapter 1.1.

It may be conceded that changing only one of 4096 $\overline{y_k}$ samples, leads to a spectral density completely different. However, since these curves have been plotted with logarithmic scale, the differences are only sensitive for the lower frequency samples. Nevertheless, if we measure the different types of noise present in the signal with a variance, the results obtained with these algorithms will be very different (see figure 8, chapter 3).

It would appear, then, that the second algorithm, despite its improper hypothesis, is the most adapted to the calculation of $\overline{y_k}$ samples, and, therefore, of the spectral density $S_y(f)$. The causes of the differences between these algorithms will be studied theoretically in chapter 2.

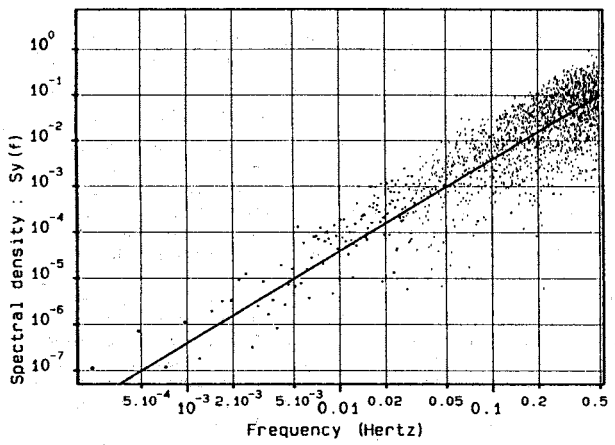


Figure 2 : Spectral density $S_y(f)$ of 4096 samples of f^2 noise. $\overline{y_k}$ samples have been calculated by the second algorithm described in chapter 1.1.

In order to compare with these calculation algorithms of the spectral density $S_y(f)$, a new method, working in the frequency domain, has been used.

1.2. Calculation of the spectral density of instantaneous normalized frequency deviation with the spectral density of instantaneous time error.

Since instantaneous normalized frequency deviation y(t) is the time derivative of instantaneous time error x(t), the spectral densities of these functions are linked by the relationship[11] :

$$S_y(f) = 4\pi^2 f^2 \, S_x(f) \qquad (4)$$

The calculation algorithm of this method consists of :

- performing FFT of x(t) data,
- multiplying each frequency sample, real or imaginary, by 2π and by the corresponding frequency,
- calculating the sum of the square of real and imaginary part of this spectrum to obtain the spectral density of instantaneous normalized frequency deviation $S_y(f)$.

The spectral density shown in figure 3 has been calculated by this method with the same x(t) data as figures 1 and 2.

This spectral density seems very similar to the one calculated with the second algorithm of the first method. There is no trace of the white frequency noise which appear in the spectral density calculated by the first algorithm of the first method. In order to compare these different calculation methods of the spectral density $S_y(f)$, their ratio have been studied, the spectral density obtained by the second method being considered as the reference.

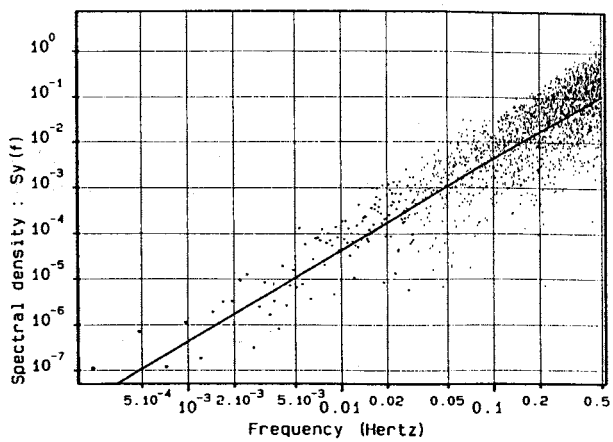


Figure 3 : Spectral density $S_y(f)$ of 4096 samples of f^2 noise. The calculation has been performed from the time error spectral density $S_x(f)$.

1.3. Comparison of the spectral densities obtained by the different calculation methods.

If we arbitrarily assume that the spectral density $S_y(f)$ calculated by the second method is the reference spectral density, it is possible to calculate transfer functions which allows one to link the spectral densities calculated by the first method to this reference spectral density. Let us call $S_{y0}(f)$ this reference spectral density, $S_{y1}(f)$ the one calculated by the first algorithm of the first method and $H_1(f)$ the corresponding transfer function, $S_{y2}(f)$ and $H_2(f)$ the spectral density and the transfer function obtained by the second algorithm of the first method (these transfer functions are the square of the modulus of the complex transfer functions $h_1(f)$ and $h_2(f)$ linking the different spectra, see chapter 3) :

$$S_{y1}(f) = H_1(f) \, . \, S_{y0}(f) \qquad (5)$$

$$S_{y2}(f) = H_2(f) \, . \, S_{y0}(f) \qquad (6)$$

The transfer functions may be written as :

$$H_1(f) = \frac{S_{y1}(f)}{S_{y0}(f)} \qquad (7)$$

$$H_2(f) = \frac{S_{y2}(f)}{S_{y0}(f)} \qquad (8)$$

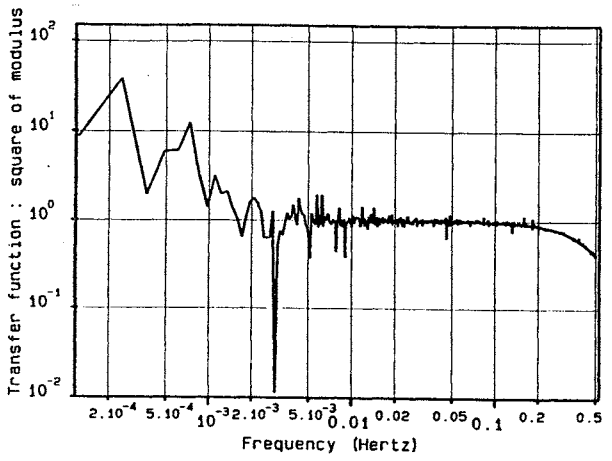
The calculation of these transfer functions will be usefull for :

- measuring the differences in comparison with the unity transfer function ;
- verifying if they are independant of the spectral density of the studied signal. Otherwise, the link between the spectral densities is composed of direct product and convolution product[13].

Figure 4 shows the graph of the transfer function $H_1(f)$ plotted with linear and logarithmic scales.

First of all, let us notice the random feature of this function which indicates that $H_1(f)$ is not independent of the studied spectral density. Indeed, if this function is plotted for an other signal, the general feature remains the same, but random fluctuations are different. However, great amplitudes for the first frequency samples (frequency $\leq 2.10^{-3}$ Hz in figure 4), perceptible on the logarithmic graph, remains systematically if the studied signal contains f^2 noise. This behaviour for low frequencies is responsible for the presence of white frequency noise in the spectral density $S_y(f)$ calculated by the first algorithm of the first method.

It may be also observed, in the linear graph, that random fluctuations seems to modulate a non-random function. This modulation, which is more perceptible for low frequency, let appear, for high frequency, a true transfer function due to calculations method. This function varies from 1, for the lower frequencies, to about 0.4 for the higher frequencies. However, the logaritmic graph shows that this transfer function may be considered as constant and equal to 1 in the case of variance measurement (variance curves are generally plotted on logarithmic graphs in order to identify easily the different power laws).



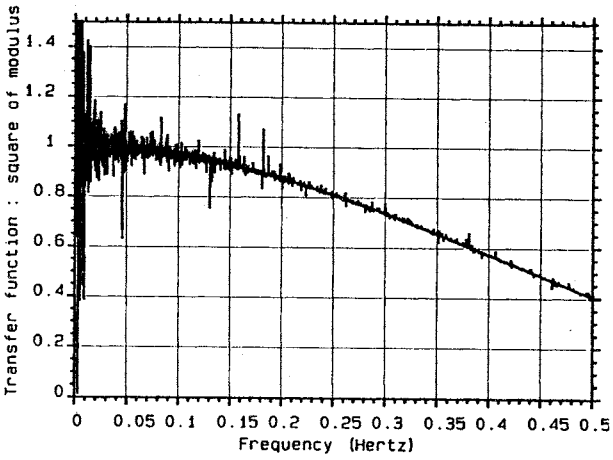


Figure 4 : Ratio of the spectral densities $S_y(f)$ calculated by using the first algorithm of the first method and by using the second method (logarithmic scale above and linear scale below).

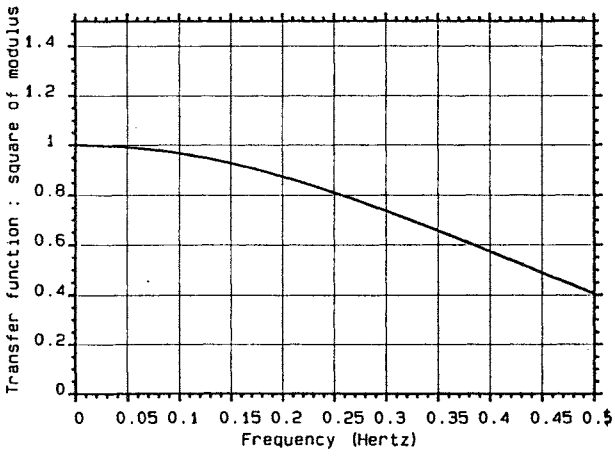


Figure 5 : Ratio of the spectral densities $S_y(f)$ calculated by using the second algorithm of the first method and by using the second method

This transfer function seems identical to the curve of figure 5 which exhibits the transfer function $H_2(f)$, ratio of the spectral density calculated by the second algorithm of the first method on the one calculated by the second method.

In this curve, which is quite smooth, there is no more random fluctuations. It seems that the fact of calculating the last $\overline{y_k}$ sample with the hypothesis of periodicity of the sequence, leads to a spectral density $S_y(f)$ which may be linked by a simple transfer function to the spectral density calculated by the second method. Moreover, the transfer function $H_2(f)$ remains always identical to the one of figure 5 whatever the time error x(t) sequence is. Furthermore, this function may be exactly fitted by the following expression :

$$H_2(f) = \frac{\sin^2(\pi\tau f)}{(\pi\tau f)^2} \qquad (9)$$

The presence of this type of function, which is frequent in signal processing, may be easily explained.

2. THEORETICAL INTERPRETATION.

In the transfer function calculated previously, it appears, on the one hand, a deterministic function which may be modelled by relation 9 and which is common to $H_1(f)$ and $H_2(f)$. On the other hand, it appears, for $H_1(f)$ only, a random modulation which depends on the studied signal.

2.1. Deterministic transfer function.

In order to distinguish the behaviours of the real and imaginary parts, let us yet study the Fourier transform of the y(t) function, and note $\mathcal{Y}(f)$ this spectrum of which the square of the modulus is the spectral density $S_y(f)$.

Since the second algorithm of the first method let appear only the deterministic transfer function, the function $y_2(t)$ is first studied. This function is the continuous extend of the $\overline{y_k}$ samples sequence calculated by this method. Relationship 3 leads us to the following expression :

$$y_2(t) = \frac{x(t+\tau) - x(t)}{\tau} \qquad (10)$$

$$y_2(t) = \int_{-\infty}^{+\infty} x(\theta) \left[\frac{\delta(t-\theta+\tau) - \delta(t-\theta)}{\tau}\right] d\theta \qquad (11)$$

$$y_2(t) = x(t) * \left[\frac{\delta(t+\tau) - \delta(t)}{\tau}\right] \qquad (12)$$

In fact, the sampling rate is τ and the number of samples is N. Moreover, for this algorithm, the x(t) values sequence is considered as periodic, with a $N\tau$ period. Let us note X(t) the true instantaneous time error x(t) function which is continuous in the measurement interval and equal to 0 outside. The x(t) measures sequence may be written as :

$$x(t) = \left[X(t)\ \text{Ш}(t/\tau)\right] * \left[\frac{1}{N\tau}\ \text{Ш}(t/N\tau)\right] \qquad (13)$$

To be rigorous, in the first quantity in bracket of relationship 13, a direct product with a $N\tau$ width gate-function should be introduced, in order to take into account the finite duration time of the measurement sequence. In the frequency domain, that implies a correlation product by a $\sin(N\pi\tau f)/(N\pi\tau f)$ function.

However, since the lobes of this function are very narrow, it may be easily neglected in order to simplify the expressions.

Then, $y_2(t)$ may be modelled by the relationship :

$$y_2(t) = \left[\left[X(t)\text{Ш}(t/\tau)\right] * \left[\frac{1}{N\tau}\text{Ш}(t/N\tau)\right]\right] * \left[\frac{\delta(t+\tau) - \delta(t)}{\tau}\right] \quad (14)$$

The Fourier transform of $y_2(t)$ may be written as[3] :

$$\mathcal{Y}_2(f) = TF\left[\left[\left[X(t)\text{Ш}(t/\tau)\right] * \left[\frac{1}{N\tau}\text{Ш}(t/N\tau)\right]\right] * \left[\frac{\delta(t+\tau)-\delta(t)}{\tau}\right]\right] \quad (15)$$

$$\mathcal{Y}_2(f) = \left[\left[TF\left[X(t)\right] * \left[\tau\text{Ш}(\tau f)\right]\right]\text{Ш}(N\tau f)\right]\left[\frac{-2j}{\tau} e^{-j\pi\tau f}\sin(\pi\tau f)\right] \quad (16)$$

It may be noticed that the $\mathcal{Y}_2(f)$ function, which is the Fourier transform of $y_2(t)$, is periodic with a $1/\tau$ period (convolution product with a $1/\tau$ period Dirac lattice), and sampled with a $1/(N\tau)$ sampling rate (direct product with a $1/(N\tau)$ period Dirac lattice). Moreover, these operations are applied to the spectrum of the X(t) function, and the transformation to the spectrum of $y_2(t)$ is due to the last factor in brackets.

The expression of the $\mathcal{Y}_0(f)$ function, of which $S_{y0}(f)$ is the square of the modulus, is obtained from the Fourier transform of the X(t) measures sequence, as it was shown in the previous chapter :

$$\mathcal{Y}_0(f) = 2\pi f\ TF\left[x(t)\right] \quad (17)$$

Substituting x(t) by its expression versus X(t) (relationship 13), we find :

$$\mathcal{Y}_0(f) = 2\pi f \left[\left[TF\left[X(t)\right] * \left[\tau\ \text{Ш}(\tau f)\right]\right] .\ \text{Ш}(N\tau f)\right] \quad (18)$$

Then, $\mathcal{Y}_0(f)$ function is also periodic and sampled, but the function which permits to obtain the spectrum of $y_0(t)$ from the spectrum of X(t) is different from the one which permits to obtain the spectrum of $y_2(t)$. Carrying out the ratio of $\mathcal{Y}_2(f)$ on $\mathcal{Y}_0(f)$, it is possible to obtain the transfer function $h_2(f)$ corresponds to the square of the modulus $H_2(f)$ in the previous chapter :

$$h_2(f) = \frac{\mathcal{Y}_2(f)}{\mathcal{Y}_0(f)} \quad (19)$$

$$h_2(f) = \frac{\frac{-2j}{\tau} e^{-j\pi\tau f}\sin(\pi\tau f)}{2\pi f} \quad (20)$$

$$h_2(f) = -j\, e^{-j\pi\tau f}\frac{\sin(\pi\tau f)}{\pi\tau f} \quad (21)$$

Performing the square of the modulus of this function, it comes :

$$|h_2(f)|^2 = \frac{\sin^2(\pi\tau f)}{(\pi\tau f)^2} = H_2(f) \quad (22)$$

which confirms the experimental result of relationship 9.

Moreover, we may verify that the real and imaginary parts of the $h_2(f)$ function are identical to the ones of the experimental function. The $h_2(f)$ function may be splitted up into :

$$h_2(f) = Re\left[h_2(f)\right] + j\ Im\left[h_2(f)\right] \quad (23)$$

$$Re\left[h_2(f)\right] = -\frac{\sin^2(\pi\tau f)}{\pi\tau f} \quad (24)$$

$$Im\left[h_2(f)\right] = -\frac{\cos(\pi\tau f)\sin(\pi\tau f)}{\pi\tau f} \quad (25)$$

Concerning experimental data, the complex samples of the $\mathcal{Y}(f)$ spectra must yet be used. Thus, as :

$$Re\left[h_2(f)\right] + j\ Im\left[h_2(f)\right] = \frac{Re\left[\mathcal{Y}_2(f)\right] + j\ Im\left[\mathcal{Y}_2(f)\right]}{Re\left[\mathcal{Y}_0(f)\right] + j\ Im\left[\mathcal{Y}_0(f)\right]} \quad (26)$$

the real and imaginary parts of $h_2(f)$ may be written as :

$$Re\left[h_2(f)\right] = \frac{Re\left[\mathcal{Y}_2(f)\right]Re\left[\mathcal{Y}_0(f)\right] + Im\left[\mathcal{Y}_2(f)\right]Im\left[\mathcal{Y}_0(f)\right]}{Re\left[\mathcal{Y}_0(f)\right]^2 + Im\left[\mathcal{Y}_0(f)\right]^2} \quad (27)$$

$$Im\left[h_2(f)\right] = \frac{Im\left[\mathcal{Y}_2(f)\right]Re\left[\mathcal{Y}_0(f)\right] - Re\left[\mathcal{Y}_2(f)\right]Im\left[\mathcal{Y}_0(f)\right]}{Re\left[\mathcal{Y}_0(f)\right]^2 + Im\left[\mathcal{Y}_0(f)\right]^2} \quad (28)$$

The curves of figure 6 show the real and imaginary parts of the experimental transfer function $h_2(f)$. They have been obtained from the same data as the ones used for the previous figures.

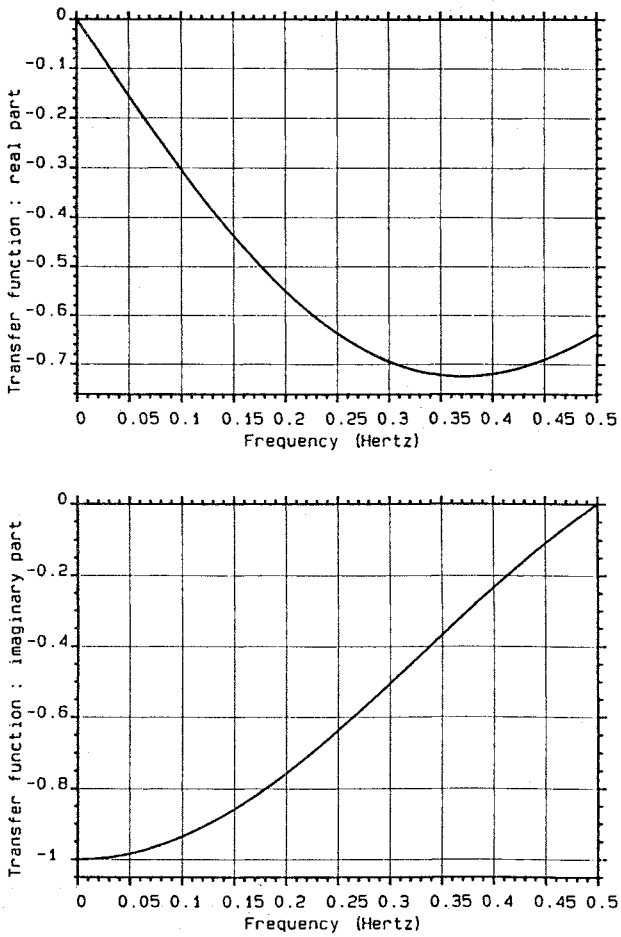


Figure 6 : Complex transfer function $h_2(f)$ calculated from experimental data (4096 samples of f^2 noise). Real part is above and imaginary part is below.

These curves are well fitted by the theoretical expressions (relationships 27 and 28). Moreover, they remain the same whatever the studied noise sequence is and whatever the number of samples is.

The deviation from the theoretical values calculated by the relationships 27 and 28 are less than 0.005 %. They are due to computer truncations.

Thus, the deterministic transfer function $h_2(f)$ is a true transfer function because it appears like a direct product in the frequency domain and does not depend on the studied signal. This is due to the fact that, for the first calculation method, the used samples are averaged values of the y(t) function. On the opposite, the second transfer function $h_1(f)$ is not a true transfer function because, in order to let it appear in a direct product in the frequency domain, it must depend on the studied signal.

2.2. Modulation of the transfer function.

To identify the modulations of the transfer function $h_1(f)$, the literal expression of this function must be written. As it was notice in the first chapter, the only difference between the second and the first algorithm of the first method consists to take or no account of the last sample $\overline{y_k}$. The use of a FFT algorithm implies that the measurement sequence is periodic. Consequently, taking no account of the last sample reverts to multiply all the samples for each period by a function equal to one excepted for the last sample $\overline{y_k}$ for which this function is zero.

This function m(t) is defined by :

$$m(t) = 1 - \left[\left[\Pi(t/\tau) * \delta(t+\tau)\right] * \left[\frac{1}{N\tau}\, \text{Ш}(t/N\tau)\right]\right] \quad (29)$$

The function $y_1(t)$, continuous extend of the calculated samples $\overline{y_k}$ by using the first algorithm of the first method is given by :

$$y_1(t) = y_2(t) \cdot m(t) \quad (30)$$

After calculating the Fourier transform of $y_1(t)$ it comes[3] :

$$\mathcal{Y}_1(f) = \mathcal{Y}_2(f) * TF\left[m(t)\right] \quad (31)$$

$$\mathcal{Y}_1(f) = \mathcal{Y}_2(f) - \mathcal{Y}_2(f) * \left[\tau\, \frac{\sin(\pi\tau f)}{\pi\tau f}\, e^{-2j\pi\tau f}\, \text{Ш}(N\tau f)\right] \quad (32)$$

It can be already remark that the $\mathcal{Y}_1(f)$ function have the same expression that the $\mathcal{Y}_2(f)$ function excepted for one term depending on all samples. The $h_1(f)$ function is given by :

$$h_1(f) = \frac{\mathcal{Y}_2(f) - \mathcal{Y}_2(f) * \left[\tau\, \frac{\sin(\pi\tau f)}{\pi\tau f}\, e^{-2j\pi\tau f}\, \text{Ш}(N\tau f)\right]}{\mathcal{Y}_0(f)} \quad (33)$$

$$h_1(f) = h_2(f) - \frac{\mathcal{Y}_2(f) * \left[\tau\, \frac{\sin(\pi\tau f)}{\pi\tau f}\, e^{-2j\pi\tau f}\, \text{Ш}(N\tau f)\right]}{\mathcal{Y}_0(f)} \quad (34)$$

As it was been a presentiment in section 1.3., the $h_1(f)$ is therefore the difference between the deterministic transfer function $h_2(f)$ of which the literal expression has been previously calculated and one term for which it is impossible to obtain a literal expression since it depends of the studied signal.

Thus, the suppression of one $\overline{y_k}$ sample conducts to introduce this term which corresponds to the summation of all samples of the $\mathcal{Y}_2(f)$ function divided by the $\mathcal{Y}_0(f)$ function. As a matter of fact, this summation is weighted by a function of the form of $\sin(\pi f\tau)/(\pi\tau f)$ centered on the frequency f argument of $h_1(f)$. However, in a first approximation, a simple summation can be calculated since all samples are located in the first lobe.

As the numerator is amost constant, therefore, this term varies as the invert of the $\mathcal{Y}_0(f)$ function. Consequently, the first frequency values of the $h_1(f)$ function for a f^2 noise seems to exhibit a f^{-2} slope (see the logarithmic graph in figure 4) and the spectral density $S_{y1}(f)$ seems to contain some white frequency noise (see figure 1). In this case of f^2 noise, it is worth noting that by using the second algorithm of the first method, the fact of multiplying only one $\overline{y_k}$ sample among a 8192 samples sequence by a factor of 0.99 is enough to produce a significative modulation term in the $h_2(f)$ transfer function (see figure 7).

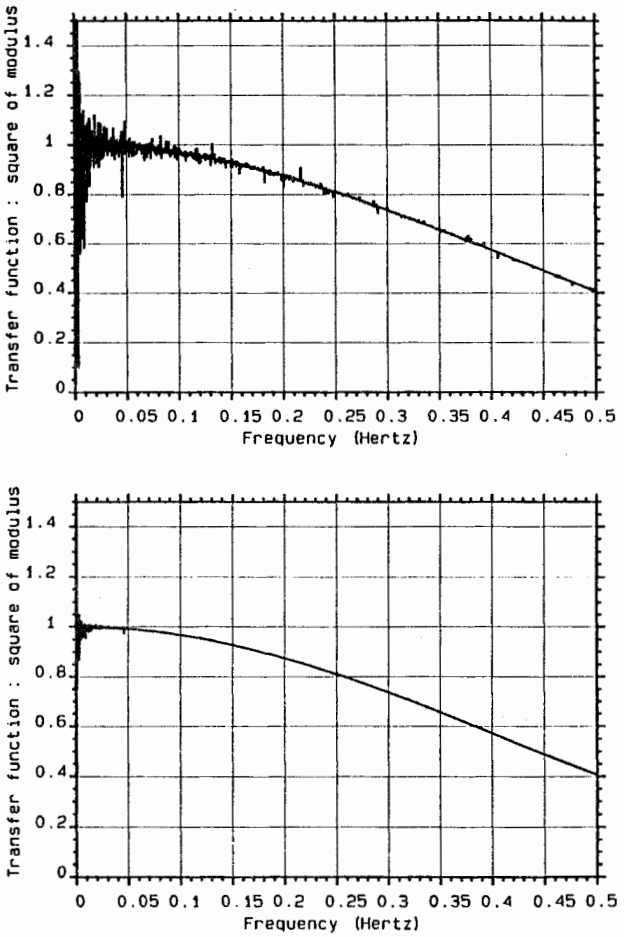


Figure 7 : Square of the modulus of $h_2(f)$ function calculated from 8192 f^2 noise samples. One sample of the corresponding $y_2(t)$ function has been multiply by a 0.9 factor, above, and by a 0.99 factor, below.

3. INFLUENCE OF CALCULATION METHODS ON VARIANCE RESULTS.

Measurements of different types of noises (f^2 noise, f^{-1} noise and f^{-2} noise) have been realized by using the modified Allan variance $\sigma_M^2(\tau)$[4][5].

The calculations have been performed in the frequency domain with the following relationship :

$$\sigma_M^2(\tau) = \int_0^\infty S_y(f)\ |H_M(f)|^2 df \qquad (35)$$

where $|H_M(f)|^2$ is the transfer function of the modified Allan variance. This function, for an integration time τ corresponding to N incremental sampling time, is defined by[5] :

$$|H_M(f)|^2 = 2\,\frac{\sin^6(\pi\tau f)}{N^2\,(\pi\tau f)^2\,\sin^2\left(\frac{\pi\tau f}{N}\right)} \qquad (36)$$

The major interest of the modified Allan variance by comparison with the simple Allan variance is the possibility to give, versus the integration time τ, a defined τ power law for each noise slope power law[4][5].

To make the calculation of the variance, it is then necessary to know the spectral density $S_y(f)$.

The results of the different calculation methods are shown on the following curves where the results of the second method are taken as reference.

On figure 8, for a studied f^2 noise, the upper variance curve has been calculated by using the first algorithm of the first method and the lower curve by using the second algorithm of the first method.

On these graphs, the effect of the transfer functions calculated in chapter 2 appears :

- for the lower values of τ, corresponding to the higher frequencies of the spectral density $S_y(f)$, a slight difference between the results of the application of the first and the second method can be noticed. This difference is due to the transfer function $h_2(f)$ which has been noted deterministic transfer function. The function $h_2(f)$ acts on the two algorithms of the first calculation method and attenuates the higher frequencies in comparison with the second method.

- for the higher values of τ and only in the upper graph, it can be noticed that the first algorithm of the first method results suddenly exhibit a change of slope to τ^{-1} corresponding to white noise[5] (see § 4.5). This change is due to the term depending on the studied signal in the $h_1(f)$ transfer function (see section 2.2). This terms varies as the inverse of the spectral density of the studied signal.

Unlike the deterministic transfer function, which have a small influence on the measurement precision, the other term of the $h_1(f)$ function applied to a f^2 noise gives results corresponding to a white frequency noise. Then, it is clear that the first algorithm of the first method must be avoided.

To verify if the second algorithm of the first method can be used for any types of noises, a comparison of the results given by the two calculation methods has been performed for a f^{-1} noise and for a f^{-2} noise (see figures 9 and 10).

As it was shown previously and although the higher scale used to plot the curve, only the first three values of variance are different. However, these differences are not high enough to yield a sensitive diminution of the measurement precision. In fact, this measurement is obtained in the case of f^{-1} noise by averaging the variance results over the different values of τ [5].

In the case of f^{-2} noise (figure 10) the difference between the two calculation methods are very small : only the first two values are not perfectly surimposed on the graph.

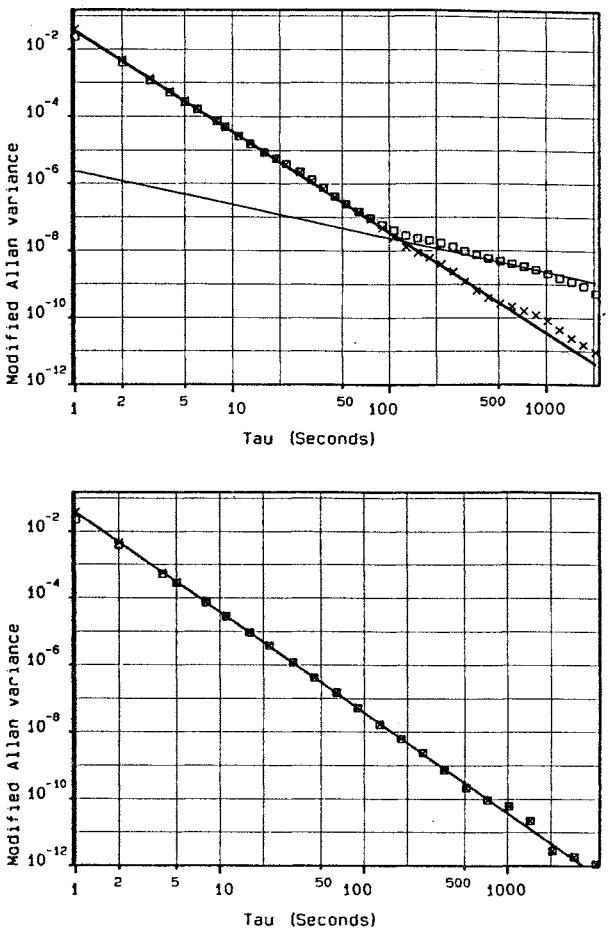


Figure 8 : Modified Allan variance applied to f^2 noise (□ first method, X second method). Above : first algorithm of the first method below : second algorithm of the first method

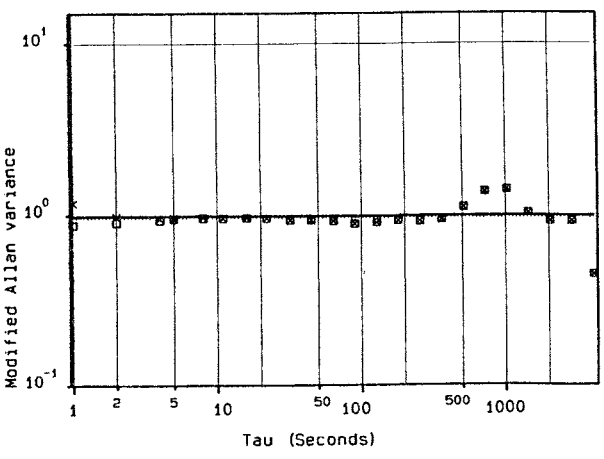


Figure 9 : Modified Allan variance applied to f^{-1} noise (□ first method, X second method).

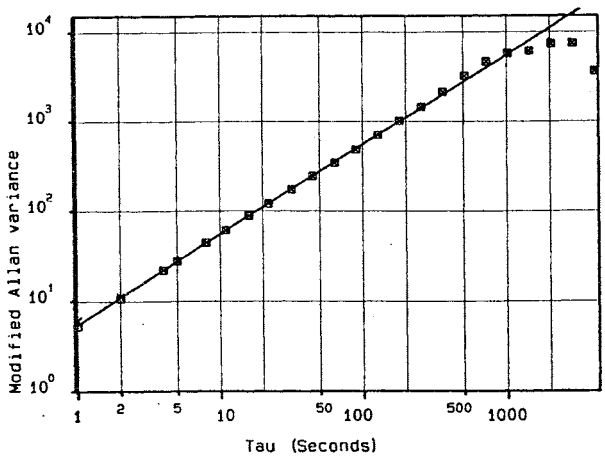


Figure 10 : Modified Allan variance applied to f^{-2} noise (□ first method, X second method).

4. CONCLUSION : PRACTICAL SOLUTIONS.

Thus, it is clear that only two methods can be used among thethree one proposed to calculate the spectral density $S_y(f)$ of a signal from instantaneous time error x(t) samples.

The first method, which is in fact a calculation method of the $\overline{y_k}$ samples since the Fourier transform is computed after, may be used each time that it is desirable to remain in the time domain.

As a matter of fact, if the used variance exhibits a transfer function corresponding in time domain to a defined calculation sequence, it is possible to obtain directly the variance from the $\overline{y_k}$ samples[11]. In this case, the only rule to follow is to calculate a number N of $\overline{y_k}$ samples for a number N of x(t) measures, the last $\overline{y_k}$ sample being obtained from the first x(t) measure and the last one (hypothesis of periodicity of the signal, see section 1.1.).

The second method, which is mathematically more rigorous, but which gives slight different results, presents the disadvantage of requiring the calculation of the Fourier transform of the x(t) values. Consequently, this second method is very adapted to the variance calculation in the frequency domain since in this method the spectral density $S_y(f)$ of the samples is directly calculated. This possibility is interesting when the used variance is only defined by its transfer function and have a unknown sequence in the time domain (high-pass variance, band-pass variance, ...).

5. REFERENCES.

[1] Rutman, J. : Characterization of phase and frequency instabilities in precision frequency sources : fifteen years of progress, Proceedings of the IEEE, vol. 66, N# 9, September 1978, pp. 1048-1075.

[2] Groslambert, J., Gagnepain, J. J., Vernotte, F., Walls, F. : A new "filtered Allan variance" and its application to the identification of phase and frequency noise sources, 43th annual Frequency Control Symposium, Denver, June 1989.

[3] Roddier, F. : Distributions et transformation de Fourier, McGraw-Hill, Paris, 1978.

[4] Allan, D. W., Barnes, J. A. : A modified "Allan variance" with increased oscillator characterization ability, Proceedings of the 35th Frequency Control Symposium, Ft. Monmouth, May 1981, pp. 470-475.

[5] Bernier, L. G. : Theoretical analysis of the modified Allan variance, Proceedings of the 41st Frequency Control Symposium, Philadelphia, May 1987, pp.116-121.

Session 8B Number 81

A SIMPLE TEMPERATURE CONTROL FOR SEMICONDUCTOR LASERS

G. D. Rovera*, A. De Marchi+, G. Santarelli*

*Laboratoire Primaire du Temps et des Frequences, Observatoire de Paris
61, avenue de l'Observatoire 75014 Paris FRANCE

+Dipartimento di Elettronica, Politecnico di Torino ITALY

ABSTRACT

The frequency of semiconductor lasers needs to be decoupled from external temperature variations in advanced metrological applications. A simple PID style temperature controller was built, the servo system was designed for a long term temperature stabilization better than 1 mK. The achievement of this level of stability was experimentally tested by measuring the frequency fluctuations of the laser itself. Experimental results show that a 5K temperature variation give a frequency shift of 100 MHz . The laser diode temperature coefficient is about 30-40 GHz/K. This means that a 1K of ambient variation is reduced to 1mK or less on the laser junction.

INTRODUCTION

Laser diodes are widely used in industrial and scientific applications. In many of these, a high stability of the output frequency is a plus; for use in metrology it's a must.

In order to operate an optically pumped Cesium beam frequency standard it is necessary to use lasers with a linewidth not too much larger than the absorption line of Cesium, that is of about 10 MHz. This may be achieved with semiconductor lasers by optical feedback narrowing [1].

The capture range of optical feedback is of the order of few hundreds of MHz [2]. This means that the frequency of the free running laser needs to remain at least within 100 MHz of the initial value during all the operation time, for line narrowing to be maintained, without the employ of more complicated techniques [3]. Such range must not be exceeded in a laboratory environment for at least a few months, notwithstanding all environmental changes and aging of the laser itself and other components. In particular the laser's frequency depends strongly on the junction temperature via variations of cavity length and refraction index of the gain material.

Since the temperature coefficient of a AsGaAl diode lasing at 852 nm is about 30-40 GHz/K, in order to obtain the required frequency stability it is necessary to control the junction temperature with an error than 2 mK. In addition, the servo system necessary to stabilize the temperature must be easy to use, inexpensive and robust.

In this paper the design and performance of such an electronic system is described. The resulting laser's frequency stability is tested by frequency measurement against the D2 line of Cesium.

SYSTEM MODEL

In order to design the temperature controller, consideration of the whole system is necessary, so that a suitable model can be devised for the thermal circuit. For this, reference is made to the particular configuration used.

The laser is a Hitachi HLP1400 mounted on a copper block which is also the interface to the Peltier element. The temperature sensor is a glass bead thermistor fixed with thermal glue in a hole drilled in the block as close as it was possible to the laser junction (about 3mm). The Peltier element is a Marlow MI1022T (1 cm2), which is squeezed between the copper block and a copper base that carries also the beam collimator. The whole fixture is contained in a metal box whose purpose is to protect the laser from dust, air currents and electrical perturbations.

The block diagram of the control system (Fig.1) shows loop elements and disturbance sources. This a classical regulator, which stabilizes the temperature of the sensing device. As our aim is stabilizing the temperature of the laser, a problem arises from the thermal resistance existing between the chip and sensor. The result is that environmental fluctuations affect the laser temperature by a factor proportional to the ratio Rj-box/Rj-block. This is illustrated by the equivalent circuit shown in Fig.2.

The copper block must be built in such a way that it would not introduce excessive thermal propagation delays. In fact the latter can negatively affect the loop performance.

The temperature measurement system is based on a thermistor, which forms one leg of a resistance bridge.

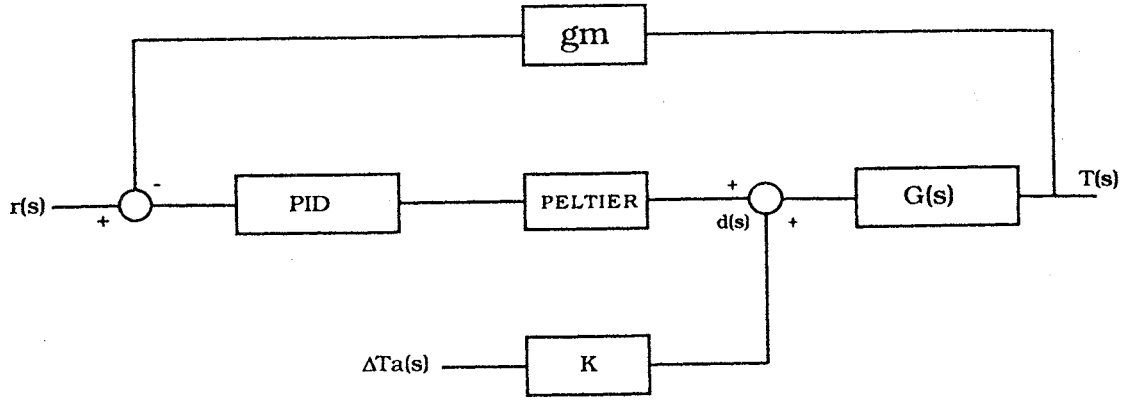


Fig 1: Schematic block diagram of the control system where: G(s) is the transfer function of the system to be controlled, gm is the thermistor measurement gain, PID is the controller itself, Peltier is the gain of heat pump and K is the gain of disturbance transfer from environment fluctuations to the system.

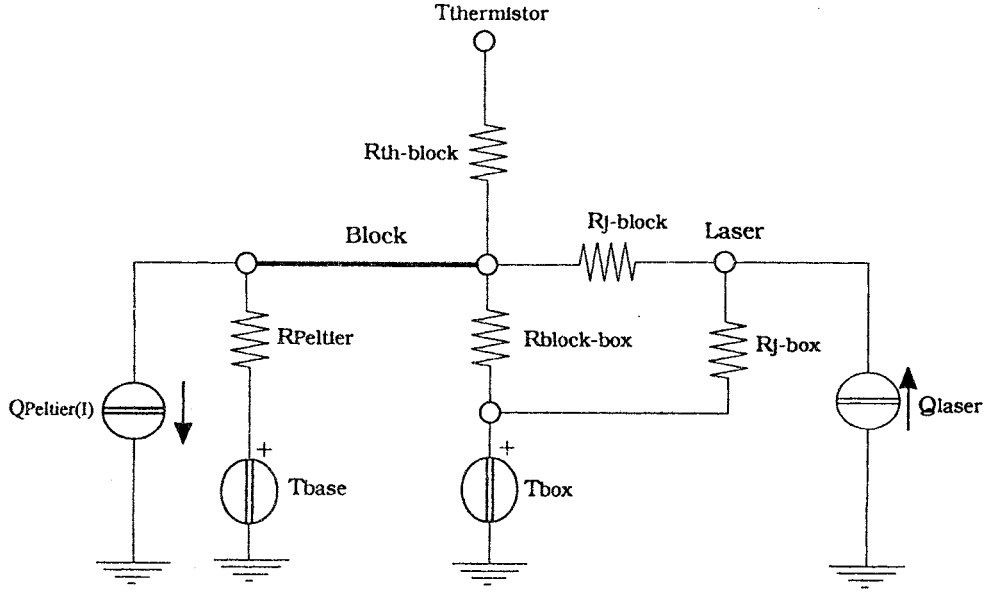


Fig 2: Equivalent electrical circuit of the thermal system. QPeltier is the heath pumped by the Peltier element. RPeltier is the thermal loss through the Peltier. Tbase and Tbox are the temperature of mounting plate and shielding box. Rblock-box is the the thermal resistance between block and box via air and radiation. Rj-block is the sum of all resistances between laser junction and copper block, but in first approximation the resistance from package to the block can be neglected with respect to the resistance between junction and package. Rj-box is the resistance between junction and surrounding block via air and radiation. Rth-block is the resistance between thermistor and block that in steady state condition may be neglected as the heat produced by the laser Qlaser.

In order to investigate the dynamic behaviour of the system, a current step was imposed to the Peltier thermoelement, and the response was measured at the thermistor temperature, the distance of the latter from its final value can be approximated by a decreasing exponential, whose time constant is given by the thermal capacity of the block and the thermal resistance between block and environment. This resistance turns out to be mainly given by the thermal losses trough Peltier element.

However, in the present work this approximation was considered too crude. In fact, the actual thermal circuit is more complicated because of the thermal response of the Peltier, the heat propagation in the copper block and the thermistor response. This complexity should be in principle modelized by a higher order system, but since at first glance it shows up mostly as a time delay in the start of the mentioned exponential, it was chosen to summarize its effects by simply introducing a small delay in the first order system. This choice is widely used in the practice of servo control synthesis, and is know to produce good results.

Another relevant feature that these measurements have shown is that the Peltier element characteristic depends on the operating point, being more efficient in the heating than in the cooling mode.

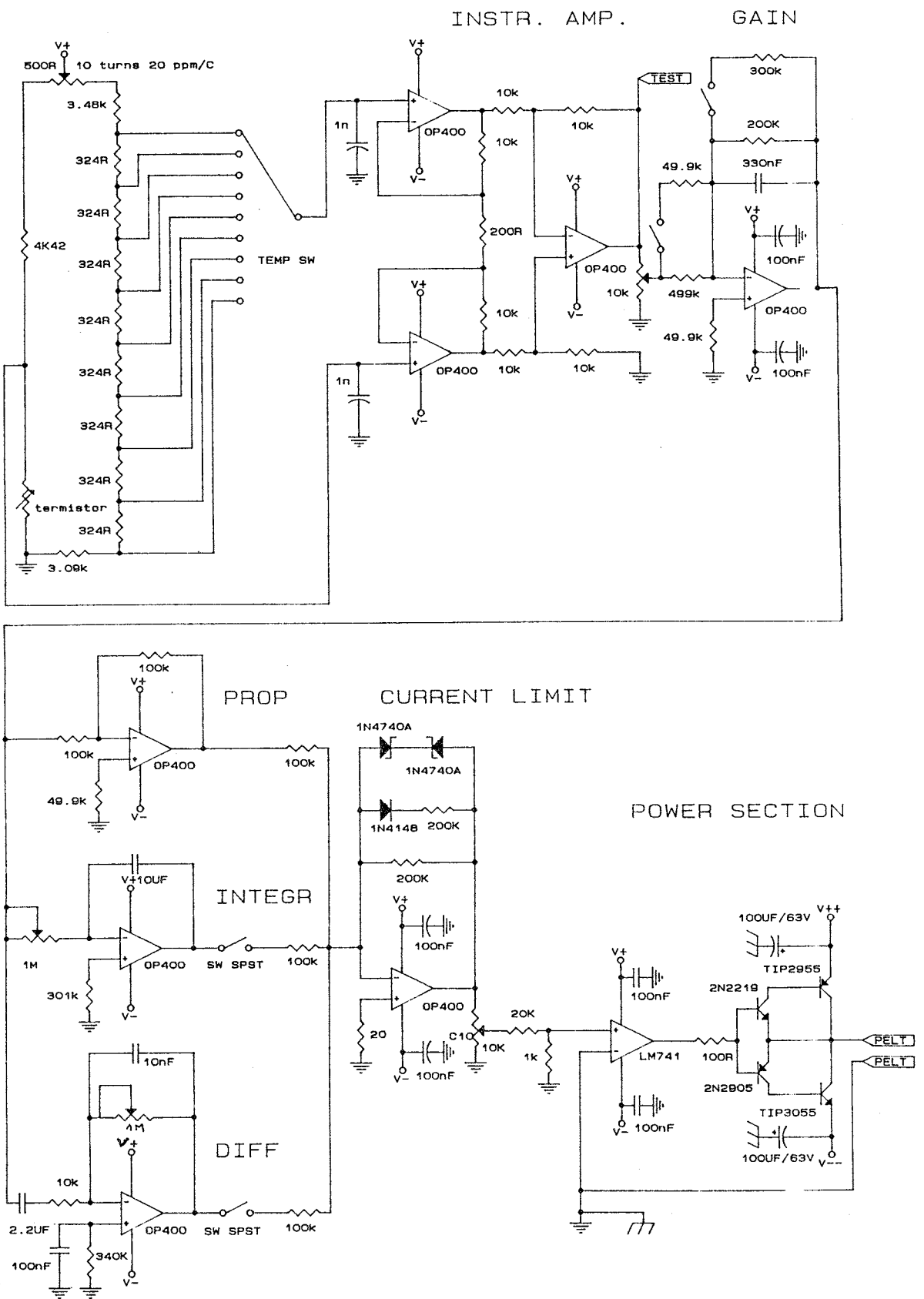


Fig 3: Schematic of controller.

CIRCUIT OVERVIEW

The schematic drawing of electrical circuit is given in Fig.3 . A 4.7 Ko glass bead thermistor highly stable in long term [4] is one of the legs of a resistor bridge where the coarse adjustment is obtained by switching resistors, and the fine regulation by a wire wound potentiometer. This structure reduces the weight of potentiometer characteristic in the overall bridge stability. The temperature range which can be covered is 10 to 50 centigrade, by steps of 5. The fine adjustment allows a resolution of 20 mK. The bridge is carefully designed to be insensitive to contact resistances, and is realized with 5ppm/K resistor because standard 50ppm/K metal film resistors would not guarantee, without being ovenized, the required temperature stability. The bridge imbalance signal is fed to an instrumentation amplifier, whose temperature drift, input resistance and CMRR are selected to obtain the necessary stability.

The controller structure adopted is that of a PID style servo system, with a transfer function $F(s)=K_P(1+\tau_D\, s+1/\tau_I\, s)$. This choice is justified by the fact that the transfer function of the thermal system is not exactly defined, and therefore is necessary to have an easy adjustment procedure available to set the control parameters. Proportional, integral and derivative actions are summed by an adder, which limits the output signal and can give a partial compensation of the Peltier element non-linear characteristic. A current driver power stage complete the circuits because the Peltier element's heat transfer is current driven.

In order to adjust the PID parameters, Ziegler-Nichols rules based on stability boundary are used [5]. These rules state that a reasonable rule-of-thunb choice for the loop constants is $K_P = 0.6\ K_r$, $\tau_I = 0.5\ T_C$, $\tau_D = 0.25\ T_C$ where K_r is the gain that gives the minimal stable oscillation when only proportional control is acting, and T_C is the period of these oscillation. A practical method for determining K_r consist in excluding derivative and integrative functions, and increasing the loop gain until small stable oscillations appears. Then τ_D and τ_I are set by adjusting the gain in the corresponding blocks.

STABILITY MEASUREMENT

The stability of the laser junction temperature was inferred from the laser output frequency variation in response to imposed external temperature disturbance. This gives an estimate of the real system performance obtained by controlling the sensor's temperature.

The light beam of the temperature stabilized laser is shone onto a Cesium atomic beam, and locked to the D2 4-5 transition signal. The error signal is then used as a measurement of the frequency difference between the Cesium line and the free running laser.

The measurement is performed by step-changing the set point of the room air conditioning system by 5 K, and recording error signal, room air, laser metal box air and laser base temperatures, and room air relative humidity. Recorded data for three of these variables are shown in Fig. 4.

These results show that the free running laser frequency variation for a 5 degree step in the set point of room temperature is smaller than 100 MHz. The inferred junction temperature variation is not more than 3 mK.

For the long term stability all we can say is that after few days of continuous operation there was no measurable frequency drift for a given room temperature.

CONCLUSION

The results reported here show that the task of stabilizing the laser temperature in such a way that the free running laser frequency wouldn't drift away more than 100 MHz in laboratory environment was meet in a satisfactory way. This was obtained with a simple and easy to use circuit.

In a case in which higher stability were needed, the best way of improving this results would be to stabilize the temperature of base plate and protective box. In this case the stability of the bridge may turn out be insufficient, and that too may therefore need to be temperature stabilized. However, the aging rate of the laser itself may ultimately vanify the efforts of obtaining ever better stabilities for the frequency of the running laser.

ACKNOWLEDGEMENT

This work is supported by BNM. The authors gratefully acknowledge A. Clairon for useful discussion on the subject and M. Lours for the technical work.

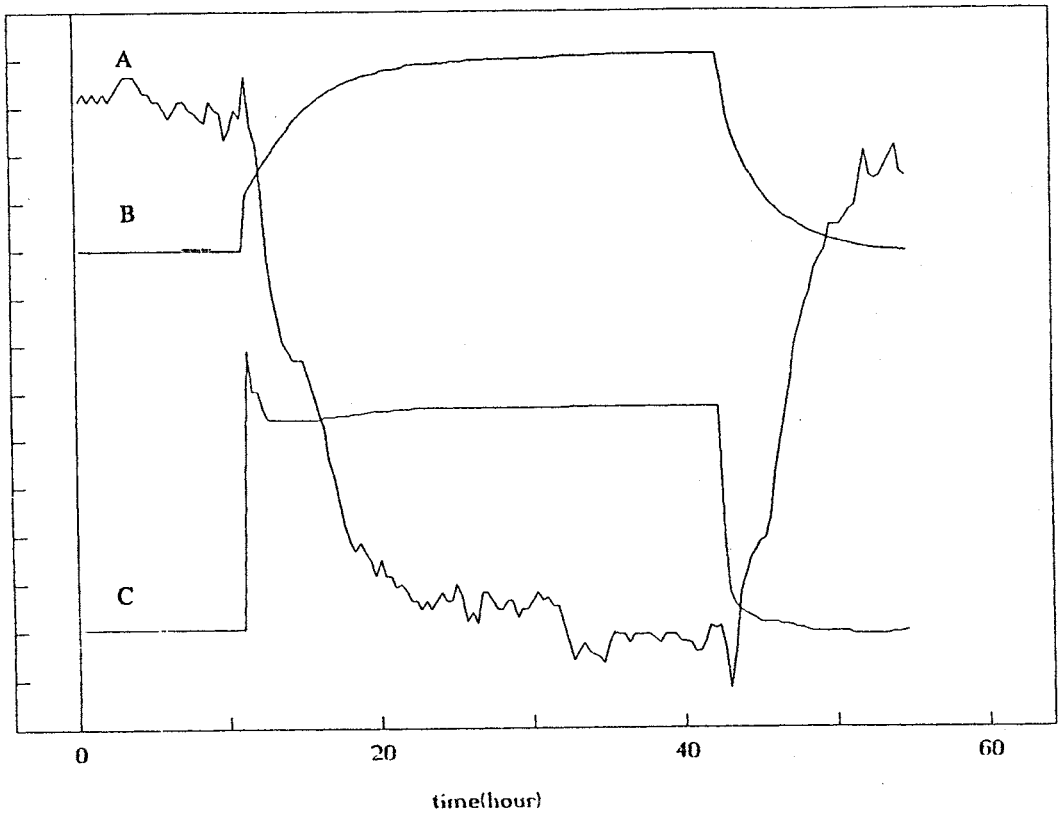


Fig 4: Experimental results: A control signal (10 MHz/div.)
B room temperature (1K/div.)
C air temperature inside the box (1K/div.)

REFERENCES

[1] B. Dahmani, L. Hollberg, and R. Drullinger, "Frequency stabilization of semiconductor lasers by resonant optical feedback" Opt. Lett. , vol. 12, no. 11, pp. 876-878, 1987.

[2] Ph. Laurent, A. Clairon and Ch. Bréant, "Frequency noise analysis of optically self-locked diode laser" IEEE J. Quantum Electronics, vol. 25, no. 6, June 1989

[3] Ph. Laurent, D. Bicout, Ch. Bréant, A. Clairon, "Frequency tuning methods for high spectral purity diode lasers self-locked on confocal Fabry-Perot cavities" Proc. of Laser Spectroscopy IX, NICOLS 89

[4] S. D. Wood, B. W. Mangum, J.J. Fillben, and S. B. Tillet, "An investigation of the stability of thermistors" J. of Research of the NBS, Vol. 83, May-June 1978.

[5] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, "Feedback control of dynamic systems" Addison Welsey ISBN 0-201-11540-9

Session 8B Number 82

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.

I. 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.

II. 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 2Ωt leads to offset errors [1,2]. Real phase modulators have small nonlinearities which generate small components of modulation at multiples of the modulation frequency. Assume that the phase modulated signal applied to the resonance 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 2×10^{-6}B or at least 114 dB below the fundamental modulation [1,2].

III. 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].

Near line center the steady state phase shift across a high Q resonator is

$$\Delta\phi = 2Q\,(\nu - \nu_0)/\nu_0, \qquad (5)$$

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

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

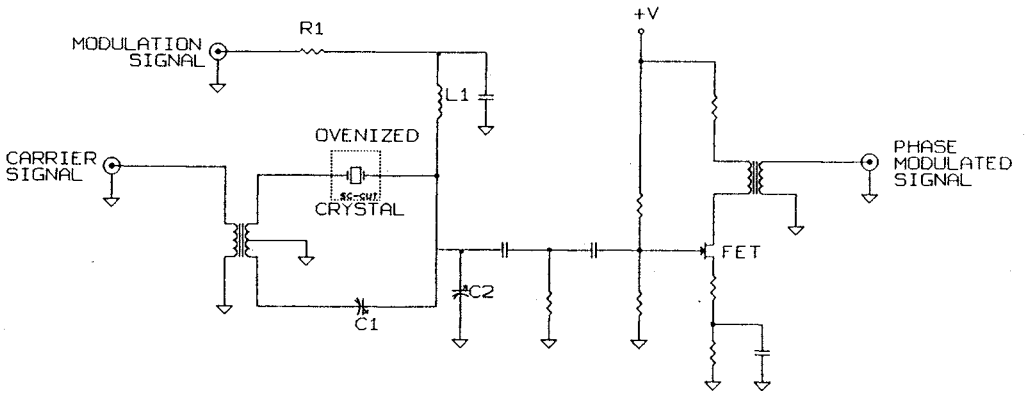


Figure 1. Block diagram of the new phase modulator.

resonator generated by applying an electric field E across the resonator results in a phase shift

$$\Delta\phi = 2Q(P_L \Delta E + P_Q \Delta E^2). \quad (6)$$

IV. 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 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

$$\Delta\phi = [2Q(P_L\Delta E + P_Q\Delta E^2)][1 + ((Q\Omega/(\pi\nu_0))^2]^{-1}, \quad (7)$$

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 ±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.

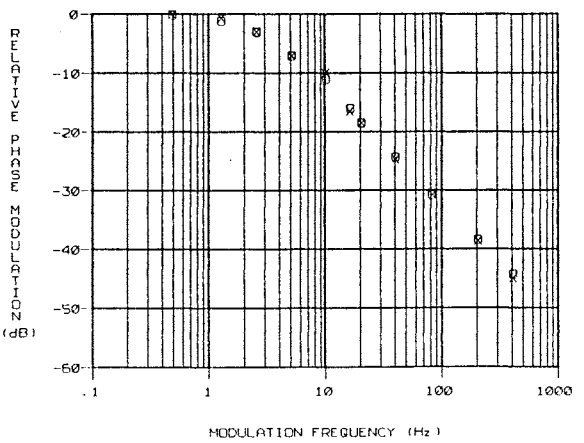


Figure 2. Relative phase modulation of the output signal as a function of the frequency of modulation. The open circles are the predictions using Eq. (7) and the half-bandwidth $\nu_0/(2Q)$ measured from transmission experiments. The solid points are the phase modulation levels obtained from phase bridge measurements [14,15].

The added phase noise is somewhat lower (typically $\mathcal{L}(100)$ = - 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 $2 \times 10^{-13}\ \tau^{-1/2}$ [3].

V. 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 50 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° of phase and using the unmodulated signal from the power splitter to drive the reference side of the mixer. 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.

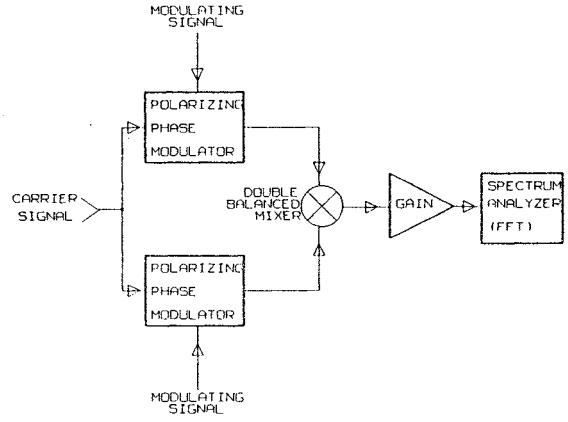


Figure 3. Block diagram of a phase bridge measurements to determine the difference in phase modulation between two phase modulators [14,15].

VI. 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 relating 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.

THE FOLLOWING PAPER COULD NOT BE PUBLISHED IN 1990 AND APPEARS IN THE 1991 PROCEEDINGS

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session Number

EQUIVALENT NETWORKS FOR PIEZOELECTRIC RESONATORS AND TRANSDUCERS

Arthur Ballato

US Army Electronics Technology & Devices Laboratory
Fort Monmouth, New Jersey 07703-5000 USA

ABSTRACT

Mechanical analogs were considered necessary in the last century for visualizing electromagnetic phenomena. Mechanisms were familiar objects, while electricity and light represented terrae incognitae. The situation was reversed in 1914 when Butterworth first used an equivalent electrical circuit to represent a mechanically vibrating system. This was followed by Van Dyke's independent discovery in 1925 that the same circuit characterized the impedance behavior of a piezoelectric resonator. Mason subsequently introduced, in the 1930s, acoustic transmission lines, mechanical ports, and piezoelectric transformers, thereby extending the circuit to encompass electro-mechanical conversion devices of wide generality. Today, the Mason equivalent circuit is universally used for bulk and surface acoustic wave device characterizations. It has also given rise to a variety of alternative formulations such as analog networks, KLM circuits, and systems models. The pervasive nature of electronics technology in the modern world has now resulted in a situation where an electrical network is often more familiar than the actual mechanical structure that it represents. The outlook has turned completely around!

A powerful motivation for using equivalent circuits to characterize piezoelectric devices is that the using equipment is often itself completely electrical in nature. The whole ensemble thus becomes subject to analysis or synthesis from a single perspective, i.e., network theory. In this paper are described a brief history of circuit equivalents through the Mason circuit, followed by a presentation of newer descriptions of piezoelectrically vibrating structures, specifically, analog networks and KLM circuits.

INTRODUCTION

The use of equivalent electrical networks to represent mechanical systems is pervasive today, particularly with respect to the depiction of piezoelectric resonators and transducers. This is a complete reversal of the situation existing less than one hundred years ago. One has only to read the works of Maxwell [01] and Kelvin [02] to see how ardently mechanical explanations were sought for electrical phenomena. Other analogies are treated in Refs. [03][04].

The situation was reversed in the case of electrically driven vibrators when Butterworth [05], and, independently, Van Dyke [06],[07] established the equivalence of the mechanically vibrating system to a certain purely electrical network, when both were viewed at the electrical input terminals.

The Butterworth-Van Dyke (BVD) [08]-[11] circuit consists of a series string of capacitor, resistor, and inductor, all shunted by a second capacitor. It represents the input immittance of a piezo vibrator in the vicinity of a single resonance. When the vibrator is not freestanding, but is instead loaded at its mechanical ports for operation as a transducer, the representation is that of the three-port shown in Fig. 1. It is due to W. P. Mason [12]. The elements are constant in value.

Mason later gave [13] an exact three-port circuit, shown in Fig. 2, for the case of a one-dimensional resonator, driven piezoelectrically in a single mode, with the applied field normal to the direction of the wave propagation. This network contains three lumped elements that are transcendental functions of frequency. He stated that these were necessary to take into account the acoustic wave propagation in the crystal. Mason's contributions ahve been pervasive in the area of equivalent circuits of bulk and surface wave resonators and transducers [14]-[22].

It was soon realized that a distinction had to be made according as the direction of the driving electric field was along or normal to the direction of acoustic wave propagation. In the latter case the electrical input circuit [23] consists only of a shunt capacitor, usually denoted C_0; in the former case, the input circuit consists additionally of a series capacitor of value $-C_0$. The case where field and propagation are perpendicular is referred to as "crossed-field," or "unstiffened;" when they are parallel it is known as "in-line," or "stiffened." In the case of a simple thickness mode excited in a crystal plate, the two situations are called "lateral-excitation," or LE (field perpendicular to the thickness direction along which the waves propagate), and "thickness-excitation," or TE. In the TE case, the self-generated field of the wave produces a reaction that leads to the negative C_0; this field also produces an effective elastic stiffness that is increased over the isagric value, hence the name "stiffened." This increase is often, but not always, absent when LE is used; the appellation "unstiffened" is therefore a misnomer and is to be used with caution.

Redwood and Lamb [24]-[28] recognized that the three lumped transcendental elements employed by Mason could be replaced by the more graphic transmission line (TL) representation shown in Fig. 3. This figure is drawn for the "in-line" case. The TL schematic chosen by Redwood and Lamb is that of a coaxial cable. An alternative schematic is that of the two-wire TL, also known as a Lecher line; Fig. 3 is redrawn in this manner in Fig. 4. The Lecher line was used extensively by H. Hertz in his electromagnetism experiments. When the transmission-line schematic is thus redrawn it is easy to see that the piezoelectric excitation occurs in series with the forces appearing at the mechanical ports. One may, in fact, split the piezo transformer into two parallel pieces and place them at the TL ends as seen in Fig. 5. The mechanical forces have here been set to zero as indicated by the short circuits placed at the mechanical ports.

Figure 5 places in evidence the concept of excitation taking place by the imposition of piezo tractions <u>at</u> the surfaces of the crystal plate, instead of volumetrically distributed throughout the crystal due to the interaction of the electric field and the dipoles. The piezoelectric portion of the constitutive relation between the elastic stress field and the electric field intensity is $T_{ij} = - e_{kij}E_k$. The force-tractions acting are the derivatives: $F_j = T_{ij,i} = - e_{kij}E_{k,i}$. The electric field suffers a delta-function discontinuity at the TE electrodes; this can be interpreted as a piezo-traction impressed at the surfaces.

In Fig. 6 are shown the physical interpretations of the various circuit elements that comprise the electrical equivalent circuit of a crystal resonator with massive electrodes. Figure 7 gives the complete network for a TE plate supporting a single piezoelectrically excited simple thickness mode. Also shown is the BVD lumped equivalent valid in a small frequency region about a single resonance. When more than one mode is driven, the piezo transformers are extended to the other TLs in the same fashion.

The resonator equivalent circuit, with no forces applied at the mechanical ports, for all three modes driven, may be bisected by the methods of network theory to give the representation of Fig. 8. Each TL is of length equal to one-half of the plate thickness; the piezo transformer turns ratios are functions of the coupling coefficients, k_m, for mode m.

By a partial-fractions expansion about the poles of the TL impedance, each mode may be represented exactly with lumped elements, as shown in Fig. 9. The circuit elements differ for each mode, of course.

Use of transmission lines to represent acoustic waves in distributed structures leads to a "building-block" approach to equivalent circuit realizations [29]. This is illustrated in Fig. 10 showing the network for a thin-film piezo layer deposited on a substrate.

In the brief discussion above, the TL circuits have been introduced without a consideration of the specific analogies between the electrical and mechanical variables. A more detailed treatment leads to the introduction of "normal coordinates" [30] to represent the TL variables, since, e.g., the three motions in an infinite plate supporting simple thickness motion represent the eigenmodes of the system, and each is accorded its own TL. At the boundaries, one must transform between the normal coordinate basis particular to the crystal type and its orientation, to laboratory coordinates. This is accomplished in network terms by the transformer arrangement of Fig. 11, where the turns ratios are the eigenvector components of each mode [29],[31].

By interconnecting transformers of the kind shown in Fig. 11, one is able to represent the juxtaposition of plates with various boundary conditions. Two media joined by welded boundaries (stress and displacement continuous across the interface) are represented in Fig. 12. Often the connections are simplified considerably, because of either the relative crystal orientations, or because of the materials, or both.

One of the virtues of the analog form is the ease with which the building-block approach can be used to accommodate more difficult situations. One such is Mindlin's problem, where mechanical surface tractions applied to a rotated-Y-cut quartz plate produce not only mechanical-acoustic motion, but also electromagnetic radiation [32]-[34]. Figure 13 models this situation.

A systems approach to equivalent networks has been taken by Hayward [35],[36]. In this method, Laplace transforms and linear systems theory are used to produce a circuit representation that is cast in signal-flow graph form.

Additional equivalent networks have been developed to model magnetostatic waves [37] and bulk acoustic wave resonators of piezoelectric semiconductors [38]. The procedure can be applied to use of finite element simulations [39].

The Krimholtz, Leedom, & Matthaei (KLM) equivalent circuits [40]-[44] are quite different in nature from all of the foregoing, and have the following advantageous features:

- o They are exact, within the one-dimensional approximation, and for a single mode type (for example, a ceramic resonator stack).
- o Arbitrary number of transducers in stack is accommodated, acoustically identical.
- o Simple network representation is produced, having:
  - series reactance or shunt susceptance (dispersive);
  - piezo transformer or immittance inverter (dispersive);
  - acoustic transmission lines (nondispersive).
- o Mechanical ports are not in series with piezo voltage.
- o Arbitrary spatial variation of piezo excitation is allowed:
  - "in-line" (stiffened modes)
  - "crossed-field" (unstiffened modes)
- o Simplifying algorithms exist for certain excitations.

The KLM networks have the following disadvantageous features:

- o Single modal type allowed, with density and stiffness constant, or averaged.
- o Members of the stack are all connected either in series, or all in parallel.
- o Circuit components are dispersive; this is unphysical.
- o Networks are equivalent only at the three ports:
  - no building-block approach is possible.
- o Complexity is traded from network schematic to the component values:
  - Fourier transforms F(k), spatial, of excitation
  - Hilbert transforms H(|F(k)|2)
- o Frequency domain operation only; e^{iwt}
  - no time domain, transient operation.

These features are summarized in the following table.

KLM CIRCUITS

Excitation	Electrical Input	TL Feed	Spatial Variation
In-Line "stiffened"	Series C and X_1		
	o transformer, Φ_p	parallel	even function
	o Z-inverter, K	series	odd function
Crossed-Field "unstiffened"	Shunt C, B_1		
	o Y-inverter, J	parallel	even function
	o transformer, Φ_s	series	odd function

F(k) = spatial Fourier transform of the excitation function; K is proportional to F(k); J and Φ_s are proportional to k times F(k); Φ_p is proportional to 1/F(k); X_1 and B_1 are proportional to H(|F(k)|2).

Figure 14 shows the KLM circuit for a thickness-stretch plate of thickness ℓ with electrodes on the major surfaces.

CONCLUSION

A brief review has been given of the genesis of equivalent networks for piezoelectric resonators and transducers. This led to a consideration of various types of representations, such as lumped element, multimode, analog, systems, and KLM types.

REFERENCES

[01] J.C. Maxwell, A Treatise on Electricity & Magnetism, 3rd Ed., 1891. See particularly pp. 228 and 431.

[02] W. Thomson (Lord Kelvin), Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light. London: Cambridge University Press, 1904. Lectures given at The Johns Hopkins University, Baltimore, MD, October 1884. See particularly Lect. I, Wed., Oct. 1 (pp.12-13), Lect. IX, Wed., Oct. 8 (pp.104-105), Lect. X, Thurs., Oct. 9 (pp. 118-119), and Lect. XI, Fri., Oct. 10 (pp.126-129).

[03] R.D. Mindlin and M.G. Salvadori, "Analogies," in Handbook of Experimental Stress Analysis, M. Hetényi, Ed. (J. Wiley & Sons, Inc., New York, 1950), Chap. 16, pp. 700-827.

[04] E. Hafner, "Resonator and Device Measurements," in Precision Frequency Control (E.A. Gerber and A. Ballato, eds.) Vol 2, Chap. 7, pp. 1-44. Orlando: Academic Press, 1985.

[05] S. Butterworth, "On Electrically-maintained Vibrations," Proc. Phys. Soc. (London), Vol. 27, 1915, pp. 410-424.

[06] K.S. Van Dyke, "The Electric Network Equivalent of a Piezo-Electric Resonator," (Abstract), Phys. Rev., Vol. 25, No. 6, June 1925, p. 895.

[07] K.S. Van Dyke, "The Piezo-Electric Resonator and its Equivalent Network," Proc. IRE, Vol. 16, 1928, pp.742-764.

[08] D.W. Dye, "The Piezo-Electric Quartz Resonator and its Equivalent Electrical Circuit," Proc. Phys. Soc. (London), Vol. 38, 1926, pp. 399-458.

[09] Y. Watanabe, "Der piezoelektrische Resonator in Hochfrequenzschwingungskreisen," Elektrische Nachrichten-Technik, Vol. 5, No. 2, Feb 1928, pp. 45-64.

[10] W.G. Cady, Piezoelectricity, (McGraw-Hill, NY, 1946), also (Dover, NY, 1964).

[11] A. Ballato, "Resonance in Piezoelectric Vibrators," Proc. IEEE, Vol. 58, No. 1, January 1970, pp. 149-151.

[12] W.P. Mason, "An Electromechanical Representation of a Piezoelectric Crystal Used as a Transducer," Proc. IRE, Vol. 23, No. 10, Oct. 1935, pp. 1252-1263.

[13] W.P. Mason, "A Dynamic Measurement of the Elastic, Electric and Piezoelectric Constants of Rochelle Salt," Phys. Rev., Vol. 55, April 1939, pp. 775-789.

[14] W.P. Mason, "A Study of the Regular Combination of Acoustic Elements, with Applications to Recurrent Acoustic Filters, Tapered Acoustic Filters, and Horns," Bell System Tech. J., Vol. 6, No. 2, April 1927, pp. 258-294.

[15] W.P. Mason, "Electrical Wave Filters Employing Quartz Crystals as Elements," Bell System Tech. J., Vol. 13, 1934, pp.405-452.

[16] W.P. Mason, "Electrical and Mechanical Analogies," Bell System Tech. J., vol. 20, no. 4, October 1941, pp. 405-414.

[17] W.P. Mason, Electromechanical Transducers and Wave Filters, (van Nostrand, NY, 1948), 2nd ed.

[18] A. Ballato, "Networks for Crossed-Field and In-Line Excitation of Bulk and Surface Acoustic Waves,", Proc. M.R.I. International Symposium XXIII, Optical and Acoustical Micro-Electronics, Polytechnic Institute of New York, Brooklyn, NY 11201, 16-18 April 1974, pp. 599-615.

[19] A. Ballato, "Piezoelectric Resonators," In Design of Crystal and Other Harmonic Oscillators, by Benjamin Parzen. Chapter 3, pp. 66-122 and pp. 432-436. John Wiley & Sons, NY 1983.

[20] T.R. Meeker, "Theory and Properties of Piezoelectric Resonators and Waves," in Precision Frequency Control. (E.A. Gerber and A. Ballato, editors) Volume 1, pp.48-118, Academic, Orlando, 1985.

[21] T. Thorvaldsson and F.M. Nyffeler, "Rigorous Derivation of the Mason Equivalent Circuit Parameters from Coupled Mode Theory," IEEE Ultrasonics Symposium Proc., November 1986, pp. 91-96.

[22] A. Ballato, "An Equivalent Circuit Appreciation of Warren P. Mason," J. Acoust. Soc. Am. Suppl. 1, vol. 85, Spring 1989, p. S19.

[23] H. Schüssler, "Darstellung elektromechanischer keramischer Wandler als Dickenscherschwinger mit piezoelektrischer und piezomagnetischer Anregung," Archiv der Elektrischen Übertragung, Vol. 22, No. 8, Aug 1968, pp. 399-406.

[24] M. Redwood and J. Lamb, "On the Measurement of Attenuation in Ultrasonic Delay Lines," Proc. IEEE (London), Vol. 103, Part B, November 1956, pp. 773-780.

[25] M. Redwood, "Transient Performance of a Piezoelectric Transducer," J. Acoust. Soc. Amer., Vol. 33, No.4, April 1961, pp. 527-536.

[26] M. Redwood, "Experiments with the Electrical Analog of a Piezoelectric Transducer," J. Acoust. Soc. Amer., Vol. 36, No. 10, Oct 1964, pp. 1872-1880.

[27] R.F. Mitchell and M. Redwood, "The Generation of Sound in Non-Uniform Piezoelectric Materials," Ultrasonics, Vol. 7, No. 2, April 1969, pp. 123-124.

[28] R.F. Mitchell and M. Redwood, "Generation and Detection of Sound by Distributed Piezoelectric Sources," J. Acoust. Soc. Amer., Vol. 47. No. 3, (Part 1), 1970, pp. 701-710.

[29] A. Ballato, "Doubly Rotated Thickness Mode Plate Vibrators," In Physical Acoustics: Principles and Methods, (W.P. Mason and R.N. Thurston, eds.) Vo.13, Ch.5, pp. 115-181, 1977. Academic Press, New York.

[30] T. Yamada and N. Niizeki, "Admittance of Piezoelectric Plates Vibrating Under the Perpendicular Field Excitation," Proc. IEEE Vol. 58, No. 6, June 1970, pp. 941-942.

[31] H.J. Carlin and A. B. Giordano, Network Theory: an introduction to reciprocal and nonreciprocal circuits, (Prentice-Hall, Englewood cliffs, NJ, 1964).

[32] J.J. Kyame, "Wave Propagation in Piezoelectric Crystals," J. Acoust. Soc. Amer., Vol. 21, No. 3, May 1949, pp.159-167.

[33] H. Pailloux,"Piézoélectricité Calcul des Vitesses de Propagation," J. Phys. Rad., Vol. 19, May 1958, pp. 523-525.

[34] A. Ballato, "Mindlin's Elastoelectrodynamics Problem," Proc. 42nd Annual Frequency Control Symposium, June 1988, pp. 6-13.

[35] G. Hayward, "A Systems Feedback Representation of Piezoelectric Transducer Operational Impedance," Ultrasonics, vol. 22, July 1984, pp.153-162.

[36] G. Hayward, C.J. MacLeod, and T. S. Durrani, " A Systems Model of the Thickness Mode Piezoelectric Transducer," J. Acoust. Soc. Am., vol. 76, August 1984, pp. 369-382.

[37] T. Koike, "A New Equivalent Circuit Approach to Magnetostatic Wave Devices," IEEE Ultrasonics Symposium Proc., October 1987, pp.221-225.

[38] A. Ballato and C.D. Bosco, "Acoustic Waves In Cubic Crystals: Networks for Semiconducting Vibrators, and Applications to Gallium Arsenide," Technical Report DELET-TR-79-9, U.S. Army Electronics R&D Command, Fort Monmouth, NJ 07703, April 1979, 45 pp.

[39] R. Lerch, "Piezoelectric and Acoustic Finite Elements as Tools for the Development of Electroacoustic Transducers," Siemens Forschungs- und Entwicklungs-berichte, Vol. 17, No. 6, 1988, pp. 284-290.

[40] R. Krimholtz, D.A. Leedom, and G.L. Matthaei, "New Equivalent Circuits for Elementary Piezoelectric Transducers," Electron. Lett., vol. 6, June 1970, pp. 398-399.

[41] D.A. Leedom, R. Krimholtz, and G.L. Matthaei, "Equivalent Circuits for Transducers Having Arbitrary Even-or-Odd-Symmetry Piezoelectric Excitation, IEEE Trans. Sonics Ultrason., vol. SU-18, July 1971, pp. 128-141.

[42] R. Krimholtz, "Equivalent Circuits for Transducers Having Arbitrary Asymmetrical Piezoelectric Excitation," IEEE Trans. Sonics Ultrason., vol. SU-19, October 1972, pp. 427-436.

[43] G.L. Matthaei, D.Y. Wong, and B.P. O'Shaughnessy, "Simplifications for the Analysis of Interdigital Surface-Wave Devices," IEEE Trans. Sonics Ultrason., vol. SU-22, Mar 1975, pp. 105-114.

[44] W.R. Smith and W.F. Pedler, "Fundamental- and Harmonic-Frequency Circuit-Model Analysis of Interdigital Transducers with Arbitrary Metallization Ratios and Polarity Sequences," IEEE Trans. Microwave Theory Tech., vol. MTT-23, November 1975, pp. 853-864; vol. MTT-24, July 1976, p. 487.

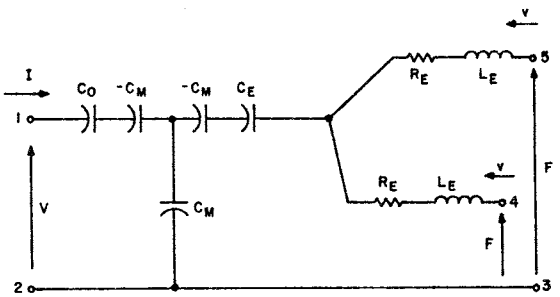


FIGURE 1. MASON: THREE-PORT, SINGLE RESONANCE, EQUIVALENT CIRCUIT.

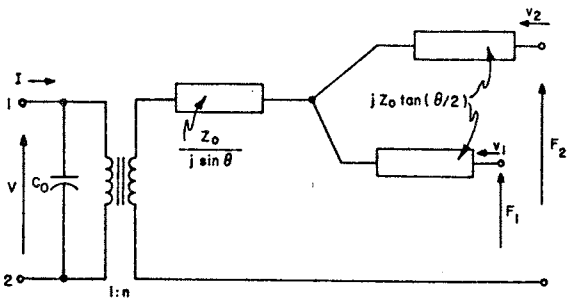


FIGURE 2. MASON: EXACT, THREE-PORT, SINGLE MODE-TYPE, EQUIVALENT CIRCUIT. APPLIED FIELD NORMAL TO WAVE PROPAGATION.

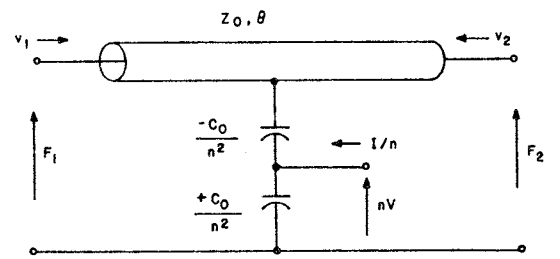


FIGURE 3. REDWOOD & LAMB: TRANSMISSION-LINE SCHEMATIC OF MASON'S EXACT, THREE-PORT, SINGLE MODE-TYPE, CIRCUIT. APPLIED FIELD PARALLEL TO WAVE PROPAGATION.

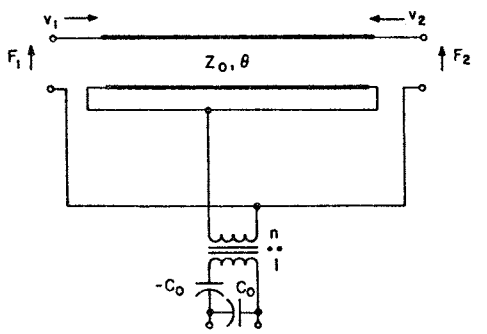


FIGURE 4.
TWO-WIRE TRANSMISSION LINE SCHEMATIC OF MASON'S EXACT, THREE-PORT, SINGLE MODE TYPE, CIRCUIT. APPLIED FIELD PARALLEL TO WAVE PROPAGATION.

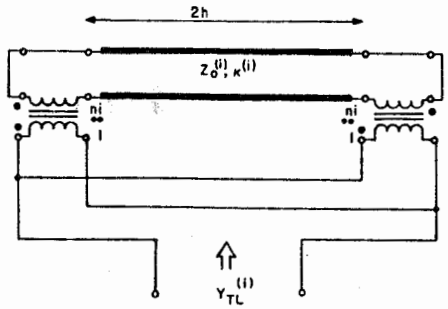


FIGURE 5.
TRACTION-FREE PLATE. REPRESENTATION OF SINGLE THICKNESS MODE, ELECTRICAL INPUT CIRCUIT OMITTED.

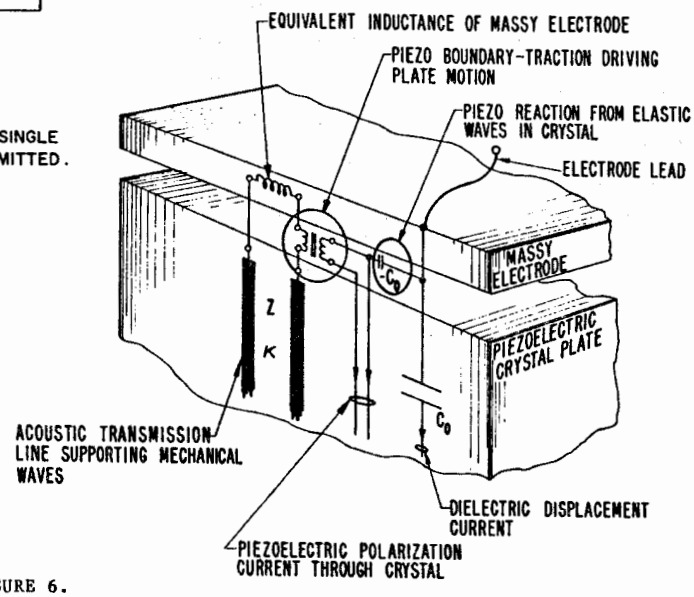


FIGURE 6.
EXPLODED VIEW OF CRYSTAL AND MASSY ELECTRODE, WITH ELECTRICAL ANALOG NETWORK SUPERPOSED

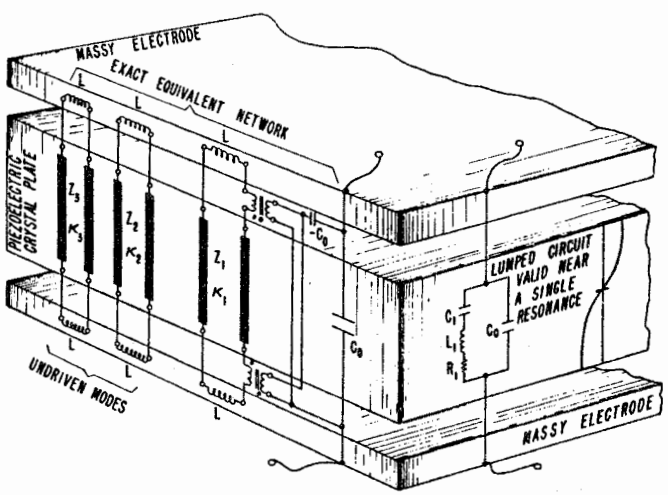
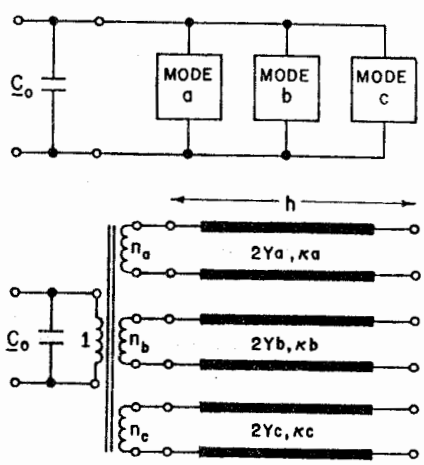


FIGURE 7.
EXPLODED VIEW OF A CRYSTAL PLATE WITH MASSY ELECTRODES, FOR THE CASE OF A SINGLE THICKNESS MODE PIEZOELECTRICALLY DRIVEN.
SUPERPOSED ARE THE EXACT AND LUMPED APPROXIMATE EQUIVALENT ELECTRICAL NETWORKS.

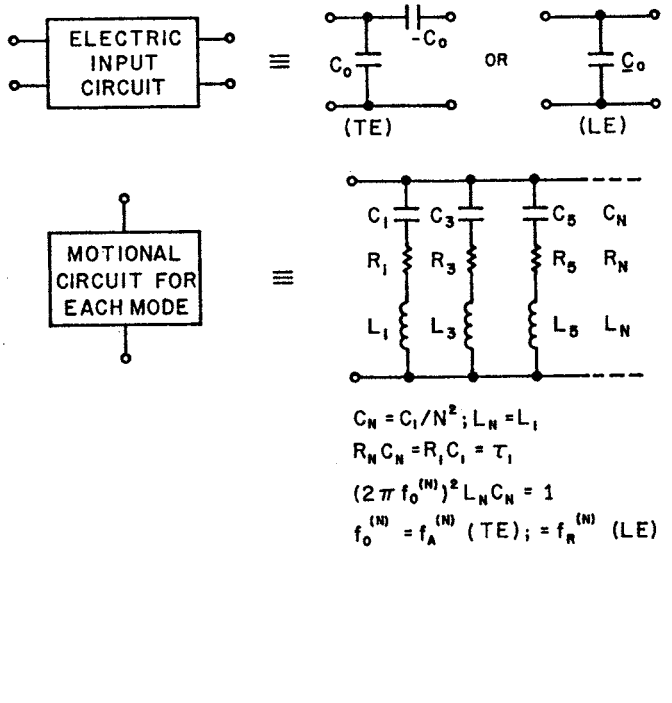


$$n_m^2 = \underline{C}_o \underline{k}_m^2 f_{Rm} (Y_m/2)^{-1}$$

FIGURE 8. BISECTED VERSION OF THREE-MODE, LATERAL-FIELD, PLATE NETWORK.

FIGURE 9.

EQUIVALENT NETWORK FOR SIMPLE THICKNESS MODES OF PLATES



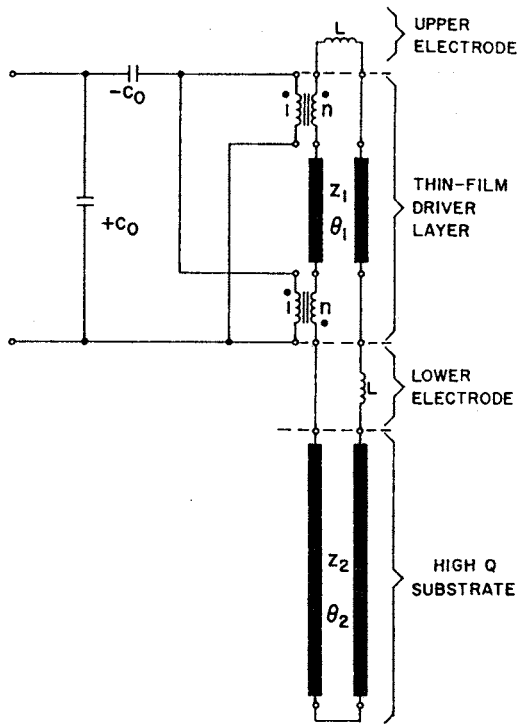


FIGURE 10. EQUIVALENT NETWORK OF COMPOSITE RESONATOR

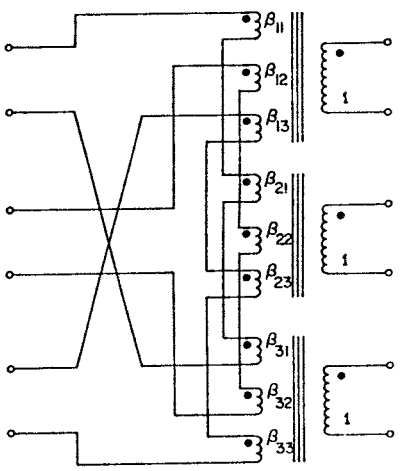


FIGURE 11. TRANSFORMER NETWORK REALIZING AN ORTHOGONAL TRANSFORMATION.

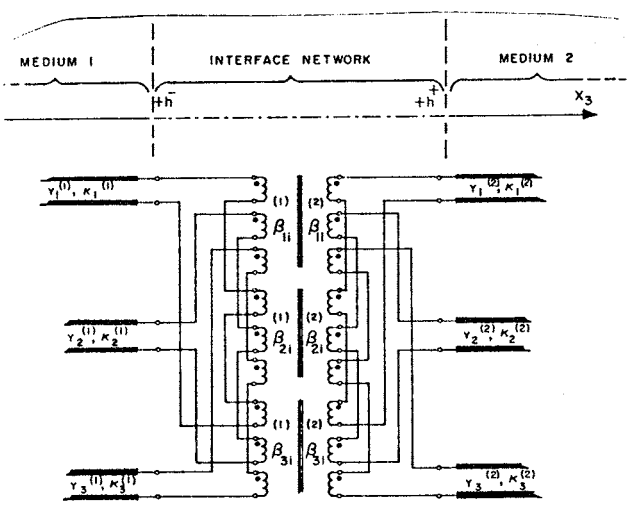


FIGURE 12.
EXACT REPRESENTATION OF MECHANICAL COUPLING AT THE INTERFACE OF TWO ARBITRARILY ANISOTROPIC MEDIA. PLANE ACOUSTIC WAVE PROPAGATION NORMAL TO BOUNDARY. PIEZOELECTRIC DRIVE TRANSFORMERS OMITTED FOR CLARITY.

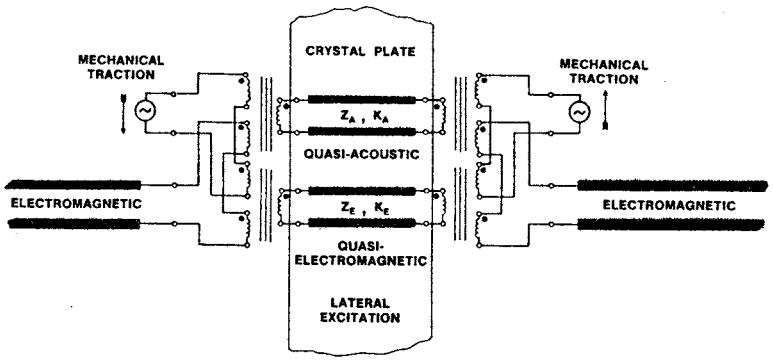


FIGURE 13. ANALOG NETWORK REALIZATION OF MINDLIN'S PROBLEM.

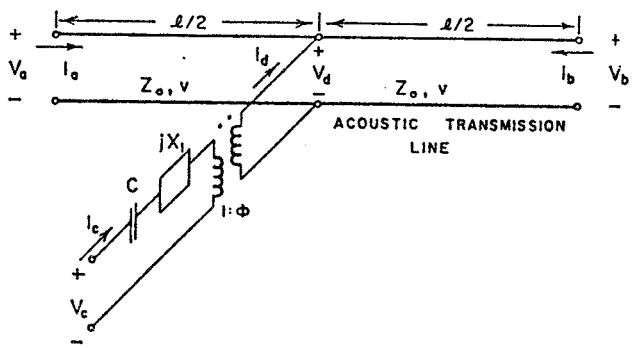


FIGURE 14. KLM EQUIVALENT CIRCUIT FOR THICKNESS-EXPANDER PLATE AND END-ELECTRODED LENGTH-EXPANDER BAR WITH EVEN EXCITATION.

INDEX DES AUTEURS - AUTHOR INDEX

AUTEUR/AUTHOR	PAGES
ABBOTT B.P.	56
ABBOUD B.	396
ADLER B.	385
ALLAN D.W.	1
AUDOIN C.	407
AZOUBIB J.	435
BAGHAI-WADJI A.R.	66, 323
BAHADUR H.	148
BALLATO A.	25
BALLANDRAS S.	50, 62
BALZER W.	224
BARILLET R.	407
BASTIEN F.	318
BAUCH A.	11, 243
BAUMONT F.	96
BEAUBIEN J.	288
BERNIER L.G.	404
BIGLER E.	62
BOY J.J.	129
BOYKO A.I.	296
BRANNSTROM E.	284
BRENDEL R.	377, 390
BRUNET M.	82
BUSCA G.	181, 404
CAPELLE B.	41, 358
CARRU H.	30
CASEY S.D.	262
CAUCHON E.	288
CEREZ P.	187
CHEN W.	337
CLAIRON A.	228
CLEMENTS A.	96
COLLIN F.	275
CROS D.	304
CUNIN P.	309
DAVIS A.	96
DAVIS D.D.	89
DAVIS J.A.	74
DE MARCHI A.	237, 456
DENERVAUD P.	105
DE ROOIJ N.	371
DETAINT J.	30, 41, 358
DIEP M.T.	105
DIMARCQ N.	187

AUTEUR/AUTHOR	PAGES
DRULLINGER R.E.	412
DULMET B.	164
ELLMEYER W.	143
ELTSUFIN E.	255
FELLNER J.	337
FELTHAM S.J.	207
FERAUDY D.	96
FOGNINI B.	110
GAGNEPAIN J.J.	314, 449
GAIGNEBET J.	96
GALLIOU S.	396
GARDIOL F.	105
GAUDRIOT L.	314
GAYGEROV B.A.	293, 296
GERARD E.	347
GEVORKYAN A.G.	205
GIORDANO V.	187
GLAZE D.J.	412
GNIEWINSKA B.	259
GODONE A.	18
GOUZHVA Y.G.	205
GRANVEAUD M.	82, 96
GROSLAMBERT J.	115, 449
GRUDLER P.	96
GUELLATI S.	228
GUILLON P.	304
HANSON W.	96
HASSANI C. El.	390
HATAT J.L.	96
HAUDEN D.	309
HAYASHI R.	249
HE J.W.	252
HEIGHWAY J.	56
HEINDORFF T.	243
HOUMMADY M.	309, 318
HOWE D.A.	89
HRUSKA C.K.	137
HUANG H.X.	252
HURRELL J.P.	21
HUTCHINS M.	74
IKEGAMI T.	190

AUTEUR/AUTHOR	PAGES
JACQUES C.	187
JAKOBY B.H.	323
JAMES B.J.	221
JEPERSEN J.	96
JESTKOVA N.D.	296
JOHNSON W.A.	21
JOLY C.	30, 41
KALLIOMAKI K.	211, 281, 401
KARUZA S.K.	21
KIRCHNER D.	96
KIRCHNER G.	96
KLAPPER H.	136
KLEPCZYNSKI W.J.	96
KOSHELYAEVSKY N.B.	415
KOSINSKI J.A.	25
KREMPL P.	143
KRUX P.	390
KURNIKOV G.P.	296
KUSTERS J.	337
LEE H.S.	412
LEONG A.	21
LESCHIUTTA S.	331
LEWANDOWSKI W.	96, 215
LIN C.F.	252
LIU J.T.	181
LOMBARDI M.	96
LOWE J.	412, 461
LUKASZEK T.J.	25
MAENO T.	249
MALLIKARJUN S.	25
MANGIN J.F.	96
MANNERMAA J.	401
MANSTEN T.	211, 281
MARIANNEAU G.	314, 377
MECKLENBRAUKER Ch.	66
MENARY P.	74
MEYER F.	82
MICHAUD A.	288
MOLLE R.	347
MORAWIEC L.	300
MORGAN D.P.	56
MOSKOVSKY A.M.	205
MOUREY M.	396

AUTEUR/AUTHOR	PAGES
NAEPFLIN P.	337
NAKADAN Y.	190
NEUBIG B.W.	385
NOVERO C.	18
OHSHIMA S.I.	190
OHTA Y.	249
OLBERT R.	343
OLIVIER M.	115
PAQUET P.	275
PEARCE P.R.	74
PETIT G.	215
PETIT P.	407
PFLAUM W.	337
PHILIPPOT E.	30, 358
PHILLIPS W.D.	228
PISTRE J.	309
PLANADE R.	309
PLANAT M.	372
POWELL W.	96
REBIERE D.	309
REICHINGER H.P.	323
RESSLER H.	96
ROBERT E.	377, 390
ROVERA G.D.	456
RUBIOLA E.	115, 237
RUSIN F.S.	293
SAGIN M.	74
SAITO H.	249
SALOMON C.	228
SANTARELLI G.	456
SCHULZKE L.	365
SCHRODER R.	11, 194
SCHUMACHER P.	123
SCHWARTZEL J.	30, 41, 358
SEBASTIAN M.T.	136
SELIC R.	143
SELIN B.	281
SHIRLEY J.H.	412
STADLER J.	143
STERN A.	255
STOERMER P.	337
SYSOEV V.P.	293, 296

AUTEUR/AUTHOR	PAGES
TAVELLA P.	435
THEOBALD G.	187
THOMANN P.	181, 249
THOMAS C.	215, 435
THORVALDSSON T.	56
THYR U.	96
TIERSTEN H.F.	157
TORRE J.M.	96
TOURDE R.	96
TREMBLAY P.	288
UHRICH P.	96
UMEZU J.	249
VALENTIN J.P.	396
VEILLET Ch.	96
VERNOTTE F.	82, 449
VIARD B.	129
VIENNET J.	407
VINCENT M.	82
VOIT F.J.	21
WALL P.K.	173
WALLS F.L.	461
WALLNOFER W.	143
WANG Y.Q.	201
WEISS M.A.	215, 442
WEISSERT T.	442
WHEELER P.J.	96
WILLEMIN P.	181
YANG D.H.	201
YELKIN G.A.	296
ZARKA A.	41, 358
ZECCHINI P.	129
ZHAI Z.C.	252
ZHANG W.Q.	252
ZHENG Y.	41, 358
ZIDEK H.P.	323
ZHOU Y.S.	157
ZIBROV A.S.	412
ZURCHER J.F.	105

Imprimé en France

par l'Imprimerie du
Conseil Général du DOUBS
Hôtel du Département
7, avenue de la Gare d'Eau
25031 - BESANCON CEDEX

Printed in France

by the Printing Office of
Conseil Général du DOUBS
Hôtel du Département
7, avenue de la Gare d'Eau
25031 - BESANCON CEDEX

Copyright by
5e Forum Européen Temps-Fréquence

Tous droits de reproduction et de traduction réservés
pour tous pays y compris l'U.R.S.S.

Imprimerie Départementale