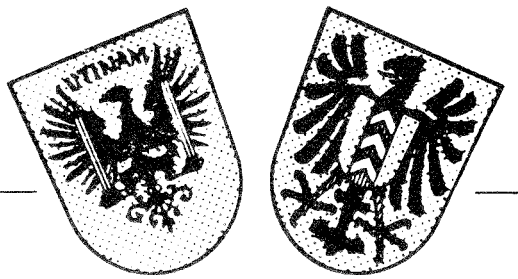


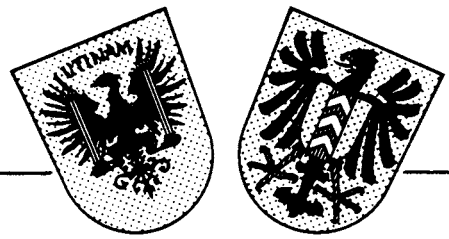
EFTF 90

Fourth European Frequency and Time Forum
Quatrième Forum Européen Fréquence et Temps

ACTES

PROCEEDINGS

13, 14, 15 March 1990
NEUCHÂTEL
Switzerland



E F T F 1 9 9 0

Fourth European Frequency
and Time Forum

Quatrième Forum Européen
Fréquence et Temps

Organised by the Scientific
Communities of Neuchâtel (Switzerland)
and Besançon (France)

Organisé par les Communautés
Scientifiques de Neuchâtel (Suisse)
et de Besançon (France)

March 13, 14, 15, 1990

13, 14, 15 mars 1990

Neuchâtel University, Aula
Espace Louis-Agassiz

Neuchâtel

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Table of contents

Introduction

Founding Committee, Organization and Secretariat, Support XIX

Scientific Committee XXI

Session chairmen XXIII

Exhibitors XXV

Opening Session

- Opening Remarks 3
G. Busca
Observatoire cantonal - Neuchâtel - Switzerland

- Welcome and official address of the Swiss Federal Office of Metrology 5
W. Schwitz
Swiss Federal Office of Metrology - Wabern - Switzerland

- The past and future of atomic time 9
N.F. Ramsey
Harvard University - Cambridge, Ma - USA

- ESA application for frequency and time* 13
M. Lopriore
European Space Agency - Noordwijk - The Netherlands

* Abstracts only

Evening Conference

- Time and Frequency in Metrology 17
B.W. Petley
National Physical Laboratory - Teddington - UK

Applications

APP-01 Navigation and positioning

1. Some aspects of space navigation system GLONASS : application in the interest of time and frequency determination..... 25
N.E. Ivanov, V.A. Salicshev, A.O. Kuprijanov
V/o Morsvjazsputnik - Moscow - USSR

Questions and answers 43

2. The GPS system : present and future 49
O. Bergene
Los Angeles Air Force Station - Los Angeles - USA

3. A civil NAVSAT overlay to GPS and GLONASS 81
C. Rosetti
European Space Agency - Paris - France

4. Applications of time and frequency to navigation and positioning 91
Ph. Hartl
University of Stuttgart - Stuttgart - FRG

Quartz, Crystals and Products

QCP-01 Review Papers

1. Innovative frequency standards : piezoelectric resonators and others 99
R.J. Besson
E.N.S.M.M. - L.C.E.P. - Besançon - France

2. Current Developments in Crystal Filters 107
R.C. Smythe
Piezo Technology Inc. - Orlando - USA

4. Quartz Crystal oscillators from their design to their performances 121
J.J. Gagnepain
C.N.R.S. - L.P.M.O. - Besançon - France

Frequency Standards

FS-01 Review Papers

1. Classical frequency standards 133
C. Audoin
C.N.R.S. - L.H.A. - Orsay - France

2. Advanced atomic frequency standards : Where are we heading ?* 143
J. Vanier
National Research Council - Ottawa - Canada

3. Characterization and measurement of frequency stability : twenty-five years of progress 145
J. Rutman, J. Groslambert
FIEE - Paris - France

Frequency Standards

FS-02 Review Papers

1. Precise Doppler tracking of interplanetary spacecraft* 151
B. Bertotti
Università di Pavia - Pavia - Italy

2. Time and relativity : Experimental and applied curved space-time chronometry* 153
C.O. Alley
University of Maryland - College Park - USA

Posters

PO-01 Short presentation of each poster by the chairman

1. Studies of a laser diode pumped Rb frequency standard 157
J.T. Liu, P. Thomann
Oscilloquartz SA - Neuchâtel - Switzerland
L. Zhang, G. Busca
Observatoire Cantonal - Neuchâtel - Switzerland

2. Estimation of some frequency shifts in an optically pumped clock taking into account modulation parameters 161
G.D. Rovera, E. de Clercq
L.P.T.F. - Paris - France
A. de Marchi
Universita di Ancona - Ancona - Italy

3. Optical pumping of cesium in applied external magnetic field : effect of changes in hyperfine coupling 163
C. Jacques,
C.N.R.C. - Ottawa - Canada
P. Tremblay, N. Cyr
University of Laval - Québec - Canada

4. Velocity distribution measurement in an optically pumped cesium beam resonator 169
A. Hamel, V. Giordano, P. Petit, C. Audoin
C.N.R.S. - L.H.A. - Orsay - France

5. GPS synchronized disciplined rubidium frequency standard* 175
J. Mc Nabb
Trak Microwave Corporation - Tampa - USA

6. A computerized time and frequency standards monitoring system 177
K. Kalliomäki, J. Mannermaa
University of Oulu - Oulu - Finland
T. Mansten
Technical Research Centre of Finland - Espoo - Finland

7. Remote control of a VLBI timing station 181
R. Ambrosini
Istituto di Radioastronomia CNR - Bologna - Italy

9. Application of time and frequency in Geodesy 187
W. Schlüter
Institut für Angewandte Geodäsie - Koetzting - FRG

10. Specification of the BVA-8600 OCXO for Space Use in ESA's Data Relay Satellite 195
L.G. Bernier, G. Busca
Observatoire cantonal - Neuchâtel - Switzerland

11. Calculated from the frequency domain : an update 197
F.L. Walls, J. Gary, A. O'Gallagher, L. Sweet, R. Sweet
NIST - Boulder - USA

12. Analysis of Fractal and Pareto - Levy Sets : Theory and Application 205
St.D. Casey
American University - Washington - USA

13. Investigations on time scale algorithms with simulated clock data* 213
H. Brandenberger, L. Prost
Office Fédéral de Métrologie - Wabern - Switzerland

14. Identification and measurement of different types of noises present in a frequency source 215
F. Vernotte
Observatoire de Besançon - Besançon - France
J. Groslambert, J.J. Gagnepain
C.N.R.S. - L.P.M.O. - Besançon - France

Posters

PO-02 Short presentation of each poster by the chairman

1. Russian radionavigation system as a time comparison method 219
T. Mansten, K. Kalliomäki
Technical Research Centre of Finland - Espoo - Finland

2. The advanced Mitrex modem with small satellite earth station 221
Ph. Hartl, K.H. Thiel
University of Stuttgart - Stuttgart - FRG
H. Reder, F. Schmidt
Dornier GmbH - Friedrichshafen - FRG

3. PRARE on GPS and/or GLONASS 225
Ph. Hartl, W. Schäfer
University of Stuttgart - Stuttgart - FRG
Ch. Reigber, W. Lechner
Deutsches Geodätisches Forschungsinstitut - München - FRG

4. SLALOM-Intersatellite ranging and Doppler measurement* 229
Ph. Hartl
University of Stuttgart - Stuttgart - FRG
Ch. Reigber
Deutsches Geodätisches Forschungsinstitut - München - FRG

5. High accuracy time management study (HAT Study) 231
D. Maeusli
ESTEC-WDI - Noordwijk - The Netherlands
U. Denskat, T. Schwammlein
Dornier GmbH - Friedrichshafen - FRG
M. Lombardi, J.M. Oberto
Matra - Toulouse - France

6. A method for determining the position of the antenna of Allen Osborne type GPS receivers 233
P. Uhrich, A.F. Paradon, R. Tourde
Observatoire de Paris - Paris - France

7. The phase stability of the synchronization pulse in a 600 km TV link system 237
J. Mannermaa
University of Oulu - Oulu - Finland

8. Automatic and fast television time comparison for the Zimmerwald satellite laser ranging station 239
P. Klöckler, Th. Schildknecht
University of Bern - Bern - Switzerland

9. Omega - VLF Signal Synthesizer 243
E. Rubiola
Politecnico di Torino - Torino - Italy

10. Renal calculi as new piezoelectric materials 245
R. Agarwal, V.R. Singh
National Physical Laboratory - New Dehli - India

11. Use of a tensorial representation of the dissolution slowness to predict etching shapes of quartz and GaAs crystals 247
C.R. Tellier, A. Brahim-Bounab
E.N.S.M.M. - L.C.E.P. - Besançon - France

12. Oscillators for real-time on-board applications of NAVSTAR-GPS 249
J.P. Lacroix
Sextant Avionique - Valence - France

13. Magnetic sensitivity of quartz oscillators 255
F. Deyzac
ONERA - Châtillon-sous-Bagneux - France

14. Linear phase filtering 259
B. D'Albaret, L. Bidart, G. Sillioc
Q.K. Piézoélectronique - Villeneuve-la-Garenne - France

Frequency Standards

FS-03 Future Prospects

1. Progress at NIST on absolute frequency standards using stored ions 267
D.J. Wineland, J.C. Bergquist, J.J. Bollinger, W.M. Itano, D.J. Heinzen, S.L. Gilbert, C.H. Manney, M.G. Raizen, C.S. Weimer
NIST - Boulder - USA

2. Atoms in optical molasses : applications to frequency standards 273
W.D. Phillips, P.D. Lett, S.L. Rolston, C.E. Tanner, R.N. Watts, C.I. Westbrook
NIST - Gaithersburg - USA
C. Salomon, J. Dalibard,
L.S.H. - C.N.R.S. - Paris - France
A. Clairon
L.P.T.F. - Paris - France
S. Guellati
Laboratoire Aimé Cotton - Orsay - France

3. Future prospects for the cold hydrogen maser 277
R.F.C. Vessot, E.M. Mattison
Smithsonian Astrophysical Observatory - Cambridge - USA
R.L. Walsworth, I.F. Silvera
Harvard University - Cambridge, Ma - USA
M.L. Dingus
Alabama Cryogenic Engineering - Huntsville - USA

Resonators, Oscillators and Materials

ROM-01 Resonator Fundamentals

2. 3-d nonlinear finite element analysis; Determination of the static force induced frequency perturbation of quartz-crystal resonators 285
K. Truempy
ETA SA - Grenchen - Switzerland

3. Determination of the complete tensor of nonlinear piezoelectricity of quartz using the resonator method 295
C.K. Hruska
York University - Downsview - Canada

4. ZT-cut HF resonator 301
J. Hermann
CSEM SA - Neuchâtel - Switzerland

Applications

APP-02 Time Transfer and Navigation

1. A review of recent advances in GPS time comparisons 307
B. Guinot, W. Lewandowski, C. Thomas
Bureau International des Poids et Mesures - Sèvres - France

2. Geostationary navigation satellites as a means of time transfer 313
S. Starker, H. Nau, J. Willner
DLR - Wessling - FRG
G.V. Kinal
International Maritime Satellite Organisation - London - UK

3. GPS/GLONASS - Receiver 321
K.H. Thiel
University of Stuttgart - Stuttgart - FRG

4. The use of ionospheric data in GPS time transfer 327
M. Weiss, T. Weissert
NIST - Boulder - USA
C. Thomas
Bureau International des Poids et Mesures
M. Imae
Communications Research Laboratory
K. Davies
National Oceanic and Atmospheric Administration

5. Towards subnanosecond GPS time transfer using geodetic processing techniques 335
Th. Schildknecht, G. Beutler, W. Gurtner, M. Rothacher
Astronomical Inst. University - Bern - Switzerland

Sensors and Actuators

SA-01 Sensors and actuators

1. Chemical etching of monolithic quartz vibrating beam sensor 349
F. Deyzac, D. Janiaud, S. Muller
ONERA - Châtillon-sous-Bagneux - France

2. Resonant magnetic-field sensor 351
E. Donzier, O. Lefort
Sextant Avionique - Valence - France
S. Spirkovitch, F. Baillieu
ESIEE - Noisy-le-Grand - France

3. Double action piezoelectric motors 355
M. Brissaud, L. Eyraud
INSA- Villeurbanne - France
J.L. Dion
University of Québec - Québec - Canada

4. New bulk wave quartz accelerometer 359
R. Bourquin, B. Dulmet, R.J. Besson
E.N.S.M.M. - L.C.E.P. - Besançon - France

Frequency Standards

FS-04 Advanced Frequency Standards

1. Laser frequency stabilization with applications to atomic frequency standards 367
K. Nakagawa, M. Ohtsu
Tokyo Institute of Technology - Yokohama - Japan

2. Operation of a cesium beam resonator optically pumped with cesium D_1 line 373
N. Dimarcq, G. Théobald, P. Cerez
L.H.A., Université Paris-Sud - Orsay - France

3. Frequency standards in the far-infrared using laser cooled atomic beams 381
N. Beverini, E. Maccioni, F. Strumia
Università di Pisa - Pisa - Italy

4. Thermalization of ions in a penning and in a Paul trap 387
F. Plumelle, M. Jardino, M. Desaintfuscien, J. Yu
C.N.R.S. - L.H.A. Université Paris-Sud - Paris - France

Resonators, Oscillators and Materials

ROM-02 SAW Devices

1. State-of-the-Art of SAW resonators 395
Ph. Defranould, L. Penavaire, S. Calisti, F. Raymond
Thomson Sintra ASM/DTAS - Valbonne - France

2. SAW low-loss filters for GHz-Applications 405
K. Anemogiannis, S. Berek
Siemens AG - Munich - FRG

3. Coupled SAW resonators - Theory and experiment 411
P. Durrant
MEDL - Lincoln - UK
J. Heighway, D.P. Morgan
Roditi International Corporation Ltd - London - UK

4. Characteristics of the various modes of acoustic waves in 36° YX-$LiTaO_3$ substrate 419
Y. Zhang, J.J. Gagnepain, D. Hauden
LPMO - CNRS - Besançon - France

Applications

APP-03 Synchronized Systems

Dedication to the memory of Philippe Monier

1. Frequency sources for space applications 429
S.J. Feltham
European Space Agency - Noordwijk - The Netherlands

2. Synchronization of telecommunications Networks 439
H.J. Bosshard, P.A. Probst
Research and Development, Swiss PTT - Bern - Switzerland

3. Distributed Master Clock System for Synchronization of a 2048 kHz Telecomm Network 451
S.M. Bass, R.E. Simons, D.W. Martin
Frequency and Time Systems, Inc. - Beverly - USA

4. Hybrid system of communication and radio determination using two geostationary satellites 463
S. Ohmori, Y. Matsumoto, E. Morikawa
Ministry of Posts and Telecommunications -
Kashima, Ibaraki - Japan
M. Wakao
R & D Center for Radio Systems Hirakawa - Tokyo - Japan

5. High Resolution Microwave Techniques in atmospheric pollution determination 469
E. Schanda
University of Bern - Bern - Switzerland

Resonators, Oscillators and Materials

ROM-03 Resonators, Oscillators and Measurements

1. PI-Network Measurement of the Load Resonance Frequency without Load Capacitor 481
B. Neubig
Tele Quarz GmbH - Neckarbischofsheim - FRG

2. Theoretical analysis of infrared ovens for tactical oscillators 487
S. Galliou, M. Mourey, J.P. Valentin
L.C.E.P. - E.N.S.M.M. - Besançon - France
R. Delaite
Institut de Génie énergétique - Belfort - France

3. Low phase noise, high frequency TCXO - Actual State-of-the-Art..... 495
L. Bidart, E. Girardet, J. Huet, P. Renoult
Q.K. Piézoélectronique - Villeneuve-la-Garenne - France

4. Electrostatic effects on quartz crystal resonators 501
R.J. Williamson
STC Technology Ltd - Harlow - UK
A.M. Harrison
STC Components - Harlow - UK

5. Amplitude and phase instabilities induced by nonlinearity and spatial dispersion in a quartz crystal resonator 509
M. Planat
C.N.R.S. - L.P.M.O. - Besançon - France

Frequency Standards

FS-05 Cesium Standards

1. Improving the long term stability of cesium beam frequency standards 517
A. De Marchi
Università di Ancona - Ancona - Italy

2. Experimental data on a new Ramsey cavity for cesium atomic clocks 523
H. de Boer, B. Fischer, T. Heindorff, R. Schröder
Physikalisch-Technische Bundesanstalt -
Braunschweig - FRG

3. Study of the correlations among the frequency changes of the contributing clocks to TAI 527
P. Tavella
Istituto Elettrotecnico Nazionale - Torino - Italy
C. Thomas
BIPM - Sèvres - France

4. Digital synchronous detector and frequency control loop for cesium beam frequency standard 543
L. Tivon Sing, J. Viennet, C. Audoin
L.H.A. - C.N.R.S. - Université Paris-Sud - Orsay - France

5. Preventive maintenance and stability of commercial cesium clocks 549
G. Fréon
Observatoire de Paris, LPTF - Paris - France

6. An Automated System for Measuring Velocity Distribution in Cesium Beam Tubes 555
S.K. Karuza, W.A. Johnson, J.P. Hurrell,
M.F. Bottjer, F.J. Voit
The Aerospace Corporation - Los Angeles - USA

Resonators, Oscillators and Materials

ROM-04 Materials

1. Progress on rotated $GaPO_4$ Y-cut BAW resonators 567
J. Stadler, P. Krempl, G. Engel
AVL List GmbH - Graz - Austria

2. Recent advances in Berlinite growth and characterization 577
E. Philippot, A. Ibanez, A. Goiffon
Université de Montpellier - Montpellier - France
X. Buisson, R. Arnaud
SICN - Annecy - France
B. Capelle, A. Zarka
Université de Paris - Paris - France
Y. Toudic, J. Schwartzel, J. Détaint
C.N.E.T. - Bagneux - France

3. New designs for resonators and filters using lithium tantalate 589
J. Détaint, J. Schwartzel, H. Carru,
R. Lefèvre, C. Joly,
C.N.E.T. - Bagneux - France
B. Capelle, Y. Zheng, A. Zarka
L.M.C. Université P. & M. Curie - Paris - France

4. Chemical etching of Berlinite 601
O. Cambon, A. Goiffon, E. Philippot,
Université de Montpellier - Montpellier - France
J.P. Aubry
C.E.P.E. - Argenteuil - France
J. Détaint
C.N.E.T. - Bagneux - France

5. Etch channels in quartz monocrystals - factors of influence 609
O. Bignon, B. Brottier
S.I.C.N. - Annecy - France

Applications

APP-04 Time Transfer : Future Prospects

1. Remote time and frequency comparisons now and in the future 619
D.W. Allan
NIST - Boulder - USA

2. Design considerations for KU-band two-way satellite time transfer terminals 631
D. Kirchner
Technical University Graz - Graz - Austria

3. Further development of the ECS time synchronization experiment* 639
O. Buzek, J. Cermak , J. Vondrak,
V. Pettiti, F. Cordara, A. Cenci
URE CSAV - Prague - Czechoslovakia

4. PRAREE - The Extended Version of the Precise Range and Range Rate Equipment 641
Ph. Hartl
University of Stuttgart - Stuttgart - FRG

5. Time transfer in the United Kingdom via a direct broadcasting satellite 643
P.R. Pearce, J.A. Davis, D.R. Allen,
B.R. Swabey, R. Jefferies, D.J. Knight
National Physical Laboratory - Teddington - UK
P. Menary
British Telecom International - Hillmorton - UK

Resonators, Oscillators and Materials

ROM-05 Material and Processing

1. Technology and characteristics of chemically milled miniature quartz crystals 653
B. Studer, W. Zingg
Micro Crystal - Grenchen - Switzerland

2. X-ray diffraction study of quartz crystals with insitu electric field 659
M.T. Sebastian, H. Klapper
Technische Hochschule Aachen - Aachen - FRG

3. Spherical trigonometry for quartz blank fast routine X-ray control 667
H. Mérigoux, J.F. Darces, D. Auguie
Université de Franche-Comté - Besançon - France

4. Recent results in study of piezoelectric resonators by X-ray topography using synchrotron radiation 675
B. Capelle, Y. Zheng, A. Zarka
Université P. & M. Curie - Paris - France
J. Détaint, J. Schwartzel
C.N.E.T. - Bagneux - France

5. The use of annular diamond saws for cutting of quartz plates to high angle accuracy 683
J. Dowsett, R.B. Spencer, P. Morley
STC Components LTD - Harlow - UK

Applications

APP-05 Time Transfer in Geodynamics and Radioastronomy

1. Applications of hydrogen masers to deep space tracking and radio science 689
R.L. Sydnor
California Institute of Technology - Pasadena - USA

2. Time and Frequency Facilities at the Applied Physics Laboratory 695
J.J. Suter
The Johns Hopkins University - Laurel - USA

3. Universal Time (UT1) Determination : Present and Future 699
W.E. Carter, D.S. Robertson
National Ocean Service, NOAA - Rockville - USA

4. The European VLBI Consortium 703
G. Setti
Università di Bologna - Bologna - Italy

5. The "Radioastron" space VLBI project 707
E. Valtaoja
University of Turku - Turku - Finland

Postdeadline Posters

1. Some Results on the line Q Degradation in Hydrogen Masers 713
L.G. Bernier, G. Busca
Observatoire Cantonal - Neuchâtel - Switzerland

2. The time and frequency system at the Matera Space Geodesy Center 715
A. Cenci, L. Garramone
Telespazio SpA - Rome - Italy
G. Bianco, B. Pernice, G. Sylos Labini
ASI - Matera - Italy

3. Phase errors distribution obtained for a class of clocks 719
M.O. Hongler, C.V. Quach
University of Geneva - Geneva - Switzerland

4. Sweeping of thin AT and SC cut quartz crystals under vacuum and air atmosphere 723
J.P. Bachheimer, P. Palleau
Université J. Fourier - Saint-Martin-d'Hères - France

Author index 725

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Founding Committee

R. Besson — *Directeur du Laboratoire de Chronométrie, Electronique et Piézoélectricité (LCEP)*
Professeur à l'Ecole Nationale Supérieure de Mécanique et des Microtechniques (ENSMM), Besançon - France

M. Ecabert — *Directeur de la Fondation suisse pour la recherche en microtechnique (FSRM), Neuchâtel - Suisse*

J.J. Gagnepain — *Directeur du Laboratoire de Physique et Métrologie des Oscillateurs (LPMO) du CNRS, Besançon - France*

P. Kartaschoff — *Ingénieur, Chef de Section, Recherche et Développement de la Direction Générale des PTT, Berne - Suisse*

A. Rémond — *Ingénieur de Recherche et Chef du Service Chronométrique de l'Observatoire de Besançon - France*

B. Schlueter — *Keres Engineering S.A., Neuchâtel - Suisse*

Organization and Secretariat

Swiss foundation for research in microtechnology
Fondation suisse pour la recherche en microtechnique (FSRM)
Rue de l'Orangerie 8

2000 Neuchâtel - Switzerland (Tel. ++38 24 52 00)
(Fax ++38 24 71 45)

With support of

Republic and Canton of Neuchâtel / City of Neuchâtel / Swiss Federal Office of Metrology, Wabern / European Space Agency/European space research and technology centre (ESA/ESTEC), Noordwijk/The Netherlands / Swiss Academy of Engineering Sciences (SATW/ASST) / Sextant Avionique, Valence/France / Telespazio S.p.A., Rome/Italy / Muller Brewery / Helvetia Insurance / Swiss Bank Corporation / University of Neuchâtel / University of Stuttgart / Jacobs Suchard Tobler, Neuchâtel / United Tobacco Factories, Neuchâtel / PTT (regional post office/telephone and telegraph - DAT) / Tourist Office, Neuchâtel

SCIENTIFIC COMMITTEE

Audoin C.	C.N.R.S. - L.H.A., Orsay	France
Ballato A.	US Army Electronics Techn. & Devices Lab., Fort Monmouth	USA
Beaussier J	ONERA, Châtillon	France
Besson R.	E.N.S.M.M. - L.C.E.P., Besançon	France
Boloix R.	Observatorio de San Fernando (ROA) San Fernando	Spain
Brunet M.	C.N.E.S., Toulouse	France
Busca G. Chairman 1990	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
Dowsett J.	STC Components LTD, Harlow Essex	UK
Duchaussoy H.	D.R.E.T., Paris	France
Gagnepain J.-J.	C.N.R.S. - L.P.M.O., Besançon	France
Graf E.	Oscilloquartz S.A., Neuchâtel	Switzerland
Granveaud M.	Observatoire de Paris, Paris	France
Guinot B.	Bureau International des Poids et Mesures, Sèvres	France
Hartl Ph.	Universität Stuttgart, Stuttgart	FRG
Henaff J.	CNET - DIT - ASC, Issy-les-Moulineaux	France
Hruska C.	York University, Downsview	Canada
Kalliomäki 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 Quarz 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

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session Chairmen

Giovanni BUSCA	*OBSERVATOIRE CANTONAL* *Rue de l'Observatoire 58, CH-2000 NEUCHATEL*	OP
John DOWSETT	*STC COMPONENTS LTD* *Edinburgh Way, GB-CM20 2DD HARLOW Essex*	ROM-03
Stephen FELTHAM	*ESA/ESTEC. RC* *Postbus 299, Keplerlaan, NL-2200 AG NOORDWIJK*	APP-03
Michel GRANVEAUD	*LPTF-OBSERVATOIRE DE PARIS* *61, av. de l'Observatoire, F-75014 PARIS*	FS-04
Philipp HARTL	*UNIVERSITAT STUTTGART* *Keplerstrasse 11, D-7000 STUTTGART 1*	APP-02
Angus HASTON	*GEC HIRST RESEARCH CENTRE* *East Lane, GB-HA9 7PP WEMBLEY, Middlesex*	ROM-04
Helmut HELLWIG	*NATIONAL INST. OF STANDARDS & TECHNOLOGY* *Bldg. 101, Room A-526, USA-20899 GAITHERSBURG, MD*	FS-03
Carl K. HRUSKA	*YORK UNIVERSITY* *4700 Keele Street, CDN-M3J 1P3 DOWNSVIEW, Ontario*	QCP-01
David J.E. KNIGHT	*NATIONAL PHYSICAL LABORATORY* *Building 5, N.P.L., GB-TW11 OLW TEDDINGTON, Middlesex*	PO-01
Bernd W. NEUBIG	*TELE QUARZ GMBH* *Landstrasse 2, D-6924 NECKARBISCHOFSHEIM 2*	SA-01
Norman F. RAMSEY	*HARVARD UNIVERSITY* *USA-02138 CAMBRIDGE, Ma*	FS-01
Gerd F. RIHA	*SIEMENS AG* *P.O. Box 83 09 52, D-8000 MUNCHEN 83*	ROM-02
Robert C. SMYTHE	*PIEZO TECHNOLOGY INC.* *P.O. Box 547859, USA-32854 ORLANDO, FLA*	ROM-01
Siegfried STARKER	*DLR* *D-8031 WESSLING*	PO-02
Richard L. SYDNOR	*CALIFORNIA INSTITUTE OF TECHNOLOGY* *4800 Oak Grove Drive, USA-91109 PASADENA, Ca*	APP-04

Pierre THOMANN	*OSCILLOQUARTZ S.A.* *Brévards 16, CH-2002 NEUCHATEL 2*	FS-05
Jacques VANIER	*NATIONAL RESEARCH COUNCIL CANADA* *Montreal Road Campus, CDN-K1A OR6 OTTAWA, Ontario*	FS-02
Fred L. WALLS	*NATIONAL INST. OF STANDARDS & TECHNOLOGY* *325 Broadway MS 576, USA-80303 BOULDER, CO*	APP-05
Gernot WINKLER	*US NAVAL OBSERVATORY* *34th Massachusetts AV NW, USA-20392-5100 WASHINGTON, DC*	APP-01
Walter ZINGG	*MICROCRYSTAL DIV. OF ETA* *Schild-Ruststrasse 17, CH-2540 GRENCHEN*	ROM-05

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Exhibitors

AUSTRON INC.	*P.O. Box 14766*	*USA-78761-4766 Austin, TX*
C.E.P.E., CIE D'ELECTRONIQUE & PIEZOELECTRICITE	*44, avenue de la Glacière*	*F-95105 Argenteuil Cedex*
CSEM S.A.	*Maladière 71*	*CH-2007 Neuchâtel*
FRISCHER ELECTRONIC S.A.	*Rue du Trésor 9*	*CH-2000 Neuchâtel*
FSRM	*Rue de l'Orangerie 8*	*CH-2000 Neuchâtel*
OBSERVATOIRE CANTONAL	*Rue de l'Observatoire 58*	*CH-2000 Neuchâtel*
OFFICE DU TOURISME DE NEUCHATEL ET ENVIRONS	*Rue Place-d'Armes 7*	*CH-2000 Neuchâtel*
OSCILLOQUARTZ S.A.	*Brévards 16*	*CH-2002 Neuchâtel*
RODITI INT. CORP. LTD.	*130 Regent Street*	*GB-W1R 6BR London*
SICN	*4, rue du Radar*	*F-74008 Annecy*
SOCIETE DE BANQUE SUISSE	*Fbg de l'Hôpital 8*	*CH-2001 Neuchâtel*
STC QUARTZ CRYSTAL	*20-22 Edinburgh Way*	*GB-CM20 2DE Harlow, Essex*
TELE QUARZ GMBH	*Landstrasse 2*	*D-6924 Neckarbischofsheim 2*
TOYO COMMUNICATION EQUIPMENT CO. LTD.	*20-4 Nishi-Shimbashi 3-Ch*	*J-Minato-Ku, Tokyo 105*
TRAK MICROWAVE CORP.	*4726 Eisenhower Boulevard*	*USA-33634-6391 Tampa, FL*
UNVERDROSS TECHNIK	*Tutzinger Hof Platz 6*	*D-8130 Starnberg*

OPENING SESSION

OP

Chairman: G. Busca

Session OP

Opening Remarks by Giovanni Busca

I am very happy to welcome all the participants with a special note for those coming from outside Europe who are contributing in a very substantial way to this Forum program and, hopefully, its sucess.

Before introducing the official session I would like to make a few remarks on the characteristics of this 4th EFTF.

This year there is an emphasis on the atomic frequency standards and their applications, even if the field of piezoelectric resonators and the associated applications is so well represented that we were forced to go to parallel sessions.

The emphasis on atomic frequency standards turns out to be timely chosen.

First, as most of you are aware, leaders of the research in the field of the atomic clocks have been honored of the Nobel Prize of Physics in 1989. And we are thankful to the founder of this field, Professor Norman F. Ramsey, to be here with us.

Secondly, we are entering a phase of open collaboration, I should say "Perestroika", for the civil applications of navigation satellite systems conceived, for military use, in the West and East and based on atomic clocks. This gigantic satellite navigation systems are the American Global Positioning System (GPS) and the Soviet Global Navigation System (GLONASS).

So I think that it will be interesting to see through this conference how basic idea, that Professor Ramsey will illustrate soon, have generated a multitude of techniques and applications in such different fields as communications, radioastronomy, geodesy, relativity and navigation.

To me this is an example of the creative power of the man oriented towards positive and interesting issues.

Session OP

Welcome and official address of the Swiss Federal Office of Metrology

W. Schwitz, Swiss Federal Office of Metrology, CH-3084 Wabern/Bern

Mr. Chairman, dear guests and participants of this 4th European Frequency and Time Forum, Ladies and Gentlemen

For the national metrology institutes worldwide the units and standards for frequency and time are without doubt of greatest importance. It is therefore a great pleasure for me as a representative of this country's metrology institute, the Swiss Federal Office of Metrology, to say a few introductory words at this widely recognized and acknowledged Forum.

Today, the units of frequency and time are - thanks to the invention of atomic clocks - by far the most precisely realizable units in our international system SI. In addition or as a consequence, they are also decisive for other units as well. The unit of length for instance, is today realized and intercompared at the highest level of accuracy by frequency standards, such as stabilized lasers. The unit of electrical voltage is reproduced at extreme precision by means of Josephson arrays relying on the Josephson frequency-to-voltage ratio.

In fact, the units of time and frequency were probably the first ones used by mankind at all. For example, the quantities of length or work were measured by time intervals like hours, days or years. These in turn were determined through the daily rotation of our planet or by astronomy.

The history of time measurements reflects the development of our culture and technology very nicely and excitingly. Shadow bars, called "gnomons" in Ancient Greece, were used to indicate hours and, already three thousand years ago, water clocks were in use. Only much later, namely at the end of the 13th century, the

first mechanical clock was invented now allowing time keeping independent of natural processes. The wheel clock became smaller and smaller spreading from public towers to royal palaces, symbolizing might and authority. The clock became even portable and only much later also affordable for everyone. And today, interestingly enough, through the definition of the meter, based on a fundamental constant, the speed of light, we are back to determine lengths by means of time intervals again!

Very impressing is the increase in accuracy of time intervall determinations during the last four hundred years. Then the most sophisticated mechanical clock was accurate to about 1 min a day. Today we achieve with commercial atomic clocks a few ns stability a day, corresponding to a progress of more than 10 decades!

The Swiss Federal Office of Metrology, established 127 years ago as the Federal Verification Office, is directly involved in frequency and time measurements only since 1982 and is now, in close collaboration with other institutions in this country like the Neuchâtel Observatory, contributing to the international atomic time.

The Federal Office of Metrology is well aware of the great importance frequency and time determinations occupy in metrology, research, navigation and technology today, to give only a few examples. Through a renewed contract with the canton of Neuchâtel including research and development in the field of frequency and time at the Observatory of Neuchâtel, more contributions to time and frequency metrology shall be possible in the future.

Besides its important quantitative side time also has a qualitative aspect. It is a measure and a quality sign of life. We all know about the problem of dividing up the available time and that it cannot be multiplied.

You already devoted more or less time for travelling to this place from all over the world. We are honored by your presence and we are convinced that international meetings like this one enable scientists and engineers to discuss and finally solve problems. We can easily conclude this is a good way - even the best way to achieve the necessary progress.

Scientists and engineers have learned to work together for a common aim and

they have therefore friends all over the world. Politicians, however, are faced with a lot of difficulties when they try to approach problems with their political tools. I therefore believe, in the future it would be an important task for scientists and engineers to exercise their influence on politicians in order to have them solving the problems in our world in an unconditional way.

Science and engineering have permitted the technical progress which is beneficial to mankind. But science and engineering too, have created the potential of powerful destructive weapons. I think, the scientists and engineers should also be urged to help bringing at last international peace.

Technical progress has to serve mankind in the legitime interest of a happy family of nations in our world. I am sure this wish inspires all of us. Let us therefore help politicians to transform this wish into reality!

Again, I am convinced that meetings like this one are of great value not only to us personally, but also to a continuing progress in the well-beeing of humanity.

I wish you therefore three days full of quality and satisfaction.

In this respect, we could not have wished a better start as the one to follow with the first talk, to be given by last year's Nobel Prize winner Norman Ramsey, who bears such great merits for the development in frequency and time measurement achieved during the last decades.

Session OP

THE PAST AND FUTURE OF ATOMIC TIME

Norman F. Ramsey
(Harvard University, Cambridge, MA 02138, USA)

ABSTRACT

The early history of atomic time and frequency standards is reviewed. The most accurate and stable present standards are described. Prospective future improvements, particularly with trapped ions and atoms are discussed.

THE PAST

The first successful magnetic resonance experiments were those of Rabi and his associates[1,2] with molecular beams in 1938. The initial experiments measured the interactions of nuclear magnetic moments with external magnetic fields, but radiofrequency spectra were soon observed that depended on internal interactions within molecules[2] or atoms[3,4]. Some of the radiofrequency spectra lines were almost independent of external conditions[5] and could therefore be used as the highly stable periodic component of an atomic clock[6,7].

In 1949 Ramsey[8,9] invented the separated oscillatory field method which provided narrower resonances, eliminated first order Doppler shifts and was useable at much higher frequencies. In 1952 Kusch, Lyons Sherwood and others[10] did some initial work on a separated oscillatory field atomic cesium clock, but the work was soon discontinued. In 1954 Zacharias[11] stimulated renewed interest in an atomic cesium clock and began[12] the development of a commercial atomic cesium clock. The first atomic beam apparatus extensively used as an actual frequency standard was constructed in 1955 by Essen and Parry[13]. From 1956 on atomic frequency standards developed rapidly with major contributions coming from a number of laboratories in many different countries[7,14]. Stabilities and accuracies of about 10^{-13} have been achieved with atomic cesium clocks[14] and the second, by international agreement, has been defined as the duration of 9,192,631,770 cesium periods.

Concurrent with the advances in atomic beam clocks, a number other atomic and molecular clocks were developed[7]. Microwave absorption devices using molecular resonances, such such as the NH_3 inversion transition, were developed by Townes and others[7]. Later Townes[15] discovered the maser principle and developed the ammonia maser which operated well but lacked the stability to be competitive.

The combination of Kastler's optical pumping technique[16] with Dicke's use of buffer gases[17] provided strong oscillations free from first order Doppler effects, so optically pumped rubidium can be used as a frequency standard. Although other atomic clocks have greater accuracy and stability, rubidium clocks are frequently used since they are much lighter and less expensive.

The atomic hydrogen maser was invented by Kleppner and Ramsey[18] and a number of scientists[7] contributed to its improvement. In the hydrogen maser, atoms in the higher hyperfine state of atomic hydrogen are stored in a teflon coated bottle inside a tuned microwave cavity where by stimulated emission they emit coherent highly stable microwave radiation. The stability of a hydrogen maser can be better than 10^{-15} over several hours.

Townes and Schawlow[19] first pointed out the possibility of applying the maser principles at infrared and light frequencies and the first successful laser was made by Maiman[20]. Since then laser developments have occurred at a rapid pace with dramatic improvements in frequency range, power and stability. Major advances came from the suppression of first order Doppler shifts by such techniques as two photon absorption spectroscopy[7] and from the development of successive chains of laser frequencies so that the laser frequencies could be calibrated in terms of cesium.

At present different atomic clocks can be chosen to fit the need. When high stability is not required, optically pumped rubidium cells can be chosen for their low cost, small size and light weight. When greater accuracy and stability is required, cesium beam tubes are usually used. For the highest stability requirements, as in long baseline radio astronomy, hydrogen masers are often required. When optical frequencies need to be measured lasers must be used even though they are not yet absolute standards.

THE FUTURE

Many improvements are being developed for the clocks that are currently widely used. Optically pumped atomic oscillators are being improved by the use of different atoms, by suitable wall coatings which permit the omission of buffer gases[7] and by combining optical pumping techniques with those of atom trapping and cooling. Possible improvements for the hydrogen maser include operation at low temperature, electronic cavity tuning, operation in a passive mode and new confinement surfaces, such as superfluid ^{4}He or other liquids. Improvements for atomic beam clocks include the use of different atoms or molecules, operation at higher frequencies, the use of two beams going in opposite directions to compensate for phase shifts, the use of lasers to select states and detect beams and better velocity definition to reduce uncertainties in second order Doppler corrections. Optical frequency standards can be improved by the use of narrower lines, by the adaptation to optical frequencies of the separated oscillatory field method[8], by improved frequency chains to compare different frequency standards and by the various ion and atom trapping and cooling techniques discussed below.

Dehmelt[21,22] first used electromagnetic ion traps in radiofrequency resonance studies. Penning traps overcome the limitations of the Earnshaw theorem by confining the ions in one direction with an inhomogeneous electric field and in the two other orthogonal directions with a uniform magnetic field. Alternatively, suitable inhomogeneous electric fields can provide focussing in all three directions in Paul or radiofrequency traps which alternately provide focussing and defocussing in each direction but with an average focussing in all directions. Ion traps have the advantage that the observed transition frequencies are approximately independent of the trapping fields. Originally the trapped ions had high kinetic energy (approximately 1 eV) and excessively high second order Doppler shifts. However, Cutler[29] and others[29] have cooled the ions by collisions with helium gas. Lasr cooling, as proposed by Wineland and Dehmelt[22,23] and by Hansch and Schawlow[24], can go even to much lower temperatures by shining intense laser light at the frequency of an allowed optical transition onto a trapped atom or ion slightly below the resonance frequency so the light pressure is greatest on the ion when it is approaching the light. Dehmelt and Wineland and others[7] have used this technique to cool trapped ions to temperatures of a few micro Kelvin where the second order Doppler shifts are negligible. This is a very promising technique for a stable clock with the principal limitation being the low ion density required to avoid space charge effects and the low density in turn reduces the signal to noise ratio.

In 1985, Phillips[25] and his associates used laser cooling of a focussed atomic beam to slow electrically uncharged atoms and even reverse their velocity. Since the Doppler shift changes as the atom slows down, either the atomic optical frequency or the laser frequency must change for the slowing to continue. Phillips did so by having the atoms pass through a region where the magnetic field gradually changed as the atom moved along its path. Alternatively, Hall[7], Wieman[7] and others[7] have changed or "chirped" the laser frequency as the atom has slowed down. The success of atom cooling permits atoms to be stored in weak traps so there has been a virtual explosion of new ideas and developments in laser trapping of atoms during recent years. Laser trapping forces on neutral atoms can arise either from the gradients of the laser electric field interacting with the induced electric dipole moment of the atom (gradient or dipole force traps) or by the transfer of momentum in the absorption and emission of radiation (spontaneous radiation or scattering force traps), with the gradient traps being intrinsically weaker. When slow atoms are introduced into a region with oppositely directed laser beams along three orthogonal directions at frequencies slightly below the resonance frequency, the atoms will be laser cooled in whatever direction they move. Although such "optical molasses" does not provide a stable trap Chu[26], Pritchard[27], Wieman[27], Cohen-Tannoudji[28] and others[29] have combined optical molasses with either gradient or radiation trapping to obtain useful atom traps. Atoms stored in such traps have been cooled to temperatures of a few micro-K.

Phillips and his associates[30] recently made the startling experimental discovery that Na atoms could be cooled to lower velocities than had previously been considedred to be the theoretical limit. For several years it was thought that there was a theoretical limit, often called the Doppler limit, below which the atoms could not be cooled by laser cooling. This theoretical limit quite reasonably occurs when $k_B/2 = \hbar\Gamma/4$ where Γ is the spontaneous emmission rate from the excited atomic state and k_B is the Boltzmann

constant. For Na, this theoretical temperature limit T = 240 μK. In contrast. Phillips and associates[30] found experimentally a cooling to 40 μK. Another theoretical limit to cooling was thought to be the recoil limit which occurs when the recoil energy $(\hbar k)^2/2M$ of the atom of mass M emits a photon of momentum $\hbar\mathbf{k}$. However Aspect, Cohen-Tannoudji and their associates have achieved[31] transverse cooling of ^{4}He atoms in the triplet metastable state to 4 μK, well below the usual one dimensional Doppler cooling limit of 23 μK and even below the recoil limit of 4 μK. These surprising results have been explained by Cohen-Tannoudji and his associates[31] as due to a velocity selective process based on coherent trapping of atomic populations that can most easily be interpreted in terms of their[32] dressed atom models.

The principal disadvantage of an atom trap is that the strong laser fields that provide the trapping also distort the energy levels and resonance frequencies of the atoms. Usually this makes it necessary to turn the trapping lasers off while the resonance is being with a consequent reduction in storage time. One possibility for increasing the effective storage time is to allow the atoms to have a small vertical component of velocity so that they rise up and then fall under gravity, in some cases passing through two coherent oscillatory fields[8] on the way up and down, as recently done successfully by Chu[33] and his associates with microwaves and as planned by others at laser frequencies.

Some atomic clocks, such as cesium beam tubes, have much better long term accuracy than short term stability in which case the shorter term stability of the clock can be greatly improved by suitably coupling it to a fly wheel oscillator of high short term stability. Very high Q cavities appear are good for this purpose and superconducting cavities are particularly promising since they are highly stable. Circuits with either optical or electrical feedback from a high Q cavity markedly diminish the noise and increase the short term stability.

Another development which may affect the future of atomic clocks is the very high observed stability of some milli-second pulsars. At least one of these appears to have a long term stability comparable to the best atomic clocks. If that continues to be the case and other equally stable pulsars can be found, it is possible that the long term definition of the unit of time might at some future time be given in terms of pulsar frequencies. However, even if such a change should occur it in no way diminishes the need for good atomic clocks since they are essential in measuring and comparing the pulsars and in the measuring other frequencies and time intervals.

It is apparent that major advances in atomic clocks can be anticipated with just the ideas that already exist, but the field is so dynamic and challenging that one one should also anticipate totally new and unanticipated ideas which will arise and greatly affect the future.

REFERENCES

1. Rabi II, Zacharias JR, Millman S and Kusch P, Phys Rev, 53(1938), 318 and 55(1939), 526.
2. Kellogg JMB, Rabi II, Ramsey NF and Zacharias JR, Phys Rev, 55(1939), 729, 56(1939), 728 and 57(1940), 677.
3. Kusch P, Millman S, Rabi II, Phys Rev, 57(1940), 765.
4. Millman S and Kusch P, Phys Rev, 57(1940), 438.
5. Kusch P and Taub H, Phys Rev,75(1949), 1477.
6. Rabi II, Phys Rev, 67(1946), 199.
7. Ramsey NF, History of Atomic Clocks, Journ of Res of NBS, 88(55)(1983), 301.,
8. Ramsey NF, Phys Rev 76(1949), 966 and Phys Today, 33(7)(1980). 25.
9. Ramsey NF, Molecular Beams, Oxford Press, (1956 and 1985).
10. Sherwood JE, Lyons H, McCracken RH and Kusch P, Bull Am Phys Soc, 27(1)(1952), 43.and Sc American 196(2)(1957), 71.
11. Zacharias JR, Phys Rev,94(1954), 751.
12. Reder FH, USASRDL Tech Rep 2230(AD 265452), (1961).
13. Essen L and Parry VL, Nature, 176(1955), 280, 284.
14. Hellwig H, Evenson KM and Wineland DJ, Phys Today, 31(1978), 23.
15. Gordon JP, Zeiger HZ and Townes CH, Phys Rev, 95(1954), 282.
16. Kastler A, Journ Phys, 11(1950), 225.
17. Dicke RH, Phys Rev, 89(1953), 472.
18. Kleppner D, Goldenberg HM and Ramsey NF, Phys Rev, 126(1962), 603.
19. Schawlow AL and Townes CH, Phys Rev, 112(1958), 1940.
20. Maiman TH, Nature, 187(1960), 493.
21. Dehmelt HG, Phys Rev, 109(1959), 381.
22. Wineland DJ and Itano W, Phys Today, 40(6)(1987), 34 and Science, 237(1987), 612.
23. Wineland DJ and Dehmelt HG, Bull Am Phys Soc, 20(1975), 60 and 61.
24. Hansch TW and Schawlow AL, Opt Comun, 13(1975),
25. Phillips WD, et al, J. Opt Soc Am, B2(1985), 1751.

26.Chu S, et al, Phys Rev Lett, 55(1985), 48 and 57(1986), 314.
27.Pritchard D and Wieman C, Phys Rev Lett, 57(1986)310.
28.Aspect A, Cohen-Tannoudji C, et al Phys Rev Lett, 57(1986), 1688.
29.Proceedings of the Eleventh Intnl Conf on Atomic Physics.
30.Lett PD, Phillips WD,et al Phys Rev Lett, 61(1988), 169.
31.Aspect A, Cohen-Tannoudji C et al Phys Rev Lett, 61(1988),
32.Dalibard J, Cohen-Tannoudji C et al Jour. Opt. Soc. Am. B2(1985), 1707.
33.Kasevich M, Riis E, Chu S and DeVoe RS, Phys. Rev. Lett.,63(1989), 612

Session OP

ESA APPLICATION FOR FREQUENCY AND TIME

M. Lopriore

Head, RF Systems Division

ESA/ESTEC (European Space Research and Technology Centre),
Noordwijk (The Netherlands)

The European Space Agency Convention notes that the Agency purpose is to 'provide for and promote, **for exclusively peaceful purposes**, cooperation among European States in space research and technology and their space applications with a view to their being used for scientific purposes and for space application systems.'

At its inception ESA had Scientific satellites to deal with (ESRO-II) but now the range of application has grown from Telecom and Earth Resources satellites to the Space Station Columbus and the Scientific Programme has expanded to include a variety of missions, some well known to the public e.g. the Giotto one.

All these applications require different on-board oscillators and complex timing : the common element remaining the space environment with it tight needs of reliability, thermal control and lifetime.

This presentation will outline the early applications at ESA for frequency and time and then review a few current and future cases that set a challenge for the future.

Evening conference (banquet)

Session Evening conference

Time and Frequency in Metrology

B W Petley
Centre for Basic Metrology
National Physical Laboratory
Teddington, Middlesex TW11 OLW
United Kingdom

Abstract.

Time and frequency are by far the most accurately measured of the dimensioned physical quantities. This means that as soon as one can find an accurate parametric relationship between frequency and the quantity of interest, that quantity may thereafter be very accurately measured in terms of frequency. Quantum physics is an increasingly fruitful source of such relationships. The present situation and the future prospects are reviewed.

Introduction.

Metrology is the phrase used to describe the art and science of accurate measurement. The SI units are almost universally used in science and technology and other unit systems have generally been defined to have an exact relationship to the SI. The system embodies seven *base units*: those of time, length, mass, quantity of electricity, thermodynamic temperature, amount of substance, and luminous intensity. There is a second class, the *supplementary units* which provide units of angle and solid angle. These are regarded as dimensioned or dimensionless according to the choice of the user. Along with these there are now some nineteen named combinations of the base units which are known as *derived units*.

There are three distinct SI time units, these are: (i) the base unit, the *second* (13th CGPM, 1967) which is the unit of time and time interval; (ii) the derived unit known as the *hertz* (11th CGPM, 1960), used to characterise the frequency of periodic phenomena; and (iii) the derived unit of activity known as the *becquerel* (15th CGPM, 1975) which is used for the stochastic phenomenon of radioactive decay to denote the number of disintegrations per second. Units such as the *minute* (60 s), *hour* (3600 s), and *day* (86 400 s) were recognised by the CIPM in 1969 as being units used alongside the SI. The *week*, *month*, and *year*, remain in common use.

Although the *metre* and *kilogram* were carefully defined by the CGPM at their first meeting in 1889, the *second* does not appear to have been explicitly defined by them. Everyone knew that it was

<<equal to the fraction 1/86 400 of the mean solar day>>

and it was the task of the astronomers to define precisely what was meant by the phrase *'mean solar day'*. By 1956, however, the CIPM decided to follow the lead of the 1955 General Assembley of the IAU, and formally decided that

<<the second is the fraction 1/31 556 925.974 of the tropical year for 1900 January 0 at 12 hours ephemeris time>>.

This somewhat inaccessible definition was adopted by the 11th CGPM in 1960, at which time the term SI was also adopted. This definition only endured until 1967 when the 13th CGPM decided that metrology needed a better unit of time than astronomy could provide and, by its resolution 1, decided that

<<The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the caesium-133 atom>>

At the same time the earlier definition was abrogated.

During the last forty years a number of quantum phenomena have been discovered which enable the measurement of a physical quantity to be largely performed in terms of a frequency measurement. Metrology is a changing science and seen from a distance it is readily apparent that:

(i) the accuracy with which we are able to make measurements has been increasing over the centuries;

(ii) that this process has been speeding up.

Other physical quantities show much the same effects - although the accuracy achieved lags that of time measurement by more than thirty years. The accuracy with which we may measure different physical quantities depends ultimately on the accuracy with which the base units are known. At present these range from the region of 10^{-14} fractional accuracy for time and frequency measurements through to the 10^{-4} to 10^{-3} accuracy of luminous intensity measurements.

We should not focus on the accuracy of measurement of the unit alone - for the bulk of users are almost certainly working well away from the unit. For almost every physical quantity we get a characteristic bath-tub shaped curve (figure 1) as we lose accuracy in going to multiples and sub-multiples of the unit. The well known Allan-variance curve is but one example of this type of curve.

Electrical Measurements and the radian per second.

The radian per second is the unit of angular frequency and, although it occurs frequently in theoretical discussions, there is little necessity to produce and measure exact angular frequencies.

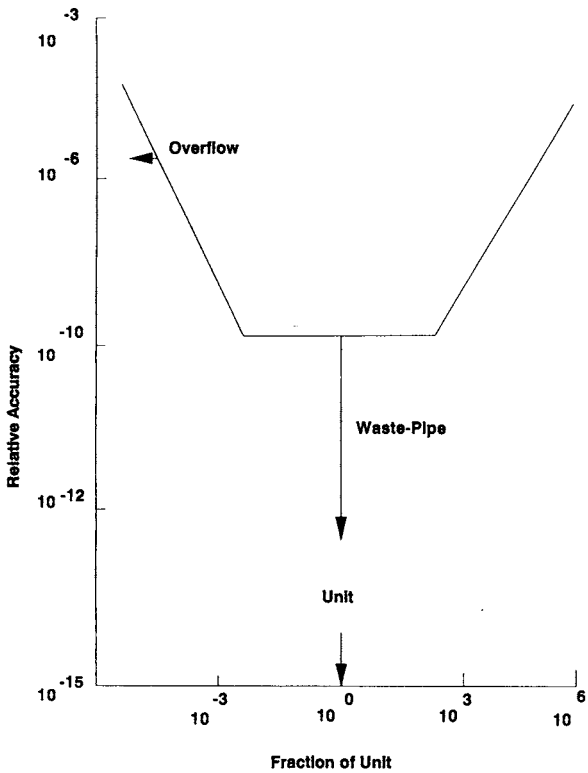


Figure 1. The metrological bath-tub: the accuracy decreases as one moves away from the unit. The overflow takes one down to the fundamental constants.

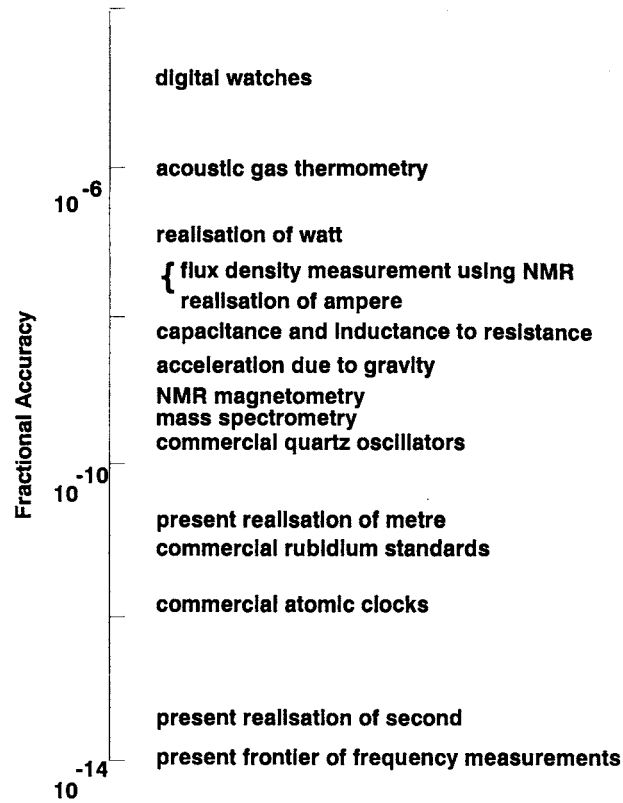


Figure 2. Some applications of time and frequency metrology.

An example of the measurement of small angular frequencies would be in gyroscopes. For use in navigation stabilities and sensitivities of better than 10^{-4} °/hr are required in ships and 10^{-2} °/hr in aircraft. For the latter this gives the required drift performance of less than one nautical mile of drift error per hour.

The optical gyroscopes make use of the Sagnac (1913) effect to achieve high sensitivities to small angular velocities. The time difference Δt in an area A and angular velocity Ω is given by: $\Delta t = 4A\Omega/c^2$. Michelson and Gale demonstrated this effect in 1925 by using an interferometer to detect the Earth's rotation rate of 15 °/hr. For A = 1 m², and Ω = 10 °/hr, ΔL (= $c\Delta t$) = 7 .10^{-13} m.

In order to detect rates of 0.050 °/hr fibre optic systems require phase detection sensitivities corresponding to about 10^{-8} λ in 10^{9} λ path length. The very best suspended electrical gyroscopes achieve 10^{-5} °/hr, while the present-day active and passive laser and fibre optic ring gyroscopes have around a 100 times poorer sensitivity. The sensitivities are normally expressed in °/hr rather than rad/s, the conversion factor being 1°/hr = 5 .10^{-6} rad/s. In laser ring gyroscopes the counter-rotating beat frequencies of <1 mHz have been measured, enabling long term stabilities of 10^{-3} °/hr to be achieved. Performance in some cases is approaching the quantum noise limit of 10^{-4} °/$(\mathrm{Hr})^{1/2}$.

A frequency in rad/s is required experimentally in the transfer from resistance to capacitance and inductance measurements. Standard capacitors and inductors are near exact multiples or sub-multiples of the henry and farad respectively, and standard resistors are similarly multiples of the ohm. The transfer across from resistance involves the equations:

$$R = \omega L; \text{ and } R = 1/\omega C$$

Obviously, if C, L, and R are all to be nearly exact multiples of the unit, then ω must also be an integral number of radians per second. An angular frequency of 10^4 radians per second is convenient for this. One often finds that AC electrical measurements specify that the measurements are made at, or within 1 Hz, of 1592 Hz. This frequency is regarded as being sufficiently close to the desired angular frequency of 10^4 radian per second. No attempt is made to synthesise the exact angular frequency.

Time and length measurements: fringe counting and fringe velocity

Length measurements may be readily converted into a problem of counting interference fringes. The definition of the metre adopted by the CGPM in 1983 invokes the relationship:

$$l = c \cdot t$$

so that the length, l, of a resonator may be derived from its resonant frequency, f, via $f = c/2l$. This is a useful non-contact method that has been applied by Gallop at the NPL to the measurement of superconducting cavities at 4.2 K with an accuracy of around 10 nm on a 50 mm diameter cylinder (2 parts in 10^7). An increasing number of such relationships are being applied at the highest levels of accuracy, particularly in the visible region where lasers may be slaved to remain tuned to a resonant mode of a cavity.

The measurement of the acceleration due to gravity poses an interesting problem for, if we have a cube corner reflector that is part of a Michelson interferometer in free fall, the laser interference fringes have a frequency $2gt/\lambda$ that is, the quantity of interest is the rate of change of frequency:

$$g = (\lambda/2)\cdot\frac{df}{dt}$$

Up to the present there has been no attempt to make the measurement other than as the time for the cube to traverse a set number of fringes (the fringe frequencies are usually less than a few MHz). In this way Faller and colleagues at JILA, and others elsewhere, have been able to measure the acceleration due to gravity to a few parts in 10^9.

The lunar ranging experiments are an example of very precise length measurement in terms of a time measurement. The Earth-Moon distance can now be monitored at the centimetre level - corresponding to a resolution of about 2 parts in 10^{11}. Similarly the orbits of satellites may be tightly mapped by timing laser pulses to provide much useful information about the higher order terms for the earth's gravitational field.

Gravity wave detectors require the detection of very small strains in the form of length changes, and detectors having strain sensitivities of between 10^{-21} to 10^{-18} are being constructed in several locations. Some of these map length changes into frequency changes and therefore make strong demands on the short-term stability of the frequency discriminator. They require frequency discrimination rather than absolute accuracy.

The realisation of the electrical units

The accuracy with which the electrical units could be realised in the past was limited by the necessity to measure the linear dimensions of coils of wire very accurately. For many years the ohm was realised via a precisely constructed mutual inductor, however, the advent of the Thompson and Lampard calculable capacitor meant that the farad could be realised instead. The capacitance so realised was $\frac{10^7}{4\pi c^2}\cdot \ln 2$ F/m, about 1.953 pF/m, and was obtained by changing a length of a screening electrode - the displacement was readily measured by counting laser fringes in a Michelson interferometer. (Frequency enters the measurement via the 1592 Hz frequency applied to the AC bridges).

In this way the realisations of the ohm have been improved to yield fractional accuracies of the order of a few parts in 10^8. It ought to be possible to improve the accuracy further, particularly by measuring the length change as a frequency by changing the length of a resonant cavity. The advent of the quantised Hall resistance for MOSFET semiconductors, operated at low temperatures and strong magnetic fields, has enabled the drift of resistance standards to be confirmed as being of the order of 0.05 $\mu\Omega$ per year - the present ability to detect such small resistance drifts and allow for them promises well for the future.

Until recently the volt was realised by some form of attracted electrode electrometer (the most accurate of these being the attracted mercury surface determination of Clothier, Sloggett and others at the NML in Australia) and the ampere via a variant on the Ayrton-Jones current balance. Kibble at the NPL proposed the method which forms the basis for recent electrical realisations of the electrical watt. Essentially a coil of wire is suspended from one arm of a mass balance such that it is partially in the field provided by a magnet (both superconducting magnets and permanent magnets are being used).

A current I, whose value is known in terms of the maintained ampere, is passed through the coil and the force on the coil balanced by adding or removing mass m, from the scale-pan. The coil is then displaced vertically with a uniform velocity v, and the induced emf E, measured in terms of the maintained volt. The product of the voltage and the current are then given by:

$$E\cdot i = mgv$$

where g is the acceleration due to gravity. Thus, the power is known in terms of both the maintained watt and the mechanical watt. The coil velocity is measured as a laser fringe frequency of about 38 kHz.

The first attempt by Kibble and others at NPL has yielded a fractional accuracy of a part in 10^7 and one may anticipate a further order of magnitude improvement - at which point it will be possible to check the stability of the kilogram and possibly replace it by some definition involving energy instead, that is, mass might become a derived unit in the longer term.

Frequency metrology and the fundamental physical constants

The involvement of the fundamental physical constants with frequency forms the basis for the use of the phrase *quantum metrology*. This is particularly appropriate in the case of the transition frequency ν between two energy levels E_1 and E_2 such that

$$\nu = \{E_1 - E_2\}/h$$

This is the equation of elementary quantum mechanics which follows from the equation of motion of a wave function ψ:

$$-ih\frac{d\psi}{dt} = H\psi,$$

given that E_1 and E_2 are eigenvalues of the Hamiltonian H. The present definition of the second means that the time unit in the wave equation is necessarily of the same character as that for the frequency - which serves to illustrate that the definition of the second has a profound effect on the equations of motion in physics, both at the macroscopic level and at the quantum level.

Other relationships are used in metrology as well, such as that in synchrotron radiation for luminous intensity measurements (which are at the 0.05% level), or the speed of sound at zero pressure v_o, which, in a gas of molar mass M, at a temperature T, is given by:

$$v_o^2 = \gamma RT/M$$

This equation is used in acoustic thermometry and for measuring the gas constant, R, (and hence the Boltzmann constant k) to a few parts per million. The quantity γ is the ratio of the specific heats is 5/2 for argon. The velocity is obtained from the resonant frequency of a cavity. The accuracy is limited by the need to extrapolate to zero pressures to remove the effects of the virial coefficients. Uncertainties in these and the temperature distribution tend to set the present accuracy limits.

Generally in metrology increased accuracy and the increased frequency associated with the measurement often go together. One can see this feature particularly with measurements of the fundamental physical constants.

Increased accuracy often comes too through increasing the measurement time since this enables noise and other effects to be averaged, the accuracy often improving as $(\text{time})^{1/2}$. In metrology the observation time is usually limited by other considerations. This is particularly the case if the measurement is part of a control process.

Magnetometry, gyromagnetic ratios and cyclotron frequency.

The relationship between the spin precession frequency and flux density, and between an ion cyclotron frequency and flux density, involve the equations: $\omega_s = \gamma_p' \cdot B$ and $\omega_c = Be/m_p$ respectively, where the γ_p' is the gyromagnetic ratio of the proton, e is the elementary charge and m_p the rest mass of the proton. The use of the prime symbol indicates that γ'_p is not an exact quantum relation, for the spin precession frequency refers to that of protons which are in a spherical sample of water and hence are subject to small temperature and sample shape dependent frequency shifts etc of some 25.689(15) parts per million. The gyromagnetic ratio of the proton has been used for some forty years as a transfer constant to enable magnetic flux densities to be measured in terms of water NMR frequencies with an accuracy limited by the accuracy of the realisation of the ampere. Other gyromagnetic ratios are known with respect to water and there are also, for example, rubidium and ^{3}He magnetometers.

The accuracy of flux density measurements by this means is limited in the first instance by the accuracy of the realisation of the volt. Beyond that the effects of the water sample shape, its temperature, and the effects of the container and impurities in the water (dissolved oxygen in particular) limit the accuracy to about a part in 10^8 - a lot depends on the nature of the substitution method involved.

Although the particle spin precession frequency and the ion cyclotron frequency are expressed above as angular frequencies, it is convenient practically to measure and express them in hertz, consequently their values are usually tabulated in hertz, as well as in radians per second.

Although not requiring accurate frequency measurement, SQUID magnetometers operate at carrier frequencies up to the microwave region and enable magnetic flux to be measured in terms of the magnetic flux quantum $h/2e$ (around 10^{-15} Wb). These are relative devices since it is difficult to measure accurately the area through which the flux is passing.

Ratio experiments and mass spectrometry

There are some important ratio measurements which may be made in terms of frequency. These involve the ratios of ion cyclotron frequencies and/or spin precession frequencies. These include the g-2 measurements for the electron, positron and muon, and the ratio of the proton to electron mass. The accuracy of such measurements has made a major leap forward through the pioneering work on ion traps by Dehmelt, Van Dyck and others at the University of Washington. Some of the improvement must also be attributed to the use of higher flux densities with correspondingly higher frequency discrimination.

Conventional mass spectrometry achieved its 10^{-9} fractional accuracy by putting very stringent geometric requirements on the ion orbit, by slits etc, so that the arrival of the ion at the detector depended critically on its starting phase with respect to the applied cyclotron frequency. Thus, in the ion substitution mass synchrometers of Lincoln Smith in Princeton and of Mamyrin in Leningrad, the ions made only two revolutions, taking around a microsecond, before being detected. This meant that the timing discrimination (or the equivalent phase discrimination) was effectively of the order of 10^{-14} s for a fractional resolution of a part in 10^8. Time measurement was not required since ion substitution was achieved entirely by changing both the applied AC and DC potentials, and AC frequencies. Ion traps are beginning to allow the properties of anti-matter to be explored and it may be that the findings of these studies will impact on metrology.

Josephson effects

The prediction of the Josephson tunnelling effects in superconductivity in 1962, and their subsequent validation at the part per million level at the University of Pennsylvania, the NPL and elsewhere by the end of 1969, meant that it had been established that there was a very powerful relation between the potential difference between two closely separated superconductors in the presence of a microwave frequency f given by: $2eV = hf$ where n was an integer, h the Planck constant and e the elementary charge. This equation has been verified in various ways to the very highest levels of precision, including independence of the operating conditions (temperature, material, frequency), and by now must be one of the most accurately verified effects in physics.

In the meantime one of the most accurate measurements involving frequency to date has been made via a null measurement by Lukens and others at Stony Brook, which involved verifying that the change in the gravitational potential of a microwave photon in traversing a vertical distance of about a centimetre was the same as the change in potential of the electron pair wave. The discrimination achieved was of the order of a part in 10^{18} - needless to say the interpretation of such a measurement is slightly open to debate! We have arrived at the situation where the volt is maintained in terms of agreed values of the Josephson effect K_J, and the ohm in terms of the quantised Hall resistance von Klitzing constant R_K.

Can we measure h directly in terms of the mechanical units?

The quantum equation $E = h\nu$ occurs so frequently that one wonders whether it can be verified directly in terms of the Joule regarded as a mechanical unit. The equation can and has, of course, been verified for cases whereby the energy involved is primarily an electrical energy. In one sense this remains true yet in another this is no longer the case as we see below.

If the voltage E_J across a Josephson junction (which is biased to be on the nth voltage step when irradiated with a microwave frequency f) is maintained at a value equal to the voltage across a von Klitzing device which is biased onto the second quantised Hall resistance step, then the power in the Hall resistor is:

$$E_J^2/R_H = n^2 f^2 h/2.$$

Today the volt is maintained via the Josephson effect K_J and the ohm via the quantised Hall resistance R_K and the numerical relationships between these and the units are exactly specified. Consequently, we see that the above realisation of the watt by Kibble and others is also a determination of h - always provided, of course, that the Josephson and von Klitzing relationships are exact.

Metrology and time variations of the fundamental constants

Following their interpretation in terms of Dirac's 'big numbers' some 45 years ago, there is always considerable interest in the question of whether or not the fundamental constants are constant. The evidence to date is that they are constant on the level of less than one percent over the age of the universe. Some measurements

have put even tighter constraints. In each case one must interpret the results with care. This is particularly the case with a dimensioned quantity since one cannot say whether it is the unit or the quantity that has varied (equally a null effect may result if they both varied by the same amount). Turneaure and Stein at Stanford were able to compare the frequency of a caesium clock against that of a high-Q superconducting microwave cavity and make tight inferences about the constancy of the fine structure constant in conjunction with other dimensionless constants. Thus the caesium frequency is represented by $h_{hf} = const.g_p(eh/2m_p)(eh/2m_e)(Z\alpha_o)^3$ where α_o is the Bohr radius, Z the caesium nuclear charge and g_p the proton g-factor. The cavity frequency depends on $f_{cavity} = const.hc/\alpha_o$, so that $\frac{f_{cavity}}{f_c} = const.g_p\left(\frac{m_e}{m_p}\right)\alpha^3$ from which Turneaure and Stein obtained:

$$1/\alpha \cdot d\alpha/dt < 3 \cdot 10^{-11} per yr$$

The particular significance of their work was that it enabled laboratory measurements over 12 days to set constraints on parameters expected to be constant over cosmological time-scales. This, along with studies of the isotropy of inertia, of Michelson-Moreley experiments, and of the black-body background radiation measurements, serves to emphasise that cosmology is important in modern metrology and *vice versa*.

There is still some debate about what happens to the fundamental constants at the high energies required by the GUT theories and the 'big bang' - but in order to discuss such possibilities one needs something else that is more stable to use as an invariant *'ruler'*!

Can we replace the present definition of the second?

During the next decade or so there is likely to be an intensive exploration of a number of other transitions with the aid of ion traps and atom traps. These narrow transitions will require interrogation sources of high spectral purity, and the required millihertz linewidths are being increasingly achieved for the optical (and near optical) laser frequencies. Many of these transitions will lie well beyond the microwave region and some may well have Q-values which are 10^8 or more times greater than that of the present generation of caesium beam and hydrogen maser frequency standards. There may be several transitions of comparable accuracy which are much superior to the accuracies achieved with caesium clocks. The advent of these will cause metrologists to consider very carefully what should be done about the present definition of the second.

There are some simple requirements for any new definition:

(i) for continuity the new unit must always lie within the reproducibility of the old;
(ii) it must be capable of being accessed by users;
(iii) it should immediately give an order of magnitude improvement in accuracy; and
(iv) it should be capable of further refinement.

The most desirable solution would appear to be a definition which was not specific to a particular transition in a particular ion or atom. This solution has already been adopted for the metre, and thereby avoided the necessity to redefine the metre each time that a higher performance stabilised laser was developed. It may, however, become very important to distinguish in the future between a time standard and a possible frequency standard - for the latter need not operate continuously!

Conclusion

As the accuracy of measurement advances throughout metrology, it will be necessary to take general and special relativity more and more into account. Consequently it will be more important to take care to use co-ordinate time, internal time, proper time, and so on, as is appropriate to the particular circumstances of the measurement. The studies of the millisecond Pulsars serve as a strong reminder that the whole of our terrestrial physics is modulated by changes in our gravitational potential such as results from the rotation of the Earth and Moon about one another and about the Sun. Hopefully we will eventually be able to detect such effects without recourse to Pulsars!

The accuracy with which we can realise the second continues to run ahead of the rest of the SI units. As we have seen, there are an increasing number of quantum effects which may be produced sufficiently cleanly in the laboratory to enable them to be used for purposes of defining or maintaining the SI units. We can expect in a general way that this process is likely to continue and that time may be the seed base unit from which the definitions of most of the other basse units will be derived.

Progress must slow eventually but we can see that there are probably some seven or more orders of magnitude to go before we reach the limits for frequency measurement - which will be reached when the averaging time for the ultimate transition is around about the working life of a metrologist. At the present rate of progress that condition may well be reached well before the centenary of the First Conference on Frequency Measurement and Control, meanwhile our endeavours to attain such accuracies will provide a lot of fun!

Abridged References.

[1] W K Clothier, G J Sloggett, H Bairnsfather, M F Currey and D J Benjamin, A determination of the volt, Metrologia, **26**, 9-46, 1989

[2] E R Cohen and B N Taylor, The 1986 evaluation of the fundamental physical constants, Rev Mod Phys, **57**, 1121-1148, 1988

[3] Faucheux, D Fayoux and J J Roland, The ring laser gyro, J Optics, vol **19**, 101-115, 1988

[4] J C Gallop, and W J Radcliffe, Shape and dimensional measurement using microwaves, J Phys E:Sci Instrum, **43**, 413-416, 1986

[5] A K Jain, J E Lukins and J -S Tsai, Test for relativistic gravitational effects on charged particles, Phys. Rev. Lett. **58**, 1165-1168, 1987

[6] B P Kibble, I A Robinson and J H Belliss, A realisation of the SI watt by the NPL moving-coil balance, NPL Report DES 88, 1988

[7] Le Systeme International d'Unites (SI), 5th Edition (BIPM: Paris 1985)

[8] I Marson and J E Faller, g - the acceleration of gravity: its measurement and its importance, J Phys E: Sci Instrum, **19**, 22-32, 1988

[9] B W Petley, The fundamental constants and the frontier of measurement, (Adam Hilger: London, 1988)

[10] B N Taylor and T J Witt, New international electrical reference standards based on the Josephson and Quantum Hall Effects, Metrologia, **26**, 47-62, 1989

[11] J P Turneaure and S R Stein, An experimental test of the time variation of the fine-structure constant, Atomic Masses and Fundamental constants 5, ed J H Sanders and A H Wapstra, 636-642, 1976

Applications

APP-01 Navigation and Positioning

Chairman: G.R. Winkler

Session APP-01 Number 1

SOME ASPECTS OF SPACE NAVIGATION SYSTEM GLONASS APPLICATION IN THE INTEREST OF TIME AND FREQUENCY DETERMINATION

Dr. Nikolay E. Ivanov, Dr. Vadim A. Salicshev
and Dr. Andrew O. Kuprijanov

v/o "MORSVJAZSPUTNIK" Satellite Communications and
Navigational Electronics Aids
1/4 U1.Rojdestvenka, Moscow 103759, USSR

At present the application of satellite navigation systems as the most perspective means for precise time and standard frequencies transmission is of worldwide importance.

Using of on-board atomic oscillators, transmitting from the satellite the precise parameters of its movement and measuring of the navigation parameters in user equipment based on the super precise measurement of the time-intervals and frequency of received signals, makes it possible to realize the navigation measurement and the measurement for time and frequency service by aid of the same hardware and software.

The first step in this direction was the development and usage of low-orbital satellite VHF systems: "Tsykada" USSR and "Transit" USA. The next great step in this activity was the realization of the GPS NAVSTAR program in USA and GLONASS program in USSR.

Both programs are in pre-operational phase at the present time. There are 8 - 10 satellites on orbit in the deployment stage. It is planned to complete the full deployment of the system in the near future. The main requirements to the navigation system are the high accuracy of the location, global coverage, continuity and reliability of operation, the possibility of free access to the user. The systems GPS and GLONASS provide a more accurate and more rapid location than the previous low-orbital systems "Transit" and "Tsykada". GPS is a well-known space navigation system and there is a lot of articles in scientific and technical papers devoted to the description of the system. Parameters of the soviet system GLONASS were presented only in 1988 at the Meeting of the Special Committee on Future Air Navigation Systems (FANS) and at the Sub-Committee on Safety of Navigation IMO (35th Session).

Under the full operational capability the system will consist of 24 satellites (21 satellites in operational status and 3 in spare status), placed on near circular orbits with altitude of 19 100 km, inclination of 64.8°, period of revolution 11 hrs 15 min. The satellites are placed in three orbital planes, equally (120 degrees apart) distributed in the equator plane. The navigational fixing is based on measurements of the pseudorange

and pseudorange rate related to satellites in the user's radiovisibility zone. User's receiving equipment is operated in the passive mode. The different Gold codes are used in GPS system for satellite signal division. To resolve this problem in the GLONASS system, the satellites emit phase-manipulated signal with the nominal carrier frequency, which is dictated by the relation:

$$f_j = f_i + (j - 1)\,\Delta f \qquad \begin{aligned} j &= 1, 2, \ldots, 24 \\ f_i &= 1602.5626 \text{ MHz} \\ \Delta f &= 0.5626 \text{ MHz} \end{aligned}$$

For the phase manipulation of the navigation signal, the signal M1 is used, which changes the phase within 180° with an error not exceeding $\pm$ 12°. This signal is obtained as a result of addition of two sequencies: pseudorandom sequence PRS A and uninterrupted sequence H1. The sequence PRS A is taken from the 7th digit of the 9 digit shift register. The shift register initial state code in the PRS A generator corresponds to the presence of "1" in all register digits. The initial symbol in PRS A period is the 1st symbol of the group 111111100..., which is repeated through 1ms (511 bit). Sequence H1 is represented as uninterruptedly formed two-second lines. The first part of the line (with duration of 1.7 s) is the binary sequence, obtained as a result of addition of two binary sequencies: digital data in the relative code with the symbol duration 20 ms and meander with symbol duration 10 ms. The transmission of the binary code of time mark (duration 0.3 s) is realized in the second part of each line. The binary code of the time mark is the truncated binary pseudorandom sequence PSP B - 30 symbols length, each symbol is 10 ms. Each 2 second line always begins with symbol 0. This symbol represents the complement to the truncated pseudorandom sequence to full one (non-truncated sequence). The structure of the emitted radiosignal assures the mutual synchronization of boundaries in two-second lines, symbols of digital data, meandre symbols, symbols of pseudorandom sequences PRS A and B (pictures 1, 2). The forward fronts of the initial symbols of pseudorandom sequence PRS A coincide with the boundaries of meandre symbols and with the boundaries of digital data symbols. As the time mark the backward front of the last symbol of pseudorandom sequence PRS B is used. This time mark corresponds to the moment, separated from the day begin (Moscow time) by an even integer number of seconds according to the satellite on-board time scale. The navigational message is transmitted in the form of the encoded binary digital data with the rate of 50 symbols per second. The data are encoded by Hemming code and converted into the relative code. The structure of the navigation message is given on the picture 3. The format of navigation message consists of so-called superframes with duration of 2.5 min (picture 4). Each superframe consists of five 30 second frames, each frame contains 15 two-second lines. Symbols of digital data are numbered from right to left. Symbols with numbers 1, ..., 8 are test symbols of Hemming code. The error of synchronization of line, frame and superframe boundaries should be no more than $\pm$ 2 ms.

Each superframe contains the following information: lines 1-4 give the within the super-

frame unchanging operative data about the satellite, from which these data are transmitted. Lines 6-15 are occupied by the information about all 24 satellites of the system: about 5 satellites in frames 1-4 and about 4 satellites in the frame 5. Every two lines contain the data about one satellite. Within the superframe a special reserve is provided for possible changes in the structure of navigation message.

It is reasonable to consider the navigation message in detail: m - line number in a frame, t_k - moment of the frame start in the current day. The amount of integer hours is registered into 5 high-order digits of the word, integer minutes into medium-order ones, the amount of 30 second intervals passed from the beginning of the current minute is written into the low-order digits.

The beginning of the day in the satellite system time coincides with the beginning of the next superframe. The data about the health of a particular satellite are given in the high binary digit in position Bn. The symbol "1" in this digit gives evidence of the satellites malfunction. The routine information within the frame relates to the moment t_b of the current day (Moscow time).

The presentation discreteness of t_b is 15 min. At moments divisible by 30 or 60 min of the on-board time-scale the change of the operational data takes place. γ_n (t) is the relative distinction of the estimated frequency value of the n-th satellite navigation signal at moment t_b

$$\gamma_n (t_b) = \frac{f_n (t_b) - f_n}{f_n}$$

where $f_n (t_b)$ - predicted frequency value of navigation signal of the n-th satellite at time $t_{(b)}$.

f_n - frequency of system reference oscillator.

Value $\tau (t_b)$ is the shift of the time-scale of n-th satellite relative to the time-scale of Glonass system. The shift equals to the phase shift of pseudorandom sequence PRS A of navigation signal relative to system reference signal for the time moment t_b expressed in seconds. The values B_n, t_b and $\gamma_n (t_b)$ are transmitted by lines 1-3, by these lines the coordinates $(X_n\ Y_n\ Z_n)\ t_b$, satellite speed and acceleration vector components $(\dot{X}_n\ \dot{Y}_n\ \dot{Z}_n)\ t_b$, $(\ddot{X}_n\ \ddot{Y}_n\ \ddot{Z}_n)\ t_b$ in the Greenwich rectangular geocentric coordinate system for the time moment t_b are transmitted. The next lines contain the following data: the calendar number of the day within the 4-year period, beginning with the leap year - N^A, correction to the system time-scale relative to the time-scale, in which ephemerides and parameters of satellite synchronization are calculated - τ_0; the number of the carrier frequency of navigation signal N^A; Greenwich longitude of the ascending node - λ_n^A; the pass time of the first ascending node of satellite $t\lambda_n^A$; correction to the mean value of the orbit inclination - $\Delta\dot{c}_n^A$; correction of the mean value of draconic orbital period of satellite ΔT_n^A ;

change of rate of the satellite period $\Delta\dot{T}_n^A$; orbit eccentricity - ε_n^A for the time moment $t_{\lambda_n}^A$; rough value of the satellite time-scale shift relative to GLONASS system scale $-\tau_n^A$; satellite conventional number in the system - n^A; generalized signature of satellite health for the moment of orbit and phase almanac - C_n insertion. The last value characterizes the fitness or unfitness of the satellite for the navigation observation.

In case of satellite illness, the message about it enters into the operational information content. The digit quantity and the digital information value range are given in Table 1.

The task of the precise transmission of the reference time and frequency signals to the mobile user is connected directly with the task to enhance the accuracy and authenticity of navigation observation. It is known that satellite navigation systems such as GPS and GLONASS separately used cannot provide a guaranteed value of the geometrical factor PDOP$\leq$ 3 in certain regions of the earth even under full deployment. There are time intervals when the geometrical factor PDOP exceeds the value of 6, that leads to an appropriate increase of the observation error. The joint usage of the GPS and GLONASS systems would allow to remove the probability of availability of any zones with higher PDOP factor. Each of the systems (GPS and GLONASS) envisages in its navigation message the information about the system integrity, which can be used for the authenticity control of the navigation. But the modern methods of reception of messages about the integrity, based on the forming of the message by the ground-based control and monitoring equipment and on the transmission of these messages for certain types of users through navigation satellites do not provide the necessary velocity of these data arrival. The authenticity control can be carried out most quickly by the user when the redundant number of satellites are available in the user's radiovisibility zone (more than 4 for three-dimensional determination). The joint application of GPS and GLONASS systems and related increase of the navigation satellite number, visible by the user, gives additional advantages for autonomous authenticity control of reception of reference time and frequency values and navigation determinations. But by the joint application of the two systems the problem of their compatibility arises. It is necessary to assess the possible consequences of distinguishing between the systems GPS and GLONASS with purpose to elaborate the recommendations for the improvement of the compatibility. Tables 2, 3, 4 give the main differences between GPS and GLONASS systems that can influence the user's equipment design and the results of navigation observations. The number of satellites in the system and the structure of the space segment determine the operational zone of the user's service. The service zone meets the requirements if such number of satellites are available in the user's visibility zone, that provides navigation determination with required accuracy, reliability and confidence in any given region of the earth. The double quantity of satellites in the joint system GPS and GLONASS will provide the possibility to eliminate practically all time intervals with the lower probability to see the number of satellites, which is sufficient to assure the reliable navigation location (6 satellites are needed to be

available in the user's visibility zone). Taking into account the practical difficulty of orbital system structure changing, it is possible to neglect the out-of-phase spatial attitude of GPS and GLONASS system navigation satellites.

The frequency division of navigation satellite signals, favoured in the GLONASS system, permits to obtain a ratio of the mutual correlation function as compared to the autocorrelation function of 48 dB (in GPS system it equals correspondingly 21.6 dB) [2]. This circumstance reduces the interference level at the point of the navigation signal reception from other satellites of the same system, this interference level may augment considerably in case of user antenna radiation pattern distortions because of the re-reflection of signal from the construction elements of the object, at which the antenna is installed. Moreover, when pseudosatellites in zones of user intensive movement are used, the mutual correlation function ratio is desired to be higher than the autocorrelation function ratio, in order to increase the navigation location confidence and to transmit the difference corrections and in view of the wide dynamic range of pseudosatellite signals for different distances of the user from the pseudosatellite.

At the same time the contemporary development stage of means and methods for signal digital processing allows to realize in digital form the synthesis of heterodyne frequencies (in accordance with the signal of each satellite in the GLONASS system) as well as the synthesis of Gold codes (in accordance with the signal of each satellite in the GPS system), after transformation of the input signal from the intermediate frequency to the discrete form. Parameters of modern means of the signal digital processing make it possible to process the GPS and GLONASS signals with all their distinctions, pointed out in table 4, using the same receiver structures. The block-circuit of such receiver, realized in the equipment "SHKIPER-N" is given in Figure 5.

GPS and GLONASS system's frequency ranges are close and this makes it possible to use the combined antenna and common input LNA in the user equipment. The optimal receiver design and appropriate software for signal processing allow to create the combined receiver at a price exceeding only a little the price of the GPS receiver.

The difference in the ephemerids information and almanac representation in GPS and GLONASS systems does not create obstacles for user's equipment operation thanks to the appropriate organizing of navigation processor software, which permits to process both data flows.

One more thing to be taken into account in the joint application of GPS and GLONASS systems is the difference between the used geodetic coordinate systems WGS-84 and SGS-85 that gives rise to error of navigation observations. Parameters of the Soviet Geocentric Coordinate System SGS-85 were reported by the Soviet delegation during the Working Group FANS meeting on April 10-21, 1989.

The activity within the bounds of the international cooperation concerning the determination of the mutual correlation of both geodetic coordinate systems, would make it possible to eliminate this error component.

Navigation signals of GPS and GLONASS systems are fixed to their system time and therefore by the mutual use of satellite signals of both systems in the user equipment, the error of the pseudorange difference measurement may appear.

$$\Delta R = c \cdot \tau$$

where τ is the system time discrepancy between GPS and GLONASS systems. The GPS system time is fixed to UTC (USNO), while the GLONASS system time is fixed to UTC (84) with error +1 mcs. The error of time coordination between UTC (SU) and UTC (USNO) can reach tens of microseconds. It should be mentioned that the error in navigation observation, caused by the pseudorange measurement error, would arise only in case (in accordance with the formula given above) of the location being carried out in a constellation including both GPS and GLONASS satellites. While locating separately on GPS and GLONASS satellite constellations, the error will appear when correlating navigation observation results with the time used in navigation complex of the mobile user. This error is proportional to the user's velocity and calculated according to the formula:

$$\Delta X = V_X \cdot \Delta\tau$$

where ΔX - navigation observation error component at any coordinate;
V_X - user's velocity component at the corresponding coordinate;
$\Delta\tau$ - error of UTC (USNO) or UTC (SU), where time-scale correlation relative to user time-scales are computed.

It is obvious that the navigation observation error is not of considerable significance.

To realize the operation with the "mixed" constellation, involving the GPS and GLONASS satellites, it is necessary to determine and to store into satellite memory of both systems the precise difference between the GPS and GLONASS time-scales for the subsequent transmission within the navigation message. This can be provided by creating an equipment complex capable to estimate the system time-scale difference, based on the reception of satellite navigation signals from two systems, to transmit the calculated values through communication channels to the national control and monitoring subsystems to supply the satellite on-board equipment with the navigation information.

The principles of satellite navigation system's construction, their structure and software organization permit, during their exploitation, to optimize their performance characteristics to the user's interest. These characteristics are accuracy, reliability

and confidence in the navigation determination at any required time, simplicity of the exploitation and compatibility with other users' equipment units.

Main sources of navigation observation errors are the following:
- ephemerides error
- frequency-time correction error to phases of range measuring satellite signals
- error due to troposphere refraction
- instrument inaccuracy error during the user motion relative to satellites (pseudorange and pseudovelocity) measurement
- measurement error due to multipath signal reception onto the user's non-directional antenna.

During satellite navigation systems operation the data are accumulated about the influence of earth gravitation anomalies, the influence of sun's and moon's attraction and other factors, which are hardly to be taken into account, such as possible emissions of materials from the satellite surface in gaseous phase etc during satellite movement. The processing and subsequent inclusion of these data for ephemeride information prediction will allow to reduce considerably the error due to ephemerides knowledge uncertainty. To take into consideration the gravitation potential anomalies more precise when predicting navigation satellite ephemerides and to determine more precise geodynamical parameters (earth rotation non-uniformity, polar motion), the quantum-optical measurements on the "Etalon" satellite, launched into GLONASS satellite orbit, may be applied. Great advantages in this activity would give the international cooperation in the joint processing of laser range measurement on satellites "Etalon", "Lageos" and "Starlet".

The inaccuracy of frequency-time corrections to signal phases, designed to be used in pseudorange measurements depends to a greater extent on the instability of the satellite on-board frequency standard. The frequency standards used at present on GLONASS satellites have instabilities of $3\text{-}5 \cdot 10^{-13}$. The conduction of further research and development directed to the creation of highly stable on-board oscillators, will permit to reduce this error component by an order of magnitude.

The ionospheric refraction error depends on season, day time, solar activity and changes within a wide range. The transmission of corrections for ionosphere state within the navigation message decreases the corresponding error component approximately by 50%. Establishing international cooperation and data exchange on ionospheric refraction would give the possibility to reduce the error.

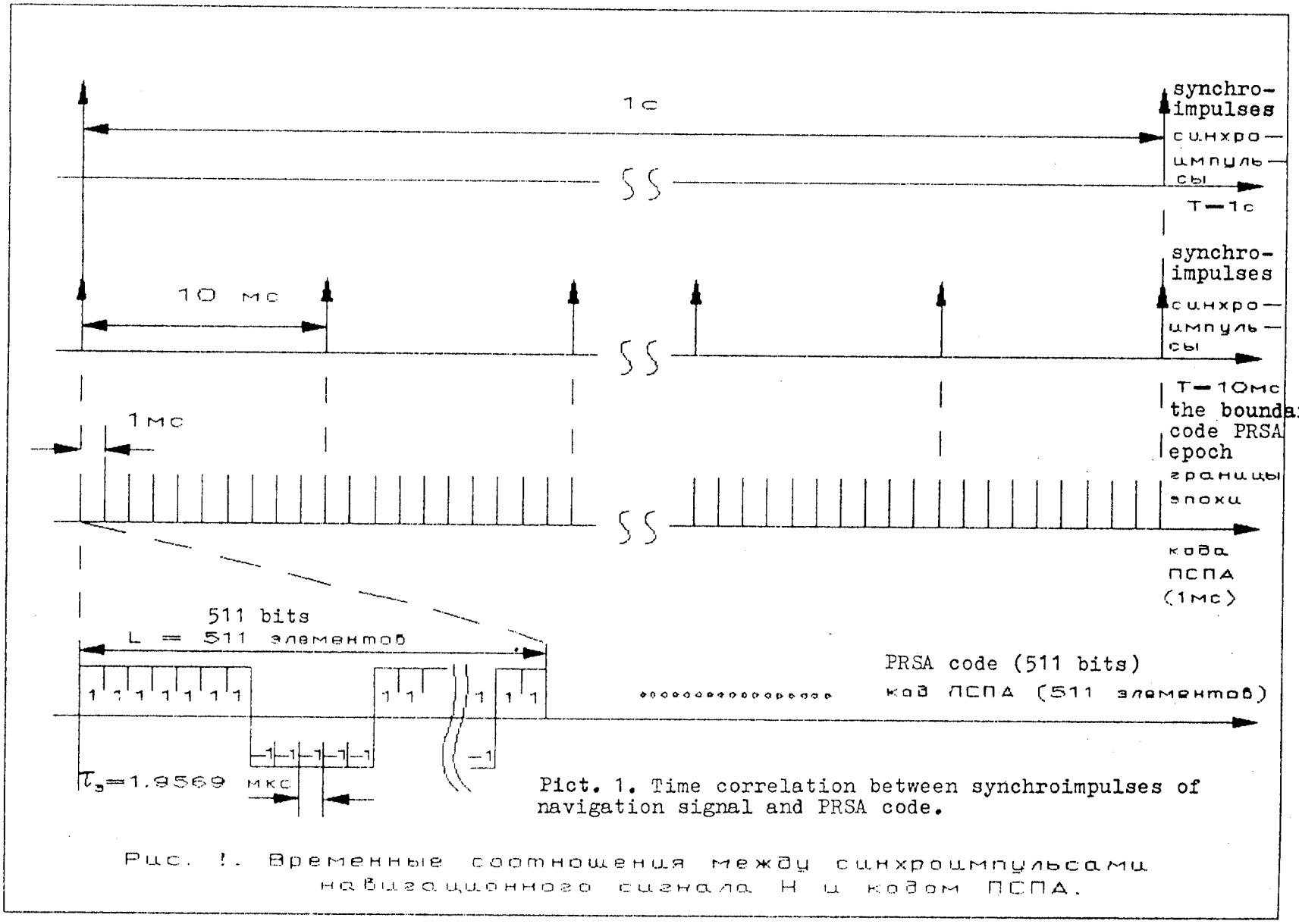
The constituent of the error due to tropospheric refraction can be taken into account to great extent based on the model of the current troposphere state. At low elevation angles the error components not taken into consideration could be great enough. Certain progress in reducing this error components can be reached by developing the multichannel version

of the user's equipment which can provide tracking of all satellite types in the visibility zone. The minimal required satellite number may be considerably increased provided the joint application of GPS and GLONASS systems becomes reality.

The common processing of measurements on all visible satellites with simultaneous inclusion of troposphere model will help to decrease this error constituent.

The user's equipment is the most flexible part of the navigation means in satellite navigation systems, the part that will be permanently improved and modified. At present most of realized equipment types are multichannel and capable to carry out the permanent tracking of pseudorange and pseudovelocity of several visible satellites, aiming at receiving maximum accuracy and authenticity of navigation observation. Progress in the development of means for digital signal processing, growth of their high rate and integration degree will permit to convert the signals into discrete form within the structure of the receiving unit close to its input.

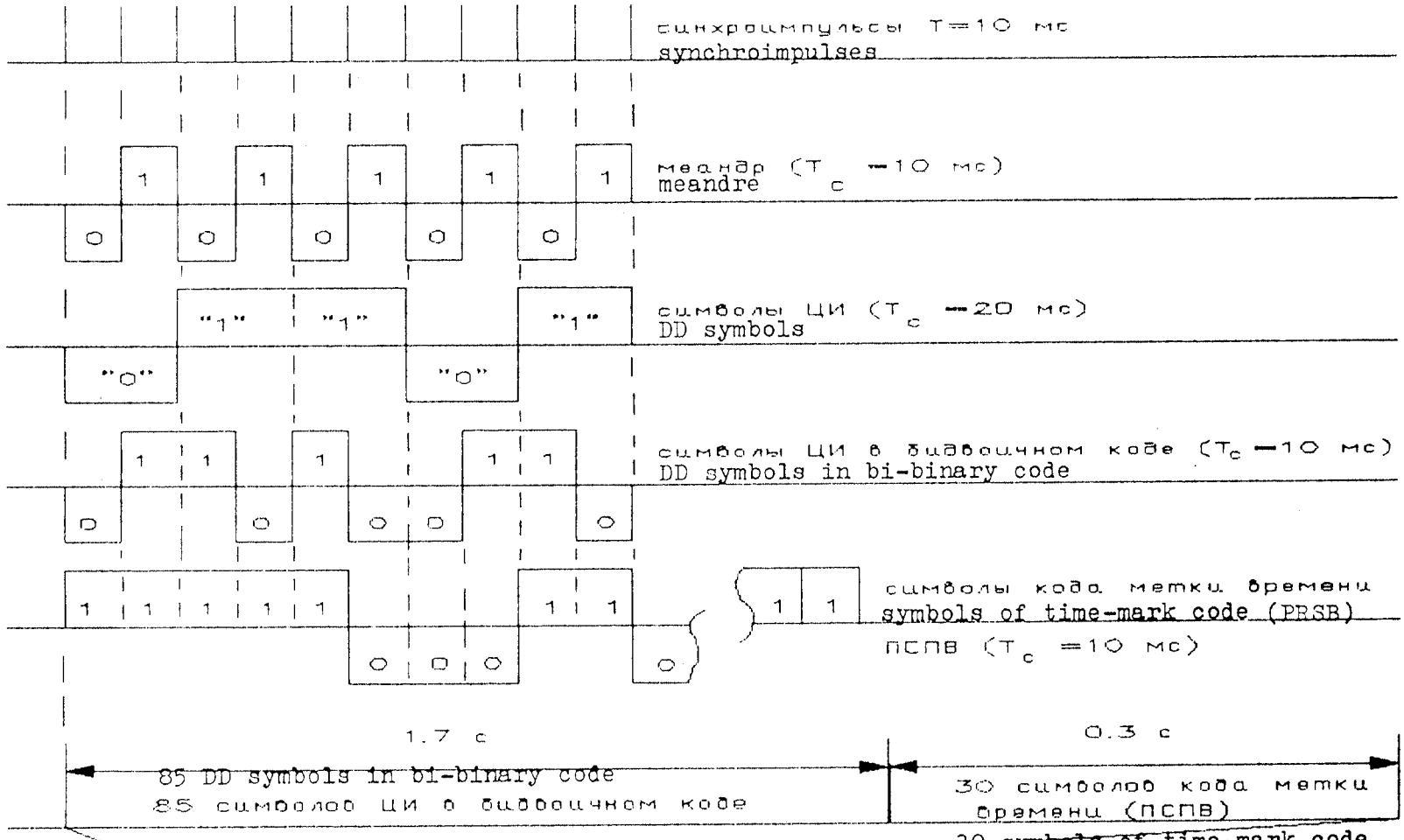
Multilevel quantization at high frequency combined with the development of high order filters, based on new achievements of electronic technology and methods of mathematic signals processing with maximum implementation of possibilities of its compatibility with other navigation measuring means, creates foundation for the progress in the direction of further reducing of measurement errors and improving design and performance characteristics of user's equipment.



Pict. 1. Time correlation between synchroimpulses of navigation signal and PRSA code.

Рис. 1. Временные соотношения между синхроимпульсами навигационного сигнала Н и кодом ПСПА.

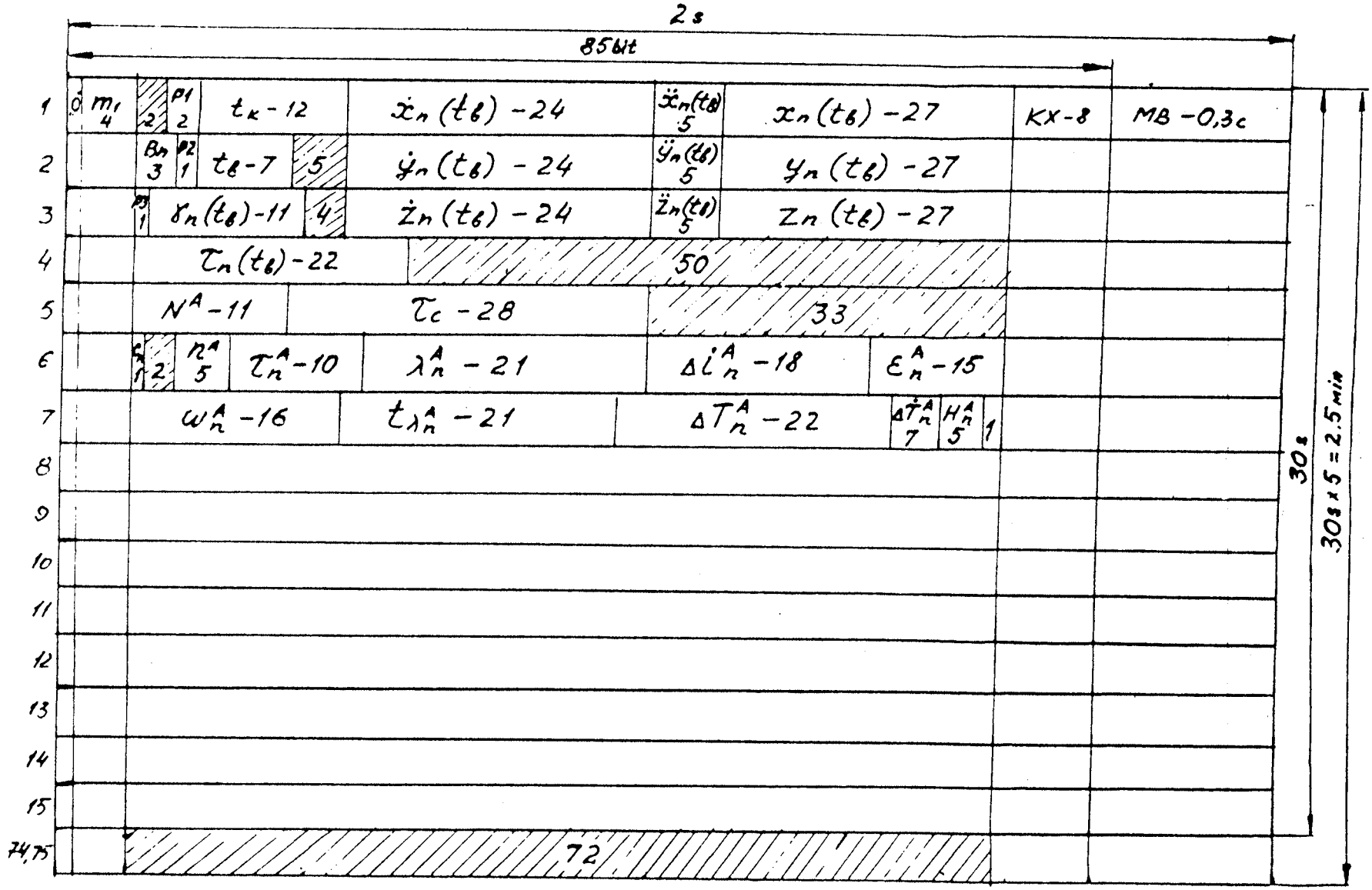
Pict. 1



Pict .2 Scheme of normalization of navigation signal in the satellite processor.

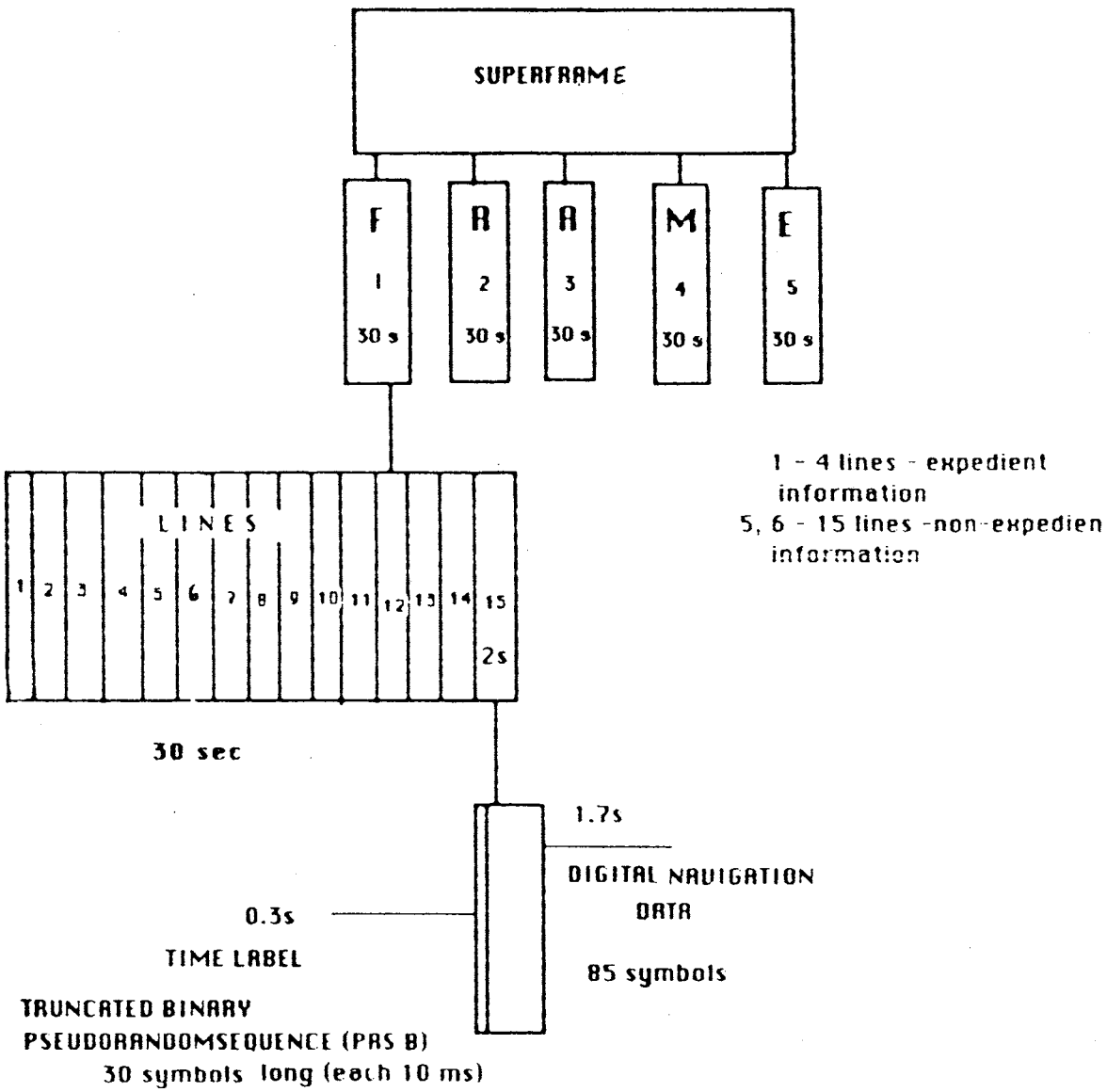
Рис.2. Схема формирования навигационного сигнала в процессоре спутника

Pict. 2

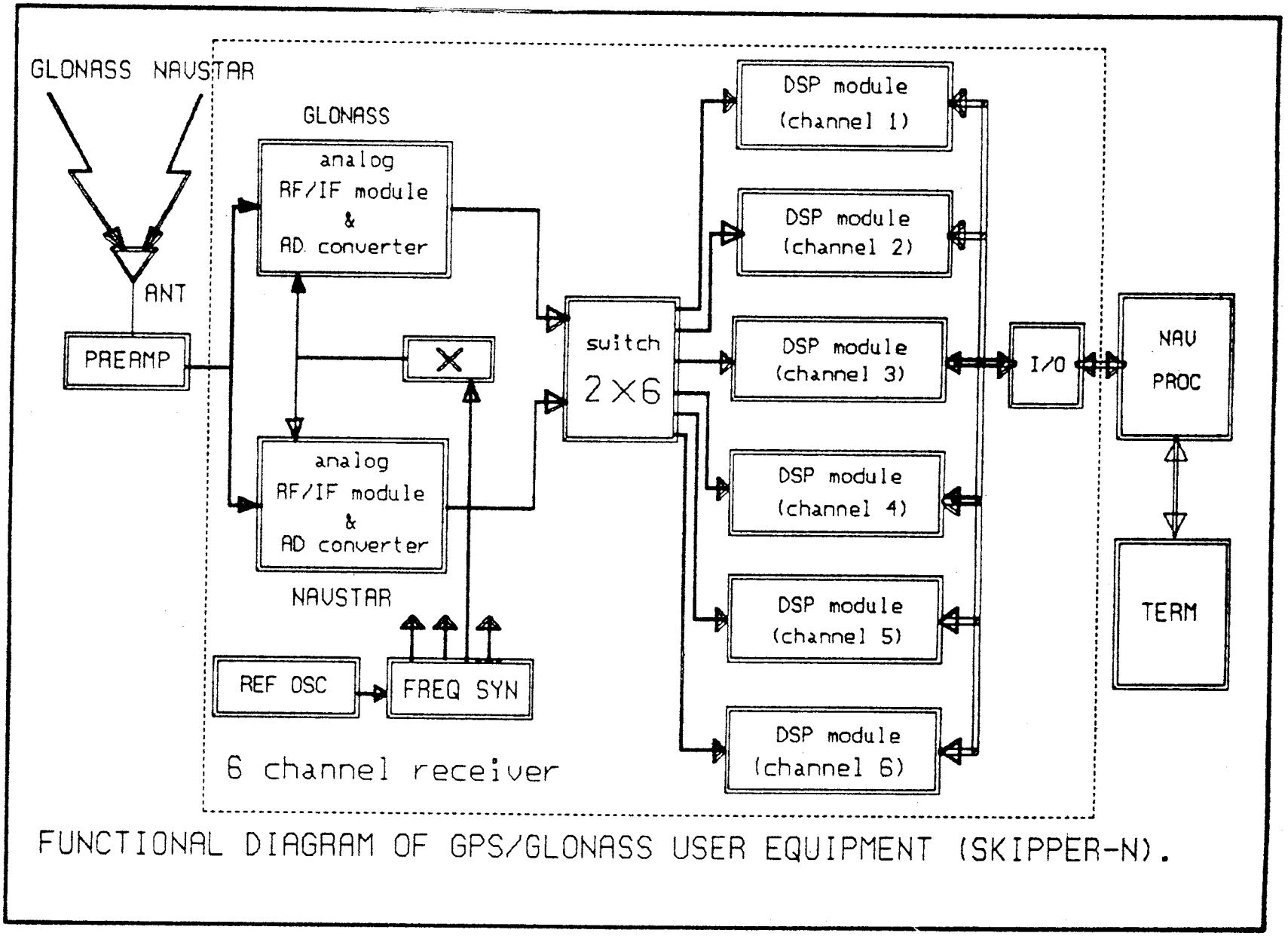

2s
85бит
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
74,75
m_1 4
2
P1 2
$t_к$ - 12
$\dot{x}_n(t_в)$ -24
$\ddot{x}_n(t_в)$ 5
$x_n(t_в)$ -27
KX-8
MB -0,3c
Bn 3
P2 1
$t_в$ -7
5
$\dot{y}_n(t_в)$ - 24
$\ddot{y}_n(t_в)$ 5
$y_n(t_в)$ -27
P3 1
$\gamma_n(t_в)$-11
4
$\dot{z}_n(t_в)$ - 24
$\ddot{z}_n(t_в)$ 5
$z_n(t_в)$ - 27
$\tau_n(t_в)$-22
50
N^A-11
τ_c -28
33
c_n^A 1
2
n_n^A 5
τ_n^A-10
λ_n^A - 21
Δi_n^A -18
ε_n^A -15
ω_n^A -16
$t_{\lambda n}^A$ -21
ΔT_n^A -22
$\Delta\dot{T}_n^A$ 7
H_n^A 5
1
72
30s
30s x 5 = 2.5 min

Pict.3

NAVIGATION MESSAGE STRUKTURE



Pict.4



FUNCTIONAL DIAGRAM OF GPS/GLONASS USER EQUIPMENT (SKIPPER-N).

Pict.5

Word capacity and numerical range in the digital data within the navigation signal structure of the "GLONASS" satellite system

Word designation	Unit of measurement	Nominal range of values	Number of binary digits	Value of low order digit(scale coefficient)	Acceptable range values
1	2	3	4	5	6
m_1	1	1 16	4	1	0 15
t_k	hour	0 23	5	1	0 31
	minute	0 59	6	1	0 63
	second	0; 30	1	30	0; 30
t_b	minute	15; 30;.... 1425	7	15	0; 15 1905
$\gamma_n(t_b)$	1	$\pm 0{,}9\times10^{-9}$	11	$2^{-40} \simeq 10^{-12}$	$\pm 2^{-30} \simeq \pm 0{,}9\times10^{-9}$
$\tau_n(t_b)$	second	$\pm 1{,}9\times10^{-3}$	22	$2^{-30} \simeq 10^{-9}$	$\pm 2^{-9} \simeq \pm 1{,}9\times10^{-3}$
$x_n(t_b), y_n(t_b), z_n(t_b)$	kilometres	$\pm 2{,}7\times10^{4}$	27	$2^{-11} \simeq 5.0\times10^{-4}$	$\pm 2^{15} \simeq \pm 3\times10^{4}$
$\dot{x}_n(t_b), \dot{y}_n(t_b), \dot{z}_n(t_b)$	kilometres/s	$\pm 4{,}3$	24	$2^{-20} \simeq 10^{-6}$	$\pm 2^{3} \simeq \pm 8$
$\ddot{x}_n(t_b), \ddot{y}_n(t_b), \ddot{z}_n(t_b)$	kilometres/s^2	$\pm 6{,}2\times10^{-9}$	5	$2^{-30} \simeq 10^{-9}$	$\pm 2^{-26} \simeq \pm 14{,}8\times10^{-9}$
τ_c	second	± 1	28	$2^{-27} \simeq 8\times10^{-9}$	± 1
N^A	day	1 1464	11	1	0 2047
n^A, H_n^A	1	1....24	5	1	0....31
λ^A_n	half-cycle (180°)	± 1	21	$2^{20} \simeq 10^{-6}$	± 1

Table 1.

Table 1 (continued)

1	2	3	4	5	6
$t\lambda_n^A$	second	0 ....44100 (0....735 min.)	21	$2^{-5} \simeq 0,03$	0....65535
Δi_n^A	half-cycle (180°)	$\pm$ 0,067($\pm$ 12°)	18	$2^{-20} \simeq 10^{-6}$	$\pm 2^{-3} \simeq \pm 0,125$
ΔT_n^A		$\pm$ 3,6x10^3 ($\pm$ 1 hour)	22	$2^{-9} \simeq 2x10^{-3}$	$\pm 2^{12} \simeq \pm 4 \times 10^3$
$\Delta \dot{T}_n^A$		$\pm 2^{-8}$	7	$2^{-14} \simeq 6,1x10^{-5}$	$\pm 2^{-8} \simeq 3,9x10^{-3}$
ε_n^A	1	0....0,03	15	$2^{-20} \simeq 10^{-6}$	0.... 0,031
ω_n^A	half-cycle (180°)	$\pm$ 1	16	$2^{-15} \simeq 3,2x10^{-5}$	$\pm$ 1
τ_n^A	second	$\pm$ 1,9x10^{-3}	10	$2^{-18} \simeq 4x10^{-6}$	$\pm 2^{-9} \simeq \pm 1,9 \times 10^{-3}$

Table 2.

Parameter	GLONASS	GPS
1. Ephemeride information presentation method	9 parameters of s/c motion in the geocentric rectangular rotated coordinate system	Helper elements of satellite orbits
2. Geodesic coordinate system	СГС — 85	WGS — 84
3. Referencing of the ranging signal phases	To the timer of GLONASS system	To the timer of GPS system
4. System time corrections relative to the Universal coordinated time	UTC [SU]	UTC [USNO]
5. Duration of the almanac transmission	2.5 min	12.5 min

Table 3.

Parameter	GLONASS	GPS
Number of satellites in the full operational system	21 + 3	21 + 3
Number of orbital planes	3	6
Inclination	64.8^{o}	55^{o}
Orbit altitude	19100 km	20180 km
Orbital period	11 h 15 min	12 h

Table 4.

Parameter	GLONASS	GPS
1. Satellites signal division method	frequency division	code division
2. Frequency band allocated	[1602.5625—1615.5] ± 0.5 MHz	1575.42±1 MHz
3. Type of the ranging code in application	PN — sequence of maximal length	Gould code
4. Number of code elements	511	1023
5. Timing frequency of code	0.511 MHz	1.023 MHz
6. Crosstalk level in two neighbouring channels	— 48 dB	— 21.6 dB
7. Synhrocode repitition period	2 s	6 s
8. Symbol number in the synchrocode	30	8

QUESTIONS AND ANSWERS

The following questions were raised in written form from the participants listed below:

QUESTIONS

1 What type of space-clock, i.e. Rb, Cs, H maser, is now in use in GLONASS?

2 What is planned for the future ?

3 How many clocks are on board of each satellitte ?

4 You mentioned instability of 3 to 5×10^{-13} for the frequency standard on board of the satellite. Is this accuracy or stability ? What is the sampling time ?

5 Is there any published description of the SGS-85 position coordinate system ?

6 Is the "health" message giving some information concerning $\Delta t = T_{SAT} - UTC(SU)$?

7 Where is the information UTC(SU) - UTC(BIH) available ? The BIH bulletin is arriving 3-4 months after the fact.

8 Is the difference $\Delta t = T_{SAT} - UTC(SU)$ kept below some given level ? If yes, what is this level ?

9 Is the clear-access code at 512 kHz available ?

10 What is the relation between GLONASS time, Moscow time and UTC(SU) ?

11 What is the relation between SGS-85 coordinate system and WGS-84 coordinate system ?

12 How many monitoring stations has GLONASS the system, and where are they located ?

13 When will the GLONASS system be fully operational ?

14 How good are broadcast ephemerides of GLONASS satellites ?

15 Is someone computing precise ephemerides of GLONASS satellites ?

16 Is someone manufacturing civil GLONASS receivers (navigation, geodesy, time) in Soviet Union ?

17 Does exist in Soviet Union a civil agency working on civil applications of GLONASS system ?

18 Will selective availability be applied to GLONASS satellites, or not ?

PARTICIPANTS

Dr. R. F.C. Vessot, Harvard Smithsonian Center for Astrophysics, Cambridge/USA

Dr. H. Hellwig, National Institute of Standards and Technology, Gaithesburg/USA

Dr. D.J.E. Knight, National Physical Laboratory, Teddington/UK

Prof. S. Leschiutta, Politecnico di Torino, Dipartimento di Elettronica, Torino/Italy

Mr. W. Leandowski, Bureau International des Poids et Mesures, Sèvres/France

Dr. G. Busca, Observatoire Cantonal, Neuchâtel/Switzerland

ANSWERS

Concerning the first 3 questions, Mr. Yuriy I. Lavkovskiy, vice-president of "MORSVJAZSPUTNIK" answered directly by telephone from Mowcow to the scientific program chairman G. Busca:

1 Cs atomic clock are presently used in GLONASS satellites.

2 For the future H masers are planned.

3 There are between 2 to 3 Cs clocks per satellite.

The other questions were answered in detail, few days after the conference, by a Fax of Mr. Lavkovskiy, as reported below.

Moscow time corresponds to UTC(SU) time. GLONASS system time is established by the Control and Management Subsystem (CMS) of the GLONASS system and is tied to UTC(S) time-scale. Time-scale of GLONASS system is the time-scale of central system's synchronizer which is a part of CMS. CMS of GLONASS system is designed to control the system time-scale in such a way that it must not differ from UTC(SU) time-scale more than a value of approximately 100 ns (root-mean-square value).

Each satellite's time-scale is periodically referred to time-scale of GLONASS system. Navigation message contains the necessary information n(tb) and c which allows the user to tie his own time-scale to UTC(SU) time-scale. The information is precise enough to do this tying of user's time-scale with a max. error + 1 μs. In GLONASS satellites there are caesium oscillators used which is reserved to provide reliability.

Control stations are placed in the USSR area. The number of control stations changes as the system is being deployed to its full structure. The system will be deployed fully by 1995.

Root-mean-square error of satellite's ephemerides does not exceed:

	according its position	according its speed
along the orbit	20 m	0,05 sm/s
in transversal direction to orbit's plane	10 m	0,1 sm/s
according to height	5 m	0,3 sm/s

Precise ephemerides of GLONASS satellites are calculated by the Control and Management Subsystem and is fed into the on-board satellite's memory.

At present time in the USSR special works are conducted to create civil receivers for different users under authority of the Ministries of Merchant Marine, Civil Aviation and others.

Questions concerning the usage of the GLONASS system are considered by the Ministries to which authority objects-user's of navigation signals belong. Glavcosmos of the USSR carries out the coordination of all questions concerning the usage.

Actions aimed at deterioration accuracy of GLONASS system are neither conducted and nor planned.

Data concerning coordinate system SGS-85 and proposals about carrying out works concerning determination of mutual correlation of SGS-85 and WGS-84 system were presented by the Soviet delegation at the FANS Committee Working Group in April 1989 in Paris.

The value of 3 to 5×10^{-13} mentioned in the report relates to instability of the satellite oscillator determined as relative random variation of the frequency during 24-hours.

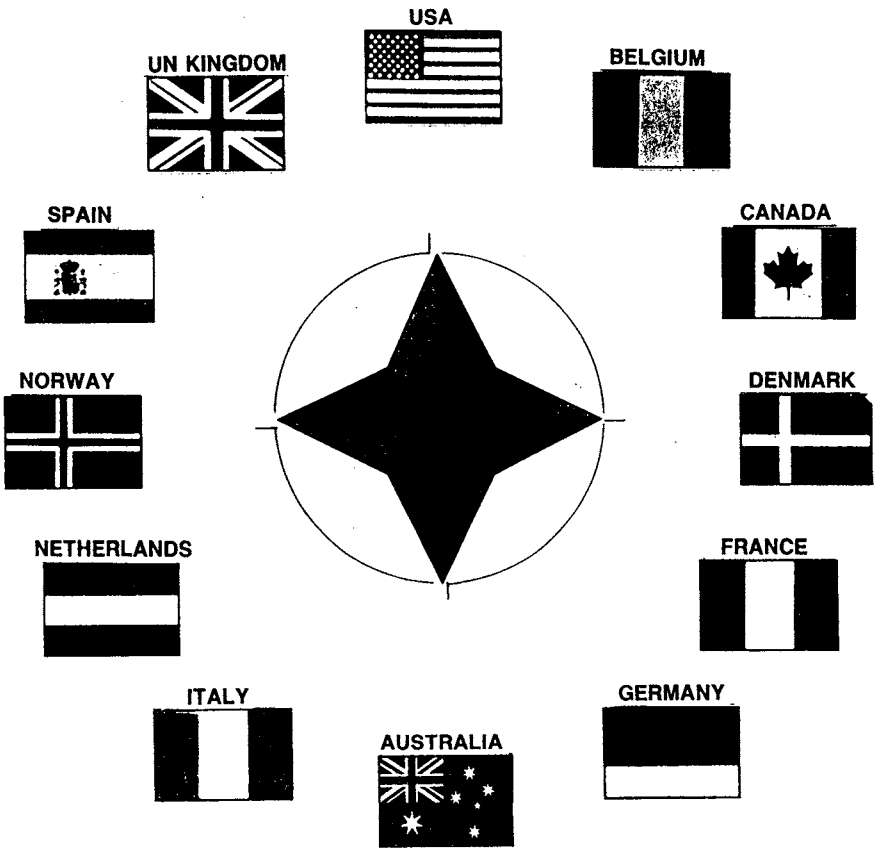
Information about the deviation of coordinated time-scales of foreign countries from UTC(SU) scale, including UTC(BIH) is contained in bulletin of State Committee on Common Time and Standard Frequencies of USSR Gosstandard.

Session APP-01 Number 2

NAVSTAR
GLOBAL POSITIONING SYSTEM
JOINT PROGRAM OFFICE

MAJOR OYVIND BERGENE, NoA
NORWEGIAN REPRESENTATIVE

USAF SPACE SYSTEMS DIVISION
LOS ANGELES, CA



NAVTAR GPS
JOINT PROGRAM OFFICE

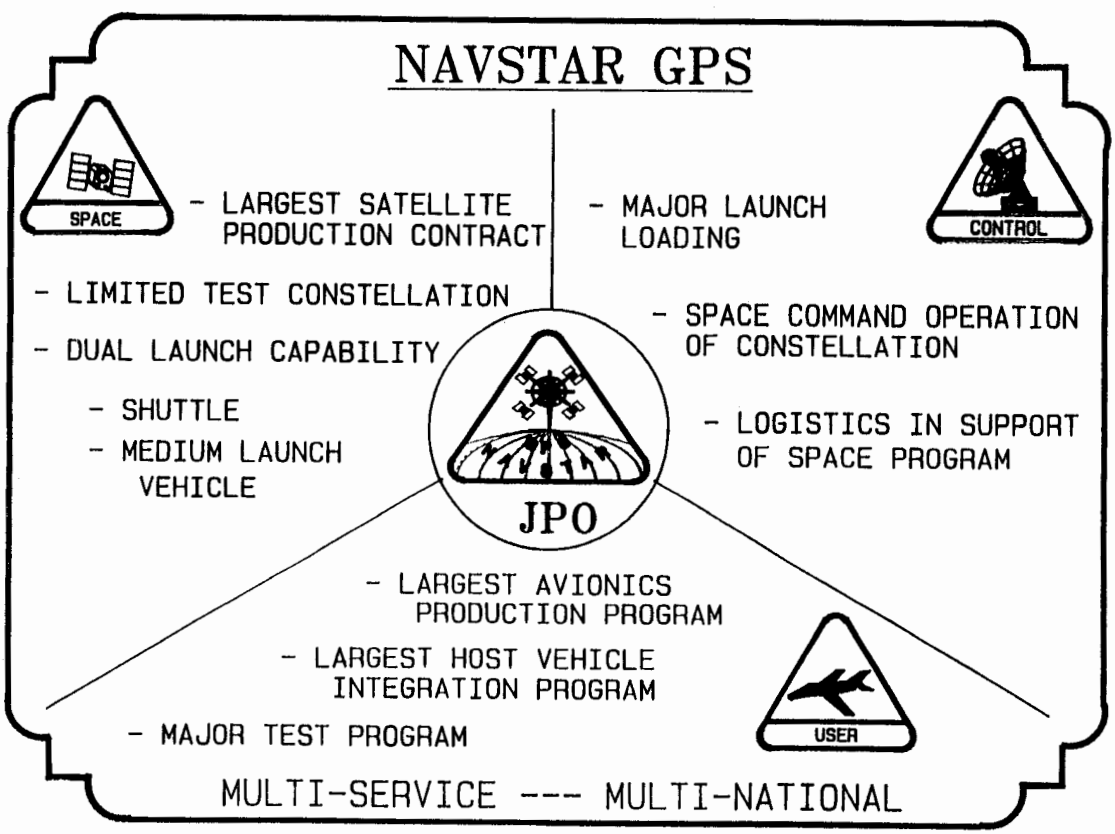
FUNCTIONS

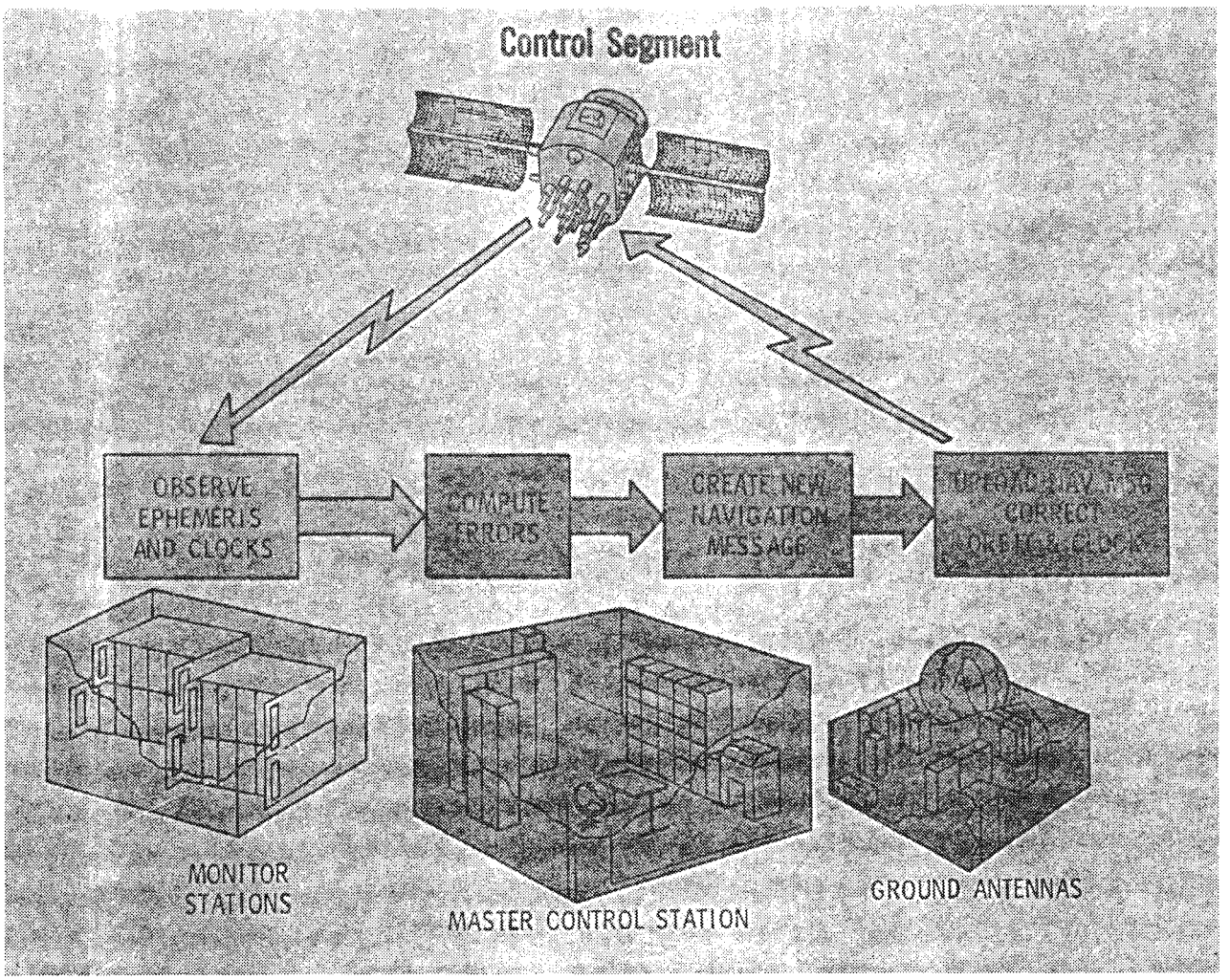
- o DESIGN/DEVELOP SYSTEM HARDWARE
- o TEST SYSTEM HARDWARE
- o BUY INITIAL PRODUCTION HARDWARE
- o TURNOVER HARDWARE TO OPERATIONAL FORCES AS SOON AS POSSIBLE

NAVTAR GPS
JOINT PROGRAM OFFICE

DOES NOT DO

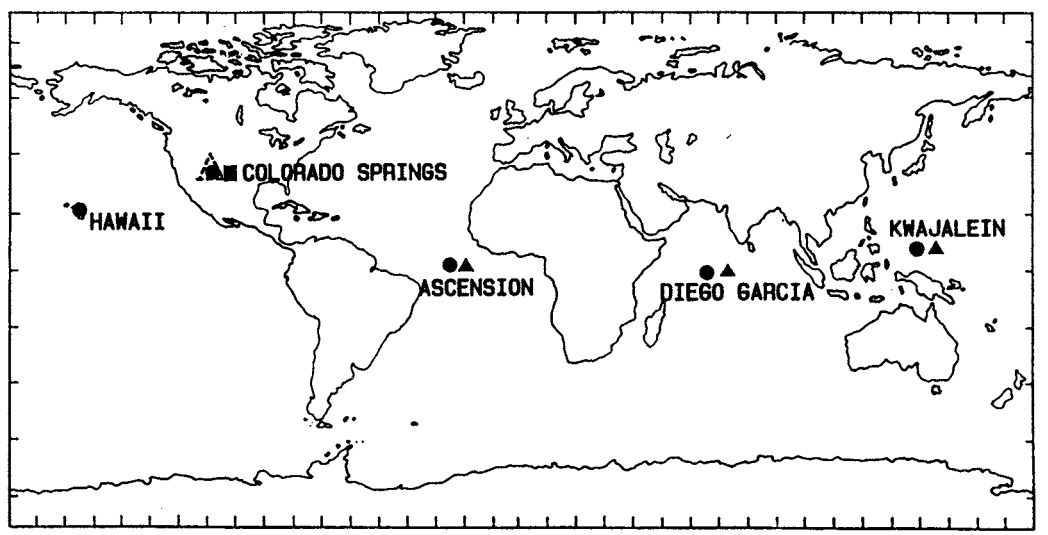
- o OPERATIONAL REQUIREMENTS
  - USING COMMANDS
- o NATIONAL POLICY
  - HIGHER HEADQUARTERS





OPERATIONAL CONTROL SYSTEM

OPERATIONS FACILITIES

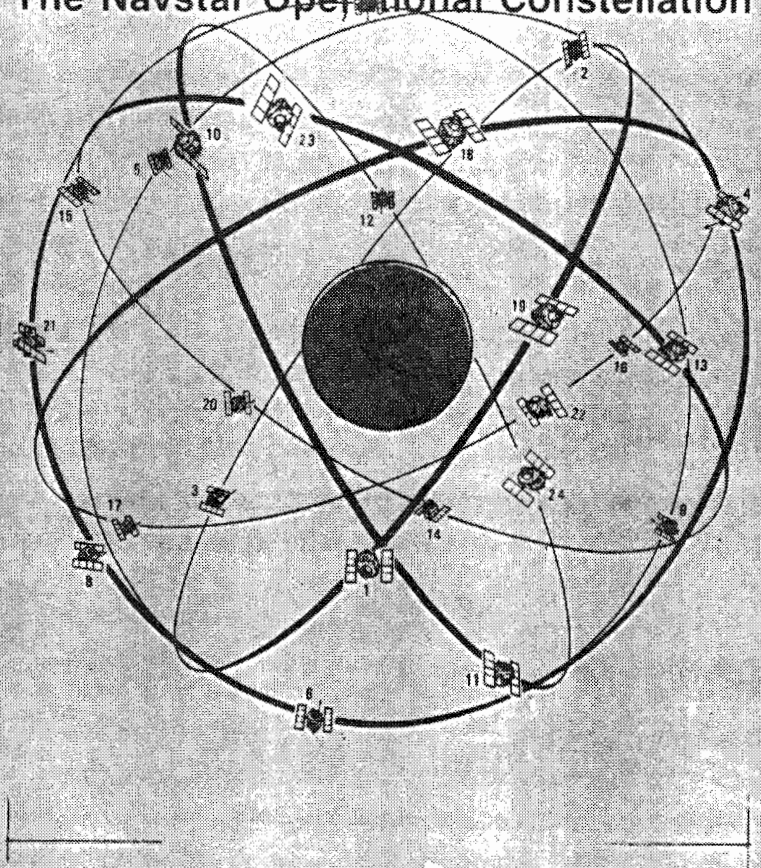


■ MASTER CONTROL STATION
- NAVIGATION ESTIMATION
- SATELLITE CONTROL
- SYSTEMS OPERATIONS

● MONITOR STATION
- COLLECT RANGE DATA
- MONITOR NAVIGATION SERVICES

▲ GROUND ANTENNA
- TRANSMIT
  -- NAVIGATION DATA
  -- COMMANDS
- COLLECT TELEMETRY

The Navstar Operational Constellation



GPS PROVIDES TWO BASIC SERVICES:

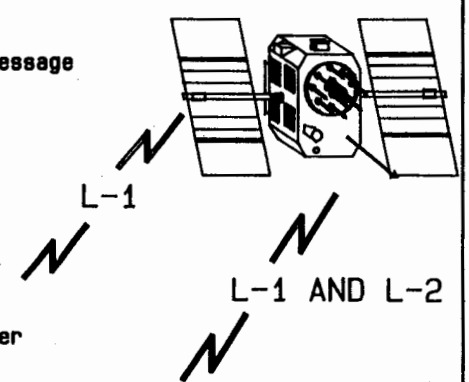
Precise Positioning Service (PPS)

Standard Positioning Service (SPS)

DATA TRANSMITTED ON TWO FREQUENCIES

Frequency L-1: C/A Code, P Code, Navigation Data Message

Frequency L-2: P Code, Navigation Data Message



SINGLE FREQUENCY RECEIVER DESIGN

Crypto Capable = PPS (30 meter) Receiver

Not Crypto Capable = SPS (100 meter) Receiver

MULTIPLE FREQUENCY RECEIVER DESIGN

Crypto Capable = PPS (16 meter) Receiver

Not Crypto Capable = SPS (100 meter) Receiver

POSITIONING/NAVIGATION CAPABILITIES USING VARIOUS RECEIVER DESIGNS

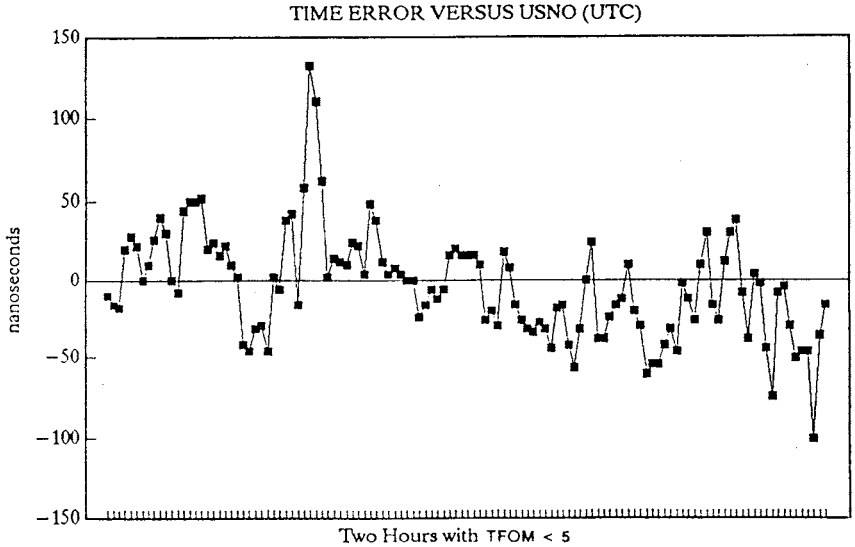
ACCURACY

POSITION	VELOCITY	TIME
6 m	0.1 m/sec	100 nsec

- 10 TIMES MORE ACCURATE THAN ANY OTHER RADIO NAVIGATION SYSTEM
- 4 TIMES MORE ACCURATE THAN CURRENT WEAPONS DELIVERY SYSTEMS

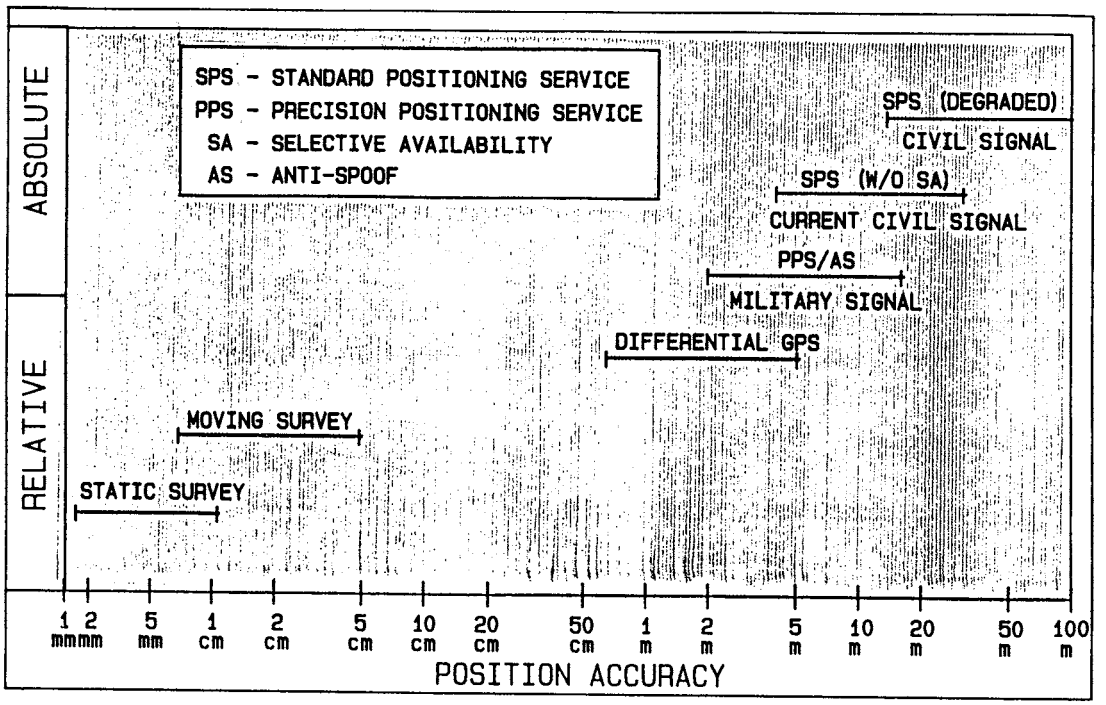
GPS 3A REC

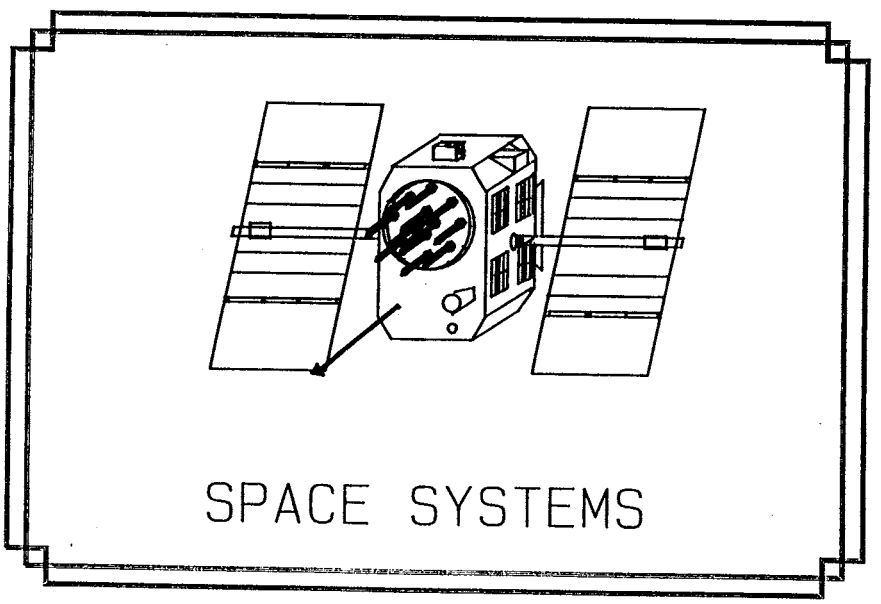
TIME ERROR VERSUS USNO (UTC)



Average error −3.7 nsec
One Standard Deviation 34 nsec

LEVELS OF GPS ACCURACY

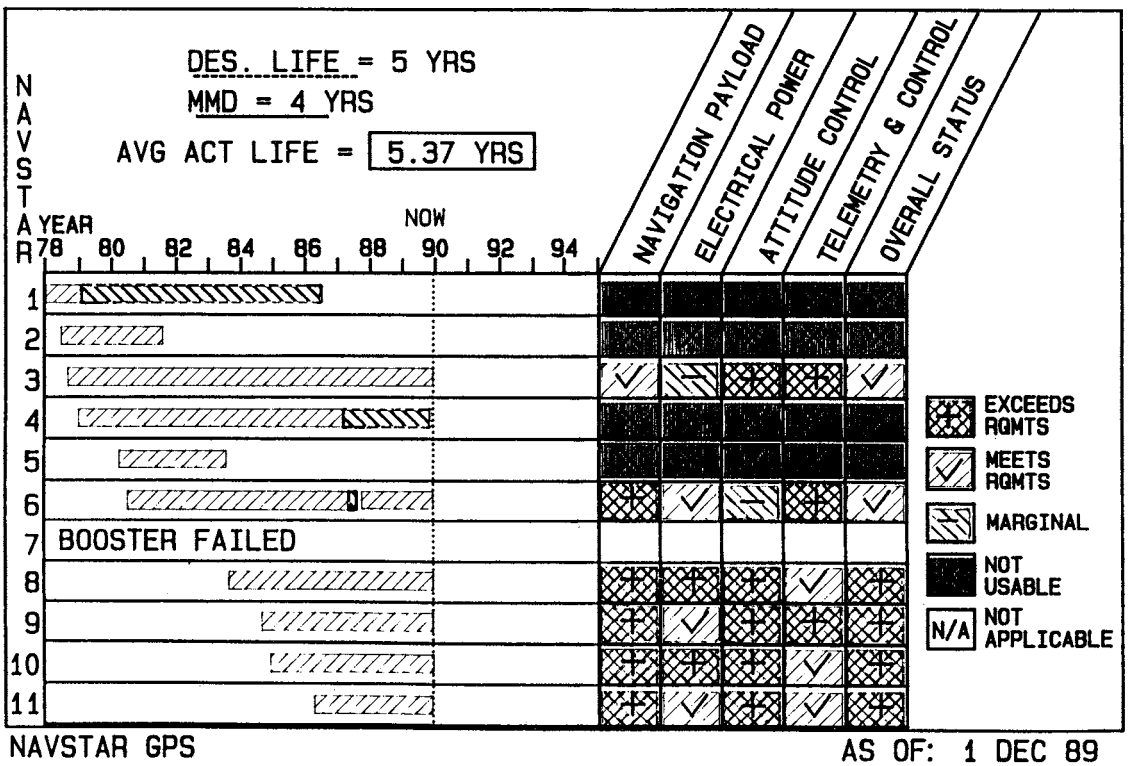




GPS SPACECRAFT

- o BLOCK I - R & D SPACECRAFT
- o BLOCK II - OPERATIONAL SPACECRAFT
- o BLOCK IIR - NEXT BLOCK CHANGE
- o BLOCK III - NEXT CENTURY

NAVSTAR
BLOCK I CONSTELLATION STATUS



BLOCK I ON-ORBIT HISTORY

NAVSTAR	LAUNCH DATE	FAILURE DATE	OPERATIONAL MONTH (1)
1	2-22-78	1-25-80 (ATOMIC CLOCK)	23
2	5-13 78	8-30-80 (ATOMIC CLOCK)	27
3	10-07-78	STILL OPERATING MARGINALLY	128
4	12-11-78	9-06-86 (ATOMIC CLOCK)	93
5	2-09-80	5-11-84 (REACT WHEELS)	51
6	4-26-80	STILL OPERATING MARGINALLY (WHEELS)	109
7	12-18-81	BOOSTER FAILURE	0
8	7-14-83	STILL OPERATING	67
9	6-13-84	STILL OPERATING	60
10	9-08-84	STILL OPERATING	57
11	10-08-85	STILL OPERATING	44
			663

TOTAL AVG OPERATING LIFE = $\frac{663}{11}$ = 60.27 MO = 5.02 YRS

(1) AS OF 31 MAY 89

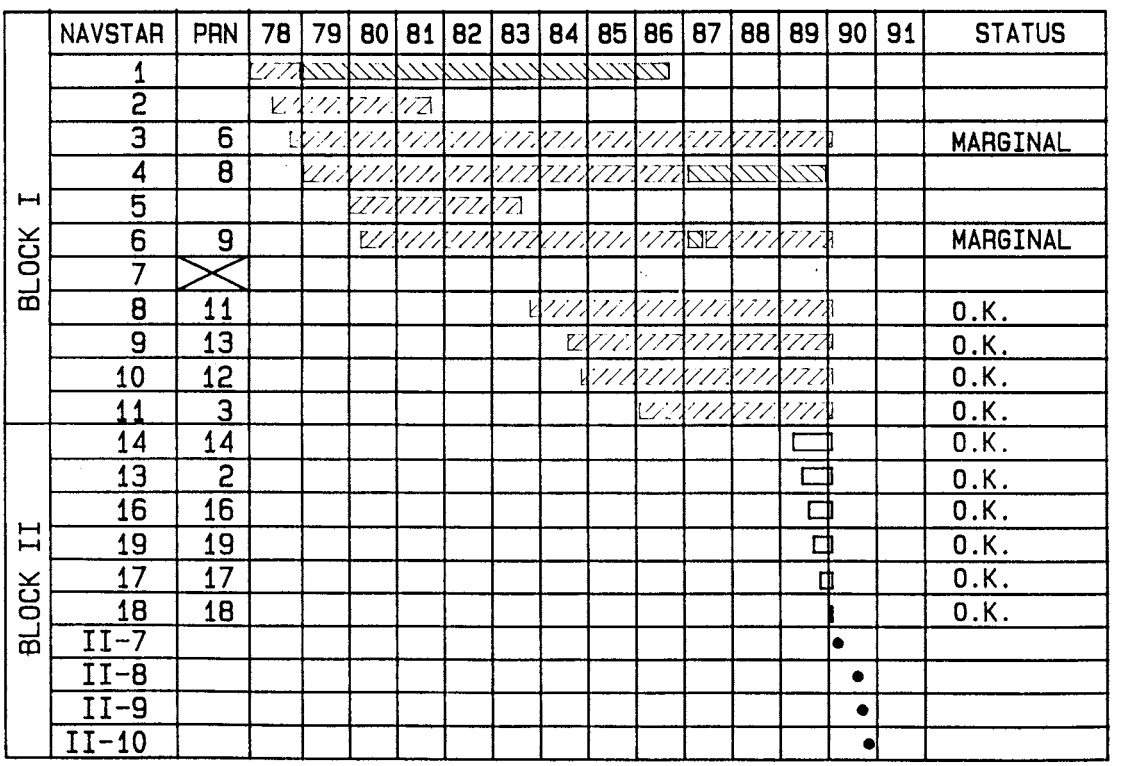
PROGNOSIS FOR REMAINING BLOCK I LIFE

- o NAVSTAR 3 POWER SYSTEM HAS REACHED END OF LIFE
  - WILL REQUIRE NAV PAYLOAD TURN OFF DURING ECLIPSE
  - IN AN OPERATIONAL ENVIRONMENT, WOULD REPLACE
- o NAVSTAR 6 HAS ONE FAILED REACTION WHEEL AND TWO OTHER SICK WHEELS
  - COULD FAIL ANOTHER WHEEL ANYTIME
- o NAVSTARS 8, 9, 10 AND 11 ARE SINGLE STRING IN ONE OR MORE SUBSYSTEMS
  - ANOTHER FAILURE IN THOSE SUBSYSTEMS LOSES SATELLITE
- o ESTIMATE OF FINAL MEAN MISSION DURATION (MMD)
  - BEST POSSIBLE - ASSUME NAVSTARS 6, 8, 9, 10 AND 11 ALL LIVE TO POWER SYSTEM FAILURE OR 10 YEARS
    - -- MMD = 6.96 YEARS
  - DESIGN - ASSUME NAVSTAR 6 LIVES 9 YRS, NAVSTARS 8, 9, 10 AND 11 LIVE 7.5 YRS (BLOCK II DESIGN LIFE)
    - -- MMD = 5.97 YEARS

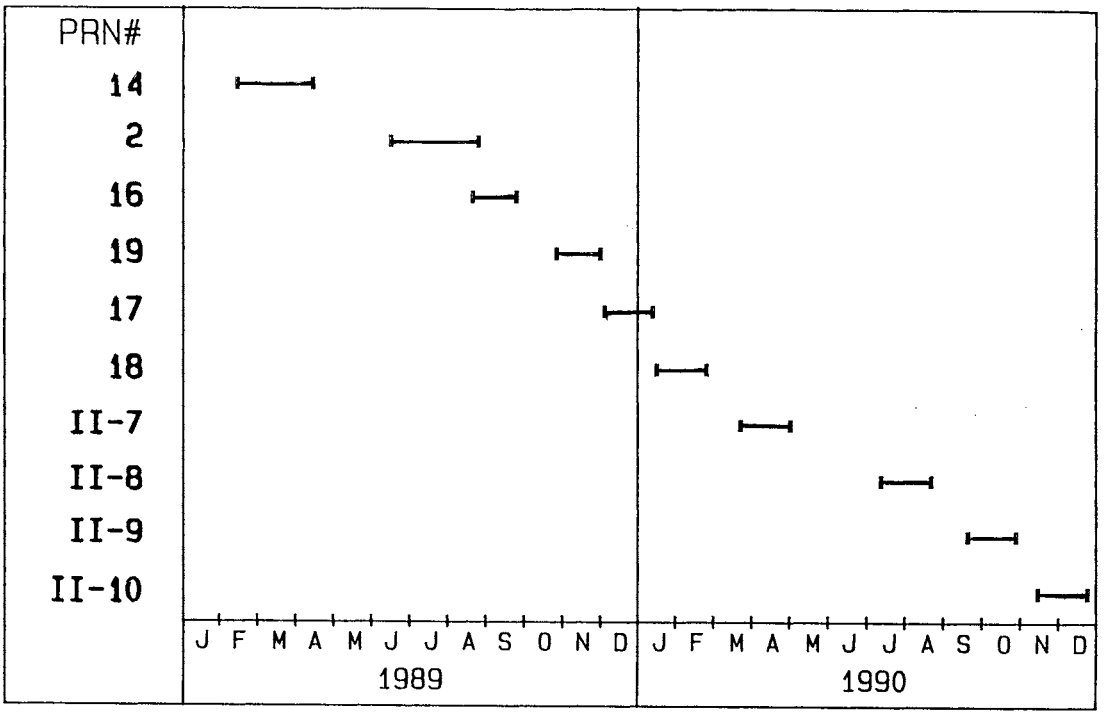
GPS BLK-II LAUNCH HISTORY

- GPS 014
  - LAUNCHED 14 FEB 89
  - OPERATIONAL 14 APR 89
  - E PLANE
- GPS 013
  - LAUNCHED 10 JUN 89
  - OPERATIONAL 18 AUG 89
  - B PLANE
- GPS 016
  - LAUNCHED 18 AUG 89
  - OPERATIONAL 18 OCT 89
  - E PLANE
- GPS 019
  - LAUNCHED 21 OCT 89
  - OPERATIONAL 23 NOV 89
  - A PLANE
- GPS 017
  - LAUNCHED 11 DEC 89
  - OPERATIONAL 5 JAN 90
  - D PLANE
- GPS 018
  - LAUNCHED 24 JAN 89
  - OPERATIONAL 14 FEB 90
  - F PLANE

NAVSTAR CONSTELLATION STATUS



BLOCK II SV LAUNCHES



PROJECTED BLOCK II SV BUILDUP DATA

ID	LAUNCH DATE	OPTIMAL PLANE POSITION	RAAN ** (DEG)	LAN ** (DEG)	ARG LAT ** (DEG)
SVN 8*	14 JUL 83	NEAR C3	87.6	116.9, 296.9	256.4
SVN 9*	13 JUN 84	NEAR C4	86.8	52.5, 232.5	129.2
SVN 10*	8 SEP 84	NEAR A3	325.5	133.6, 313.6	173.9
SVN 11*	9 OCT 85	NEAR C1	87.7	173.7, 353.7	9.7
14 II-1*	14 FEB 89	E1	197.3	149.4, 329.4	101.9
13 II-2*	10 JUN 89	B3	17.3	28.3, 208.3	219.7
16 II-3*	18 AUG 89	E3	197.3	167.1, 87.1	337.4
19 II-4*	21 OCT 89	A4	317.3	15.5, 195.5	314.2
17 II-5*	11 DEC 89	D3	137.3	191.8, 11.8	306.7
18 II-6*	11 JAN 90	F3	137.3	126.7, 306.7	176.4
21 II-7	8 MAR 90	B2	17.3	322.7, 142.7	88.7
20 II-8	12 JUN 90	D2	197.3	197.8, 17.8	198.7
15 II-9	7 JUL 90	E2	257.3	169.2, 349.2	21.4
23 II-10	20 NOV 90	D1	137.3	67.4, 247.4	57.9
24 II-11	22 JAN 91	F1	257.3	224.8, 44.8	132.7
25 II-12+	2 APR 91	B1	17.3	93.8, 273.8	350.7
26 II-13	16 JUL 91	F2	257.3	289.9, 109.9	262.9
22 II-14	13 AUG 91	A2	317.3	248.3, 68.3	59.7
27 II-15	19 NOV 91	E4	197.3	218.8, 38.8	240.7
28 II-16	14 FEB 92	A1	317.3	358.8, 178.8	280.7
29 II-17++	7 APR 92	C2	77.3	57.8, 237.9	158.7++

* - REPHASED SV
** - EPOCH = 1 JUL 90 0 (HR) 0 (MIN) 0 (SEC)
\+ - APPROXIMATE WORLDWIDE 2-D (HDOP<28) REACHED
++ - APPROXIMATE WORLDWIDE 3-D (HDOP<10) REACHED

NAVSTAR

GLOBAL POSITIONING SYSTEM

CONSTELLATION BUILDUP

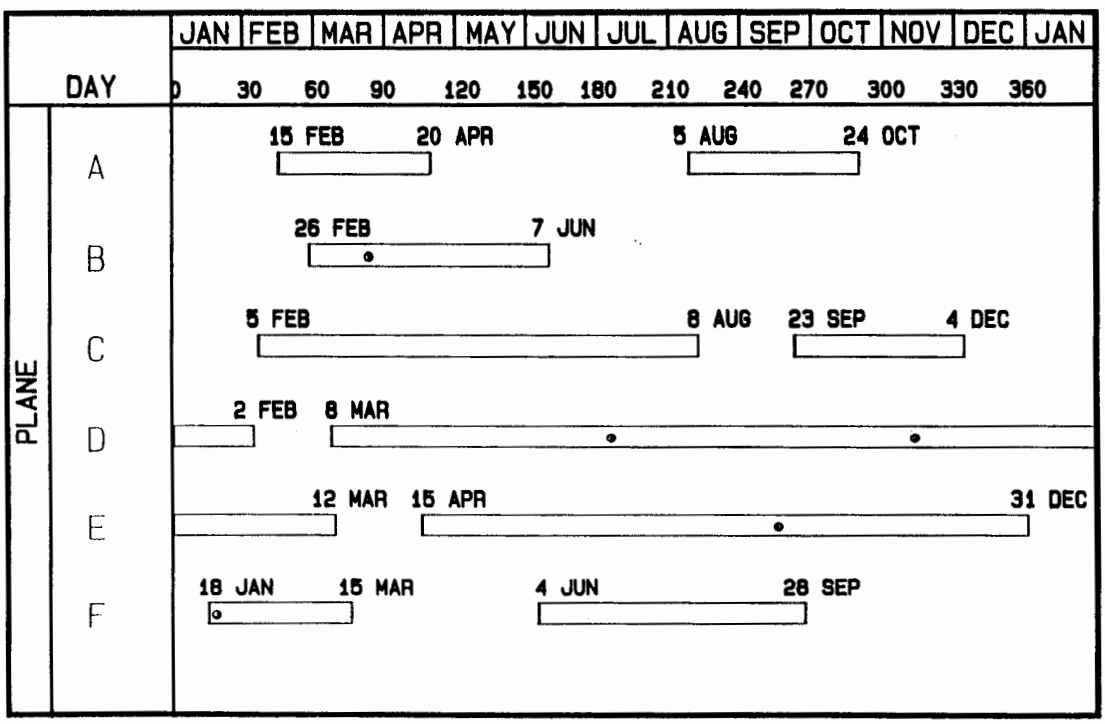
OVERVIEW

- o MANY DIFFERENT OPERATIONAL GPS CONSTELLATIONS HAVE BEEN CONSIDERED
  - 12 TO 24 SATELLITES
  - 3 AND 6 PLANES PLUS ONE PLANE FOR EVERY SATELLITE ("ROSETTE")
  - 0 TO 65° INCLINATIONS
  - 20,000 TO 120,000 Km ALTITUDES
- o KEY RESULTS
  - VIRTUALLY ALL CONSTELLATIONS CAN PROVIDE 16m WORLDWIDE 24 HOUR SEP
  - A 24 SPACECRAFT CONSTELLATION CAN PROVIDE WORLDWIDE COVERAGE WITH NO LOCATION WHERE PDOP PDOP EXCEEDS 6
  - A 21 SPACECRAFT CONSTELLATION CAN PROVIDE WORLDWIDE COVERAGE WITH NO LOCATION WHERE PDOP EXCEEDS 11
  - TWO CONSTELLATIONS WITH THE SAME INCLINATION, ALTITUDE, NUMBER OF SPACECRAFT AND PLANES CAN HAVE SIGNIFICANTLY DIFFERENT WORLDWIDE COVERAGE AND MAXIMUM PDOP

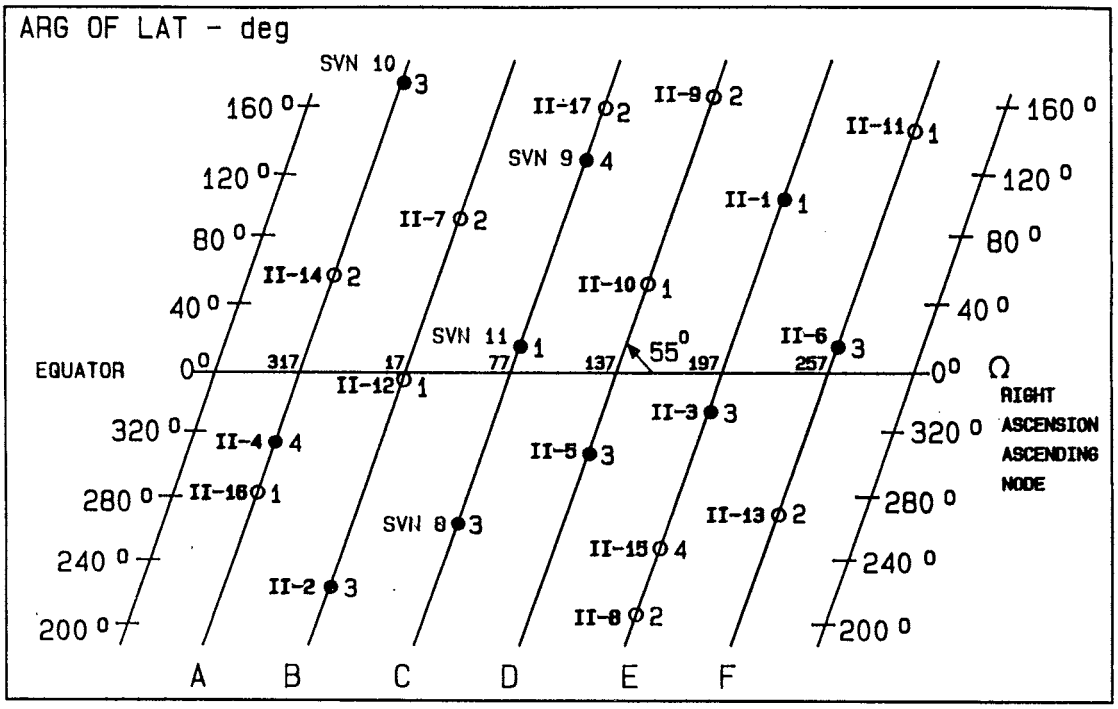
BLOCK II BUILDUP RATIONALE (1989-1992)

- o PHASE 1 FIRST 6 BLOCK II SVs WERE POSITIONED PRIMARILY TO AUGMENT YUMA COVERAGE IN SUPPORT OF O, T, & E TESTING
  - BLOCK I AND BLOCK II SV REPHASING
- o PHASE 2 LAUNCH DIRECTLY INTO OPERATIONAL LOCATIONS IN THE OPTIMAL 21 SV CONSTELLATION
- o PHASE 3 TRANSITION TO A 24 SATELLITE CONSTELLATION

LAUNCH OPPORTUNITIES - 1990

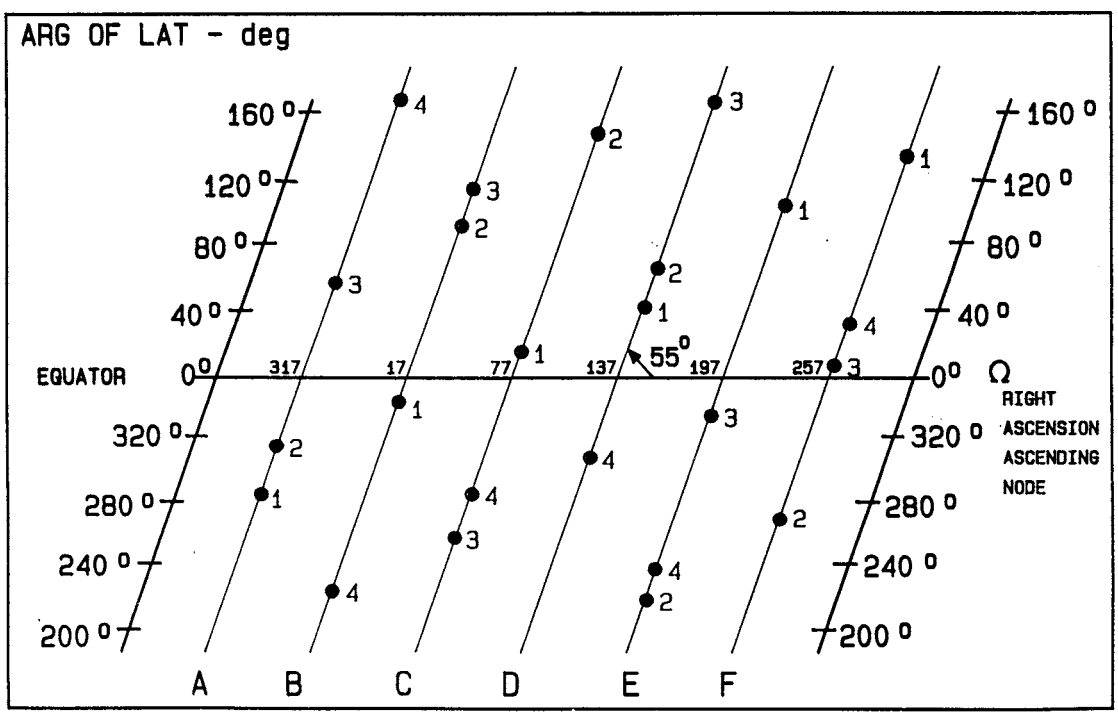


OPTIMAL 21 SATELLITE CONSTELLATION
(UP TO 3-D - 1992)



AS OF: JAN 1990

21 + 3 SV CONSTELLATION
(POST 1992)



AS OF: JAN 1990

GPS POSITION ERRORS

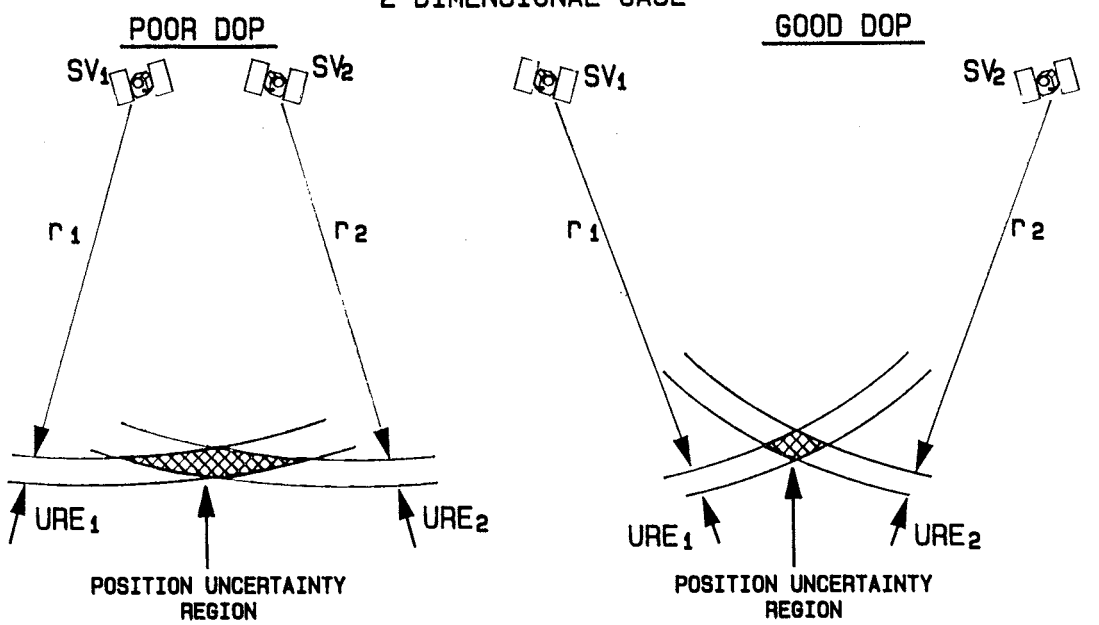
- o CAUSED BY ERRORS IN DETERMINING DISTANCE FROM USER TO GPS SPACECRAFT. THIS ERROR IS CALLED THE USER RANGE ERROR (URE). ERROR SOURCES ARE:
  - EPHEMERIS
  - SATELLITE CLOCK
  - INONOSPHERE/TROPOSPHERE
  - USER SET
  - SELECTIVE AVAILABILITY
- o URE IS BASED ON ESTIMATES OF ERROR SOURCES AND IS A STATISTICAL PARAMETER
- o GEOMETRY OF SATELLITES USED IN DETERMINING POSITION MAGNIFIES THE EFFECT OF URE
  - DILUTION OF PRECISION (DOP)
  - POSITION DOP (PDOP)
  - HORIZONTAL DOP (HDOP)
  - ALL DOPs ARE FUNCTIONS OF SATELLITE AND USER LOCATION AND NUMBER OF OPERATING SATELLITES
- o GIVEN A SATELLITE CONSTELLATION, DOP CAN BE COMPUTED BY SIMULATION FOR ANY LOCATION, OR ALL LOCATIONS, OVER 24 HRS

URE BUDGET

	P-CODE (1σ) METERS	C/A CODE (1σ) METERS
SPACE	3.5	3.5
CONTROL	4.3	4.3
USER EQUIPMENT		
IONO	2.3	5.0-10.0
TROPO	2.0	2.0
RECEIVER NOISE	1.5	7.5
MULTIPATH	1.2	1.2
OTHER	0.5	0.5
TOTAL	6.6	10.8-13.9

GPS POSITION ACCURACY

2 DIMENSIONAL CASE



INSTANTANEOUS 3-D POSITION ACCURACY = (PDOP) x $\left[\frac{URE_1^2 + URE_2^2 + URE_3^2 + URE_4^2}{4}\right]^{1/2}$

INSTANTANEOUS 2-D POSITION ACCURACY = (HDOP) x $\left(\frac{URE_1^2 + URE_2^2 + URE_3^2}{3}\right)^{1/2}$ (assuming no error in 4th measurement)

HANDY CONVERSION FORMULAS FOR GPS NAVIGATION ACCURACY

OPTIMAL 21 SPACECRAFT CONSTELLATION, 5° MASK ANGLE

1. USER RANGE ERROR (URE) IS 1σ (67% PROBABILITY) ERROR ALONG LINE OF SIGHT FROM USER TO SPACECRAFT

2. IF ALL SPACECRAFT HAVE SAME URE, WORLDWIDE ACCURACIES ARE:

   A) CEP = 1.16 URE (50% OF HORIZONTAL PLANE VALUES FALL WITHIN CIRCLE OF THIS RADIUS)

   B) 2dRMS = 2.67 URE (95% OF HORIZONTAL PLANE VALUES FALL WITHIN CIRCLE OF THIS RADIUS) (2σ 2-D ERROR) (CIVILIAN 2-D ACCURACY)

   C) SEP = 2.00 URE (50% OF 3 DIMENSIONAL VALUES FALL WITHIN THE SPHERE OF THIS RADIUS)

   D) 3-D RMS = 2.57 URE (67% OF 3 DIMENSIONAL VALUES FALL WITHIN THE SPHERE OF THIS RADIUS) (1σ 3-D ERROR)

SUMMARY

- o STAND ALONE GPS LOCATION ACCURACY IS URE x PDOP IF ALL SPACECRAFT HAVE SAME URE
  - URE CONTROLLED BY CONTROL SYSTEM AND IS A STATISTICAL PREDICTION
  - PDOP IS DETERMINED BY THE LOCATION AND NUMBER OF OPERATING SPACECRAFT AND USER LOCATION AND TIME
  - ACCURACY FUNCTION OF USER LOCATION AND TIME
- o PMD ACCURACY REQUIREMENTS DO NOT APPROPRIATELY DEFINE CONSTELLATION CHARACTERISTICS
  - LESS THAN 18 SPACECRAFT COULD PROVIDE 16m SEP WORLDWIDE
- o DEFINITION OF DEGRADED ACCURACY AREAS ARBITRARY AND GIVES THE WRONG MESSAGE TO USERS
  - MANY MISSIONS CAN BE ACCOMPLISHED WITH PDOP LARGER THAN 6
  - SOME MISSIONS (BLIND BOMBING) REQUIRE PDOP LESS THAN 6
  - GPS/INS INTEGRATION SIGNIFICANTLY IMPROVES ACCURACY IN AREAS OF LARGE PDOP
  - UE IMPROVEMENTS COULD SIGNIFICANTLY IMPROVE ACCURACY FOR STAND ALONE OPERATIONS IN AREAS OF LARGE PDOP
    - -- CLOCK RIDE
    - -- LOW COST ATOMIC STANDARD
  - SHOULD EXPAND INTERACTION WITH KEY USERS ON GPS ACCURACY ISSUES

POTENTIAL DEVELOPMENTS IN GPS CONSTELLATION

1. BLOCK I/BLOCK II REPHASING
2. CONSTELLATION ALTITUDE CHANGE
3. 4 SATELLITE TIME TRANSFER

REPHASING OBJECTIVES & CONSTRAINTS

o OBJECTIVES

- REPHASE 4 BLOCK I AND 6 BLOCK II SATELLITES FROM THE CURRENT TEST CONFIGURATION TO SELECTED OPTIMAL 21 SV CONSTELLATION LOCATIONS TO ACCOMPLISH THE FOLLOWING:

  1) MAXIMIZE WORLDWIDE 2-D AND 3-D NAVIGATION COVERAGE
  2) MINIMIZE THE SIZE AND NUMBER OF MANEUVERS REQUIRED
  3) SCHEDULE MANEUVERS TO MINIMIZE THE IMPACT ON MCS OPERATIONS

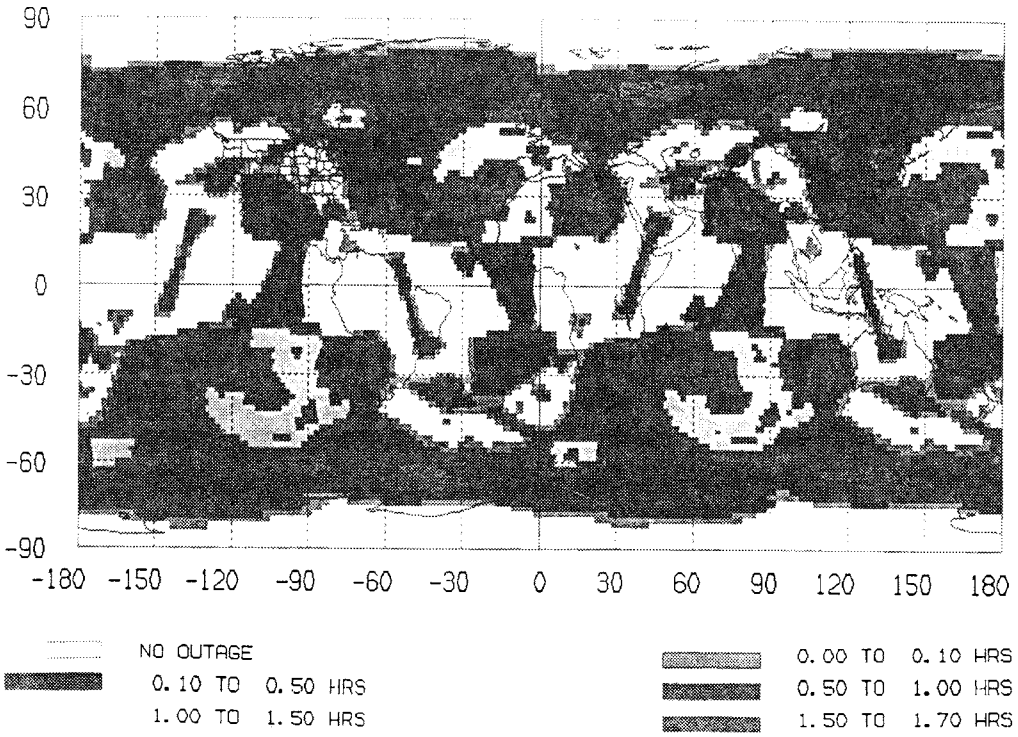
REPHASING OBJECTIVES & CONSTRAINTS (CONT'D)

o CONSTRAINTS

- TO AVOID COVERAGE DEGRADATION, REPHASING IS NOT RECOMMENDED UNTIL AT LEAST 6 BLOCK II SVs ARE ON ORBIT
- MANEUVERS NEED TO BE SCHEDULED SO AS NOT TO IMPACT O, T & E TESTING SIGNIFICANTLY
- BLOCK I MANEUVERS NEED TO BE SCHEDULED TO AVOID THE MAY - AUGUST SOLAR ECLIPSE SEASON

3-D NAV 21 SV CONSTELLATION WITHOUT REPHASING

EPOCH=1990. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- A1 A2 A3 B1 B2 B3* C1 C2 C3 D1
D2* D3* E1* E2 E3* F1 F2 F3 A4* E4 C4

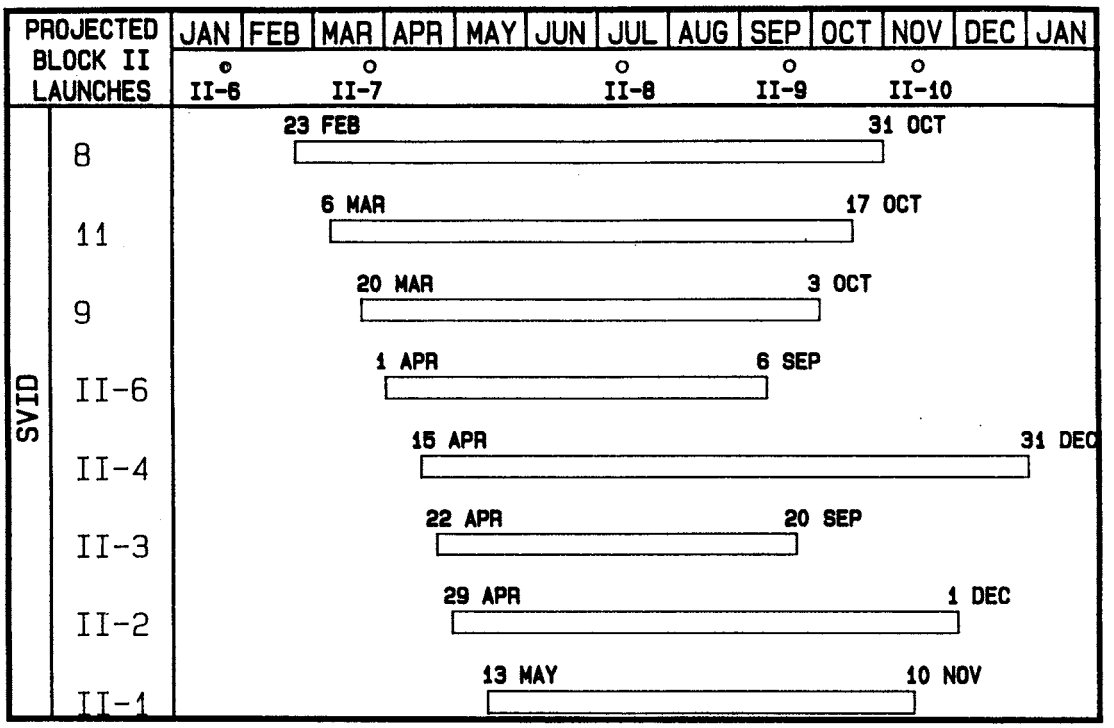


REPHASING MANOEUVRE DATA

SVID	NEW LOC	NODES (DEG)	IN-PLANE SHIFT (DEG)	NODAL DRIFT (DEG/DAY)	DELTA-V (FT/SEC)
8	256.4	117, 297	120.6	.21	-2.65
9	129.2	53, 233	33.6	.025	-0.32
10	173.9	134, 314	-1.9	-	-
11	9.7	174, 354	-43.3	-.13	1.6
II-1	E1	149, 329	50	.13	-1.37
II-2	B3	28, 208	-55	-.13	1.42
II-3	E3	267, 87	37	.12	-1.32
II-4	A4	15, 195	-177	-.34	3.56
II-5	D3	191.5, 11.5	6	-	-
II-6	F3	169, 349	-	-	-

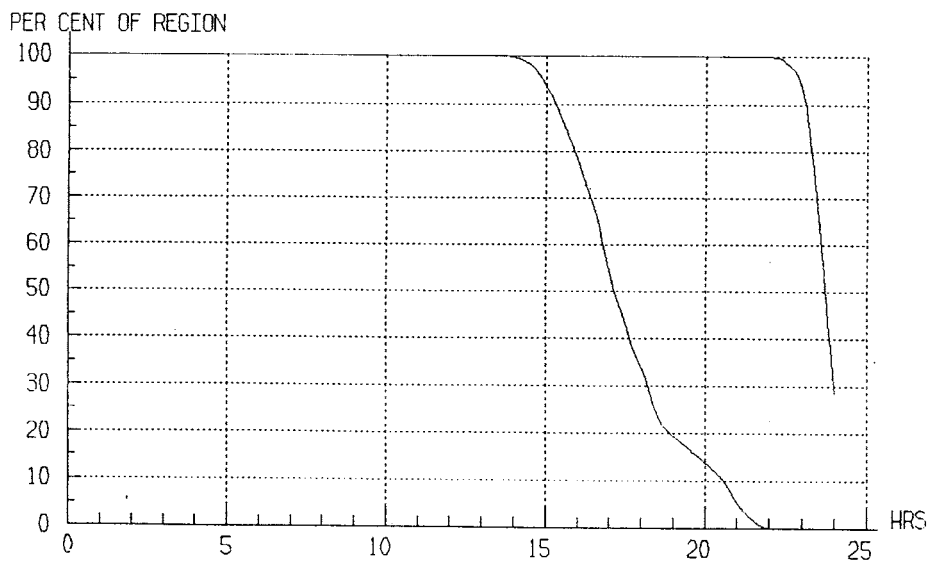
(REF DATE - 7/1/90)

REPHASING MANOEUVRE TIMELINE - 1990



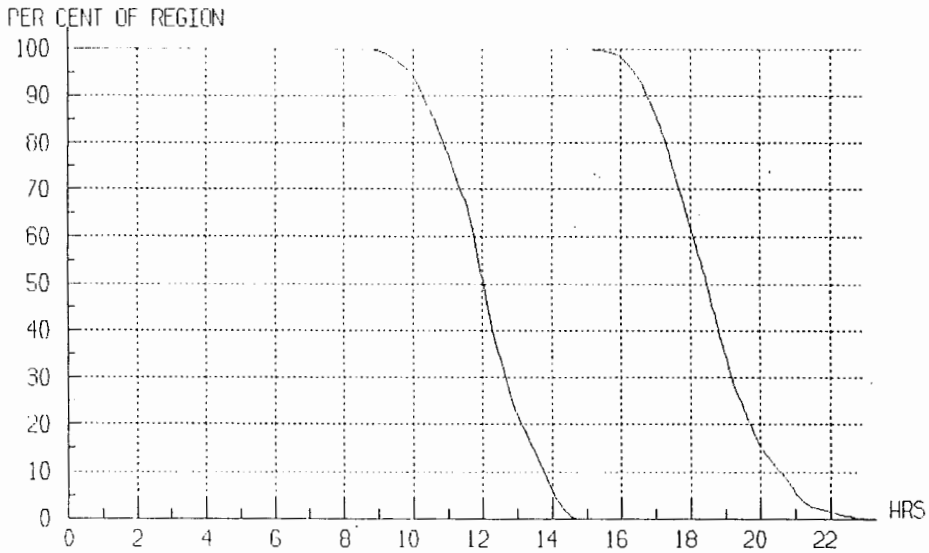
2-D NAV COMPARISON BEFORE AND AFTER REPHASING (HDOP=4)

EPOCH=1990. 12. 20. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN8 SVN9 SVN10 SVN11 E1 B3 E3 A4 D3 D2
B2 E2 F3 D1



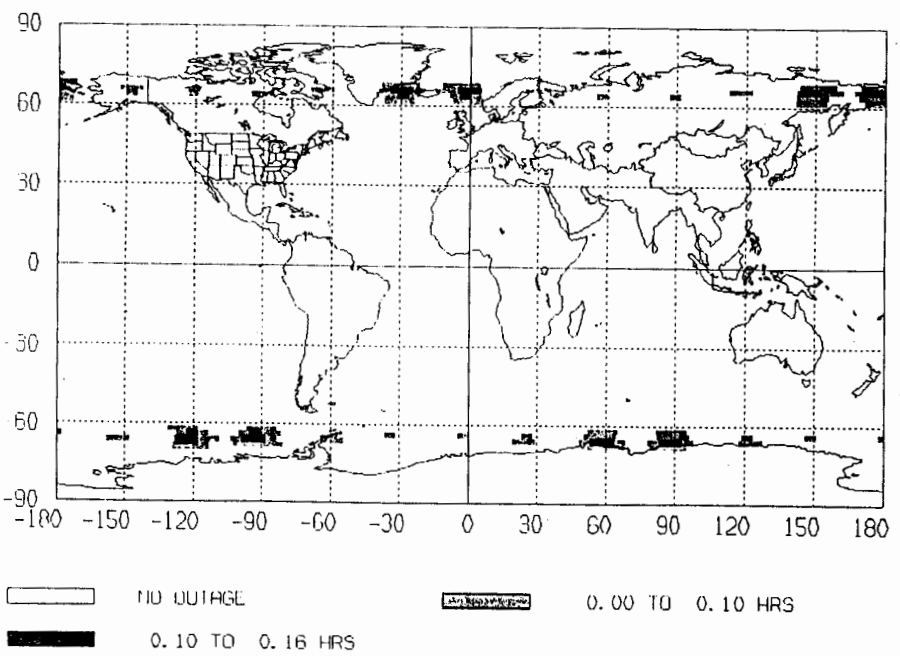
3-D NAV COMPARISON BEFORE AND AFTER REPHASING (PDOP=6)

EPOCH=1990. 12. 20. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN8 SVN9 SVN10 SVN11 E1 B3 E3 A4 D3 D2
B2 E2 F3 D1



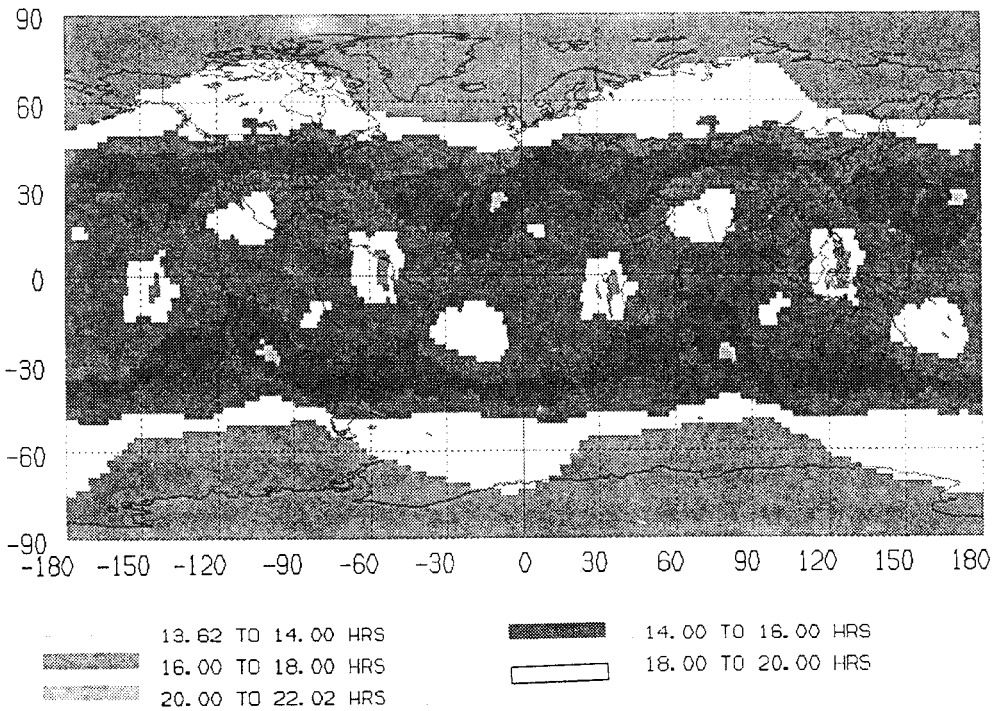
3-D NAV OPTIMAL 21 SV CONSTELLATION

EPOCH=1990. 7. 1. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- A1 A2 A3 B1 B2 B3 C1 C2 C3 D1
D2 D3 E1 E2 E3 F1 F2 F3 D4 E4 C4



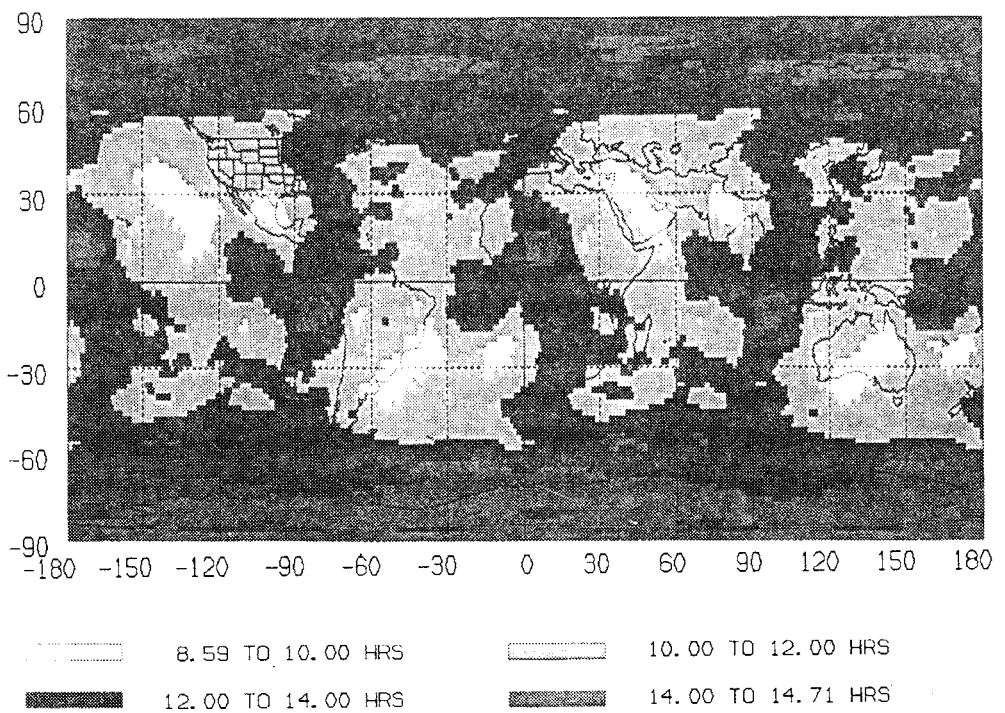
2-D NAV 6 BLK-I + 6 BLK-II (HDOP=4)

EPOCH=1990. 2. 15. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN3 SVN6 SVN8 SVN9 SVN10 SVN11 II-1 II-2 II-3 II-4
II-5 II-6



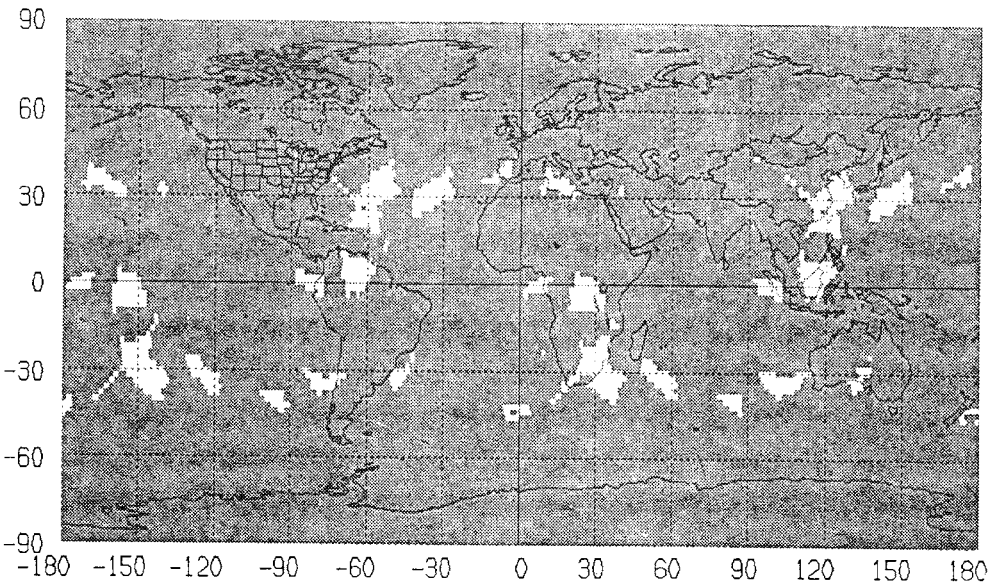
3-D NAV 6 BLK-I + 6 BLK-II SVS (PDOP=6)

EPOCH=1990. 2. 15. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN3 SVN6 SVN8 SVN9 SVN10 SVN11 II-1 II-2 II-3 II-4
II-5 II-6



2-D NAV 4 BLK-I + 10 BLK-II SVS (HDOP=4)

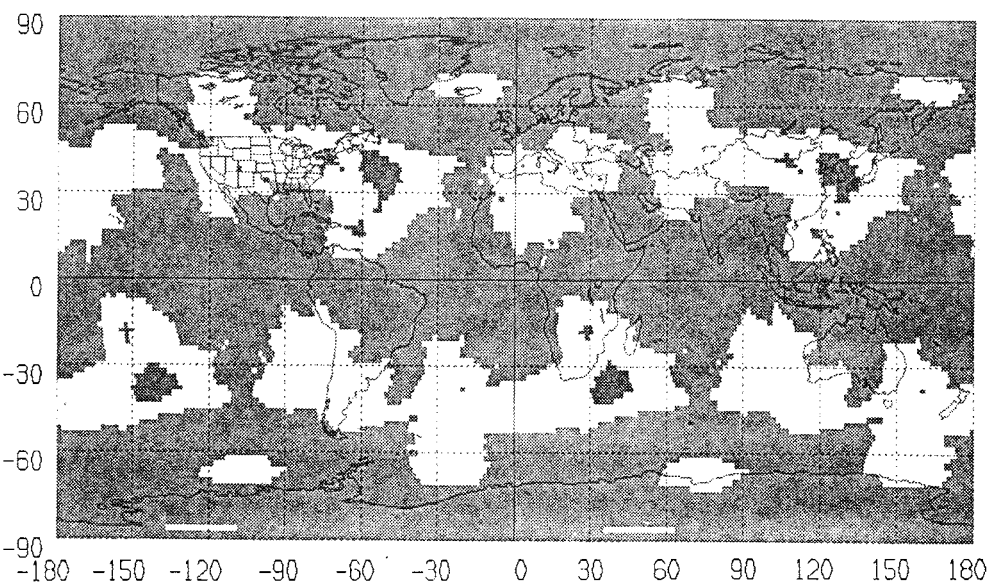
EPOCH=1990. 12. 20. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN8 SVN9 SVN10 SVN11 E1 B3 E3 A4 D3 D2
B2 E2 F3 D1



21.98 TO 22.00 HRS
22.00 TO 23.00 HRS
23.00 TO 24.00 HRS

3-D NAV 4 BLK-I + 10 BLK-II (PDOP=6)

EPOCH=1990. 12. 20. 0. 0.0 0.0 FROM 0.00 HRS TO 24.00HRS. EL= 5.0
LATITUDE FROM -90.0 TO 90.0 LONGITUDE FROM -180.0 TO 180.0
SATELLITES- SVN8 SVN9 SVN10 SVN11 E1 B3 E3 A4 D3 D2
B2 E2 F3 D1



14.00 TO 16.00 HRS
16.00 TO 18.00 HRS
18.00 TO 22.97 HRS

CONCLUSIONS

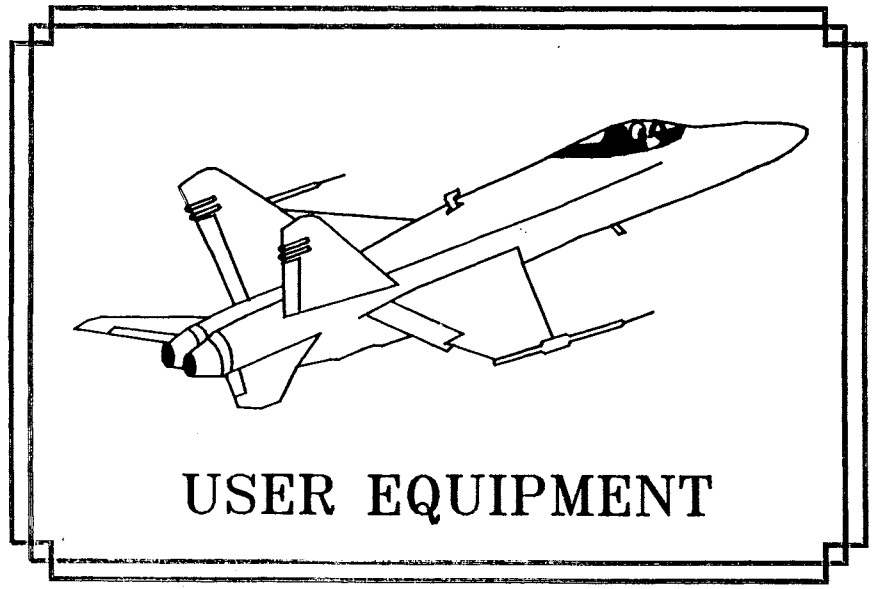
- NEAR FULL WORLDWIDE 2-D AND 3-D NAVIGATION COVERAGE IS EXPECTED NET APRIL 1991 AND APRIL 1992 RESPECTIVELY
- REPHASING IS NECESSARY TO ACHIEVE A DEPLOYMENT OF THE OPTIMAL 21 SV CONSTELLATION
- THE SLOW DRIFT OF THE CONSTELLATION FOR 8-10 MONTHS WILL NOT HAVE AN ADVERSE IMPACT ON 2-D COVERAGE AND A VERY MINIMAL ADVERSE IMPACT ON 3-D COVERAGE AT YUMA FOR ABOUT 1-2 MONTHS
- SATISFYING THE PMD 16 METER SEP 3-D SYSTEM POSITIONING ACCURACY REQUIREMENT IS NOT EXPECTED TO BE A PROBLEM

BLOCK IIA vs IIR DESIGN

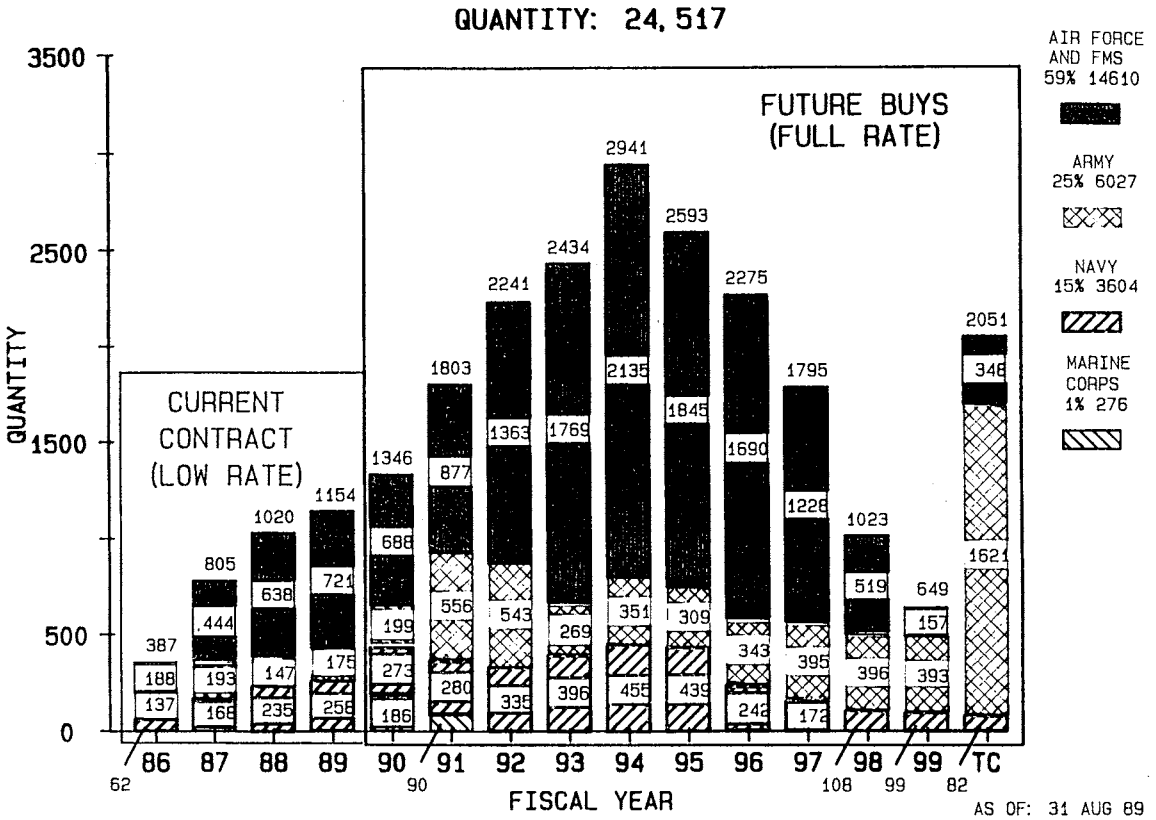
- INCREASED SYSTEM AUTONOMY
- LESS THAN SIX METER URE 180 DAYS AFTER LOSS OF GROUND SYSTEM PROGRAM OFFICE THROUGH USE OF CROSSLINK RANGING
- INCREASED SURVIVABILITY
  - JCS HARDENING REQUIREMENT INCREASED BY 10 TIMES
- INCREASED RELIABILITY, PRODUCIBILITY, SUPPORTABILITY AND INTEGRATION

CROSSLINK RANGING/ON-BOARD NAVIGATION DATA PROCESSING

- o CONCEPT
  - EACH SPACECRAFT PERIODICALLY MEASURES DISTANCE TO OTHER SPACECRAFT
  - EACH SPACECRAFT TRANSMITS MEASUREMENTS TO OTHER SPACECRAFT. MAY TRANSMIT DATA TO GROUND IF DESIRED
  - EACH SPACECRAFT COMPUTES CORRECTIONS TO PRE-LOADED NAVIGATION MESSAGES. OCS PROVIDES PRE-LOADED MESSAGES. 7 MONTHS OF MESSAGES UPLOADED ONCE PER MONTH
- o BENEFITS
  - MAINTAINS NAVIGATION ACCURACY FOR 6 MONTHS WITHOUT GROUND SUPPORT. SURVIVABLE CONTROL NOT REQUIRED. NO USER MODIFICATIONS REQUIRED
  - MINIMIZES UPLINK JAMMING CONCERNS
  - ONE UPLOAD/SPACECRAFT/MONTH INSTEAD OF 3 PER DAY REDUCED OCS WORKLOAD
  - REDUCES NEED FOR OVERSEAS ASSETS TO SUPPORT NAVIGATION UPLOADS
  - CAN PROVIDE IMPROVED NAVIGATION ACCURACY



GPS PHASE III UE RECEIVER QUANTITIES



PLATFORM INSTALLATIONS

FY88	FY89	FY90	FY91	FY92	FY93
EP-3E		RU-21H	MH-47E	UH-1	OH-58D
B-52G		OV-1E	CH-47D	EH-60A	► AH-64A
F-111A/E	EF-111A	RC-12H	MH-60K	► CH-53D	► SH-2G
F-16C/D (P)	VC-25	HH-65A	HH-2D	► F/A-18	► AH-1J
RC-135	TR-1	VH-3	► AV-8B	RH-53D	UC-12
MH-53J	UH-60A	C-9B	MH-53E	► CH-53E	► V-22
		ES-3	HH-60J	HH-60H	C-130
		E-2C	SH-60F	SH-60B	C-130 SCNS
		A-6G	VH-60	C-20	F-15D
		C-29 (P)	HU-25A	OV-10D	C-141B
		AC-130H	KC-130	E-6A	A-7D/+
		MC-130E	VP-3	HC-130	UH-1N
		MC-130H	RP-3D	F-14D	EC-135
			P-3C	US-3A	WC-135
			EA-6B	S-3B	B-1B
			EC-130C-IV	E-8	
			AC-130U	KC-10	
			F-16C/D (R) ►	F-16A/B	
			F-15E	F-111D/F	
				E-4	
				C-17A	
				MH-60G	
				KC-135A-R	
				A-10	
				RF-4C	
				F-4G	
				F-15	
				E-3A	

AIR FORCE

NAVY

ARMY

► SMALLER RECEIVER REQUIRED

GPS USER EQUIPMENT PRODUCTION

- o TOTAL PRODUCTION BUY: 26,000 UNITS
- o LOW RATE INITIATED: 1986
- o FULL RATE (~3000 UNITS/YR): 1990

DEMONSTRATED GPS PERFORMANCE

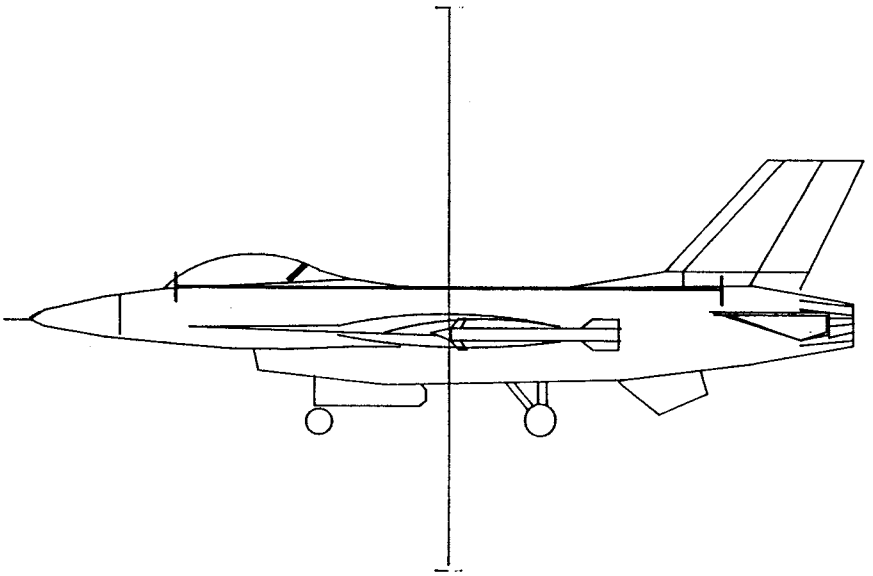
NAVIGATION

SH-60B	< 7 METERS SEP
F-16	< 10 METERS SEP
F-111	< 10 METERS SEP
B-52	< 10 METERS SEP
UH-60	TEST IN PROGRESS
MANPACK	TEST IN PROGRESS

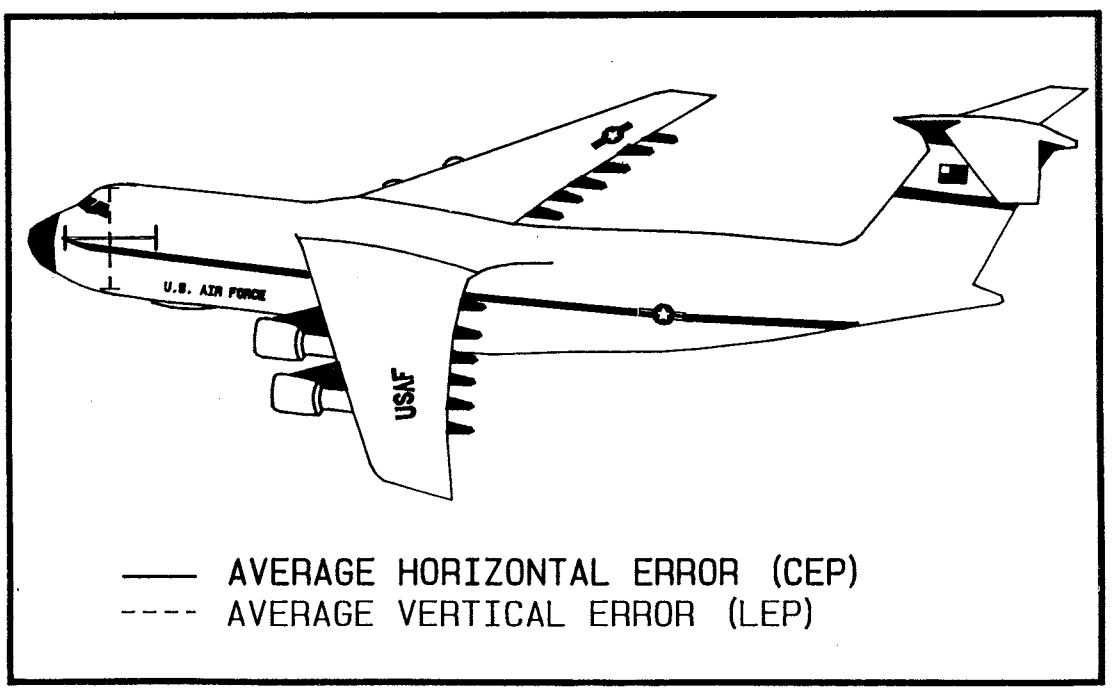
RELIABILITY (AS OF 8 DEC 88)

- AF - 572 HOURS MTBF BASED ON 25,000 HOURS
- NAVY -2129 HOURS MTBF BASED ON 34,000 HOURS
- ARMY - TESTING IN PROGRESS

F-16 15 NOVEMBER 1988
GPS AVERAGE VERTICAL ERROR (LEP)
GPS AVERAGE HORIZONTAL ERROR (CEP)

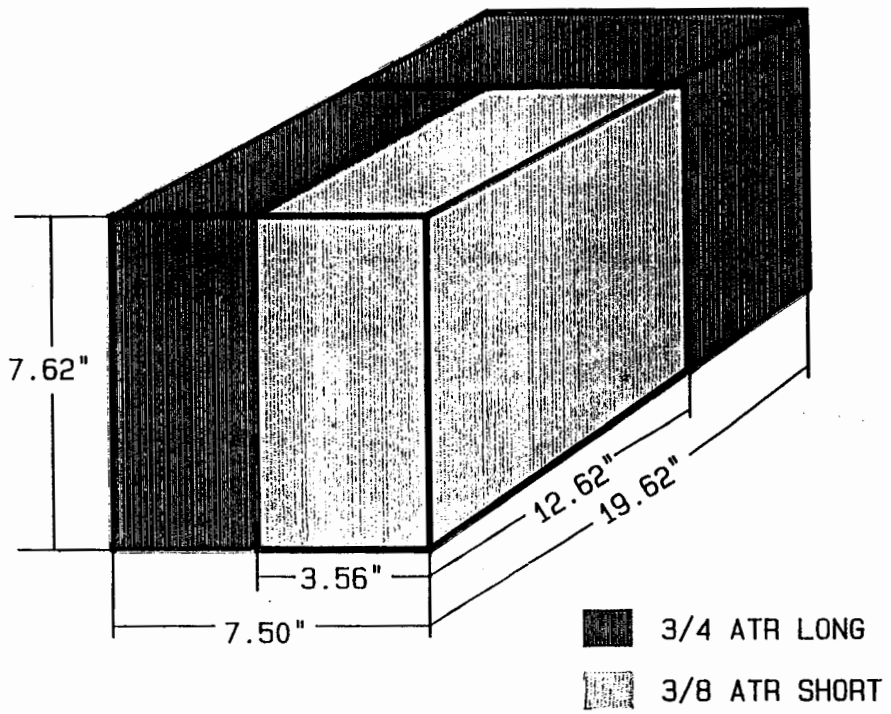


C-5 RELATIVE GPS ACCURACIES



3/4 ATR LONG vs 3/8 ATR SHORT

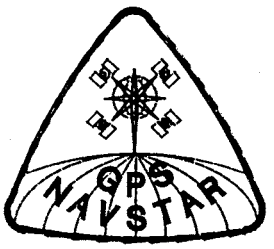
(REF DOC ARINC STD 404)



STRATEGY IMPLEMENTATION

3/8 ATR RECEIVER (RCVR 3A-M):

- o MAY CONDUCT AN NDI QUALIFICATION PROGRAM
  - LIMITED DEVELOPMENT PROGRAM MORE LIKELY
  - CONTRACTOR SAMPLES BEING TESTED AT YPG
- o CONTRACT AWARD FY 1990
- o QUALIFIED SOURCES BY FY 1992
  - DOWNSELECT AND AWARD PRODUCTION CONTRACT
  - PRODUCTION HARDWARE AVAILABLE FY 1993
- o CURRENTLY DEFINING TECHNICAL REQUIREMENTS



SELECTIVE AVAILABILITY

(SA)

SA/A-S DEFINITIONS

- SELECTIVE AVAILABILITY (SA)

  METHOD OF DENYING UNAUTHORIZED REAL-TIME USE OF FULL GPS ACCURACY PROVIDED IN PRECISE POSITIONING SERVICE (PPS)

- ANTI-SPOOFING (A-S)

  METHOD OF PROTECTING AGAINST HOSTILE IMITATION OF PPS SIGNAL

SELECTIVE AVAILABILITY/ANTI-SPOOFING POLICY

- BLOCK I SATELLITES EXEMPT
  - NO OASD POLICY CONCERNING BLOCK I SATELLITES
  - CURRENTLY NO PLANNED SA ON BLOCK I SATELLITES
  - U.S. RESERVES THE RIGHT TO IMPLEMENT SA
- BLOCK II SATELLITES
  - ALL SATELLITES WILL HAVE SA ENABLED AS SOON AS DECLARED OPERATIONAL
  - U.S. RESERVES RIGHT TO IMPLEMENT A-S
  - ON-ORBIT TESTING WILL BE PERFORMED FOR EACH SATELLITE, WITH THE OPTION FOR SA AND/OR A-S TURN-ON AT ANY TIME THE HEALTH BIT IS SET GOOD

GPS USER EQUIPMENT CONFIGURATIONS

- SPS: CONTAINS NO CRYPTOGRAPHY
- PPS: CONTAINS CRYPTOGRAPHY
  - EMBEDDED HARDWARE/SOFTWARE
  - PPS-SECURITY MODULE (PPS-SM)
  - AUXILIARY OUTPUT CHIP (AOC)

POSITIONAL ACCURACIES IN SA/A-S ENVIRONMENT

OPERATING MODE		AUTHORIZED USERS ***		UNAUTHORIZED USERS	
SA	A-S	P(Y)-CODE	C/A-CODE ONLY	P-CODE	C/A CODE ONLY
OFF	OFF	16 m	30 m	16 m	30 m
ON**	OFF	16 m	30 m	72 m	76 m
OFF	ON	16 m	30 m	30 m *	30 m
ON**	ON	16 m	30 m	76 m *	76 m

NOTE: THE P(Y), C/A, AND P- CODE ACCURACIES ARE IN METERS, SPHERICAL ERROR PROBABLE (SEP) (50TH PERCENTILE). NAVIGATION ACCURACY OF 76 m SEP IS EQUIVALENT TO 100 m 2D RMS (HORIZONTAL).

* ASSUMES DEFAULT TO C/A-CODE WHEN P-CODE IS UNAVAILABLE

** ASSUMES MAXIMUM NORMAL SA ERRORS INTRODUCED

*** AUTHORIZED MEAINING COMSEC AUTHORIZED TO RECEIVE THE GPS CRYPTOVARIABLES (CVs) AND SA/A-S ALGORITHMS

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-01 Number 3

A CIVIL NAVSAT OVERLAY TO GPS AND GLONASS

C . Rosetti
European Space Agency / HQ

INTRODUCTION

Two national (military) satellite navigation systems are at present under development. One of them

(GPS) is deployed by the U.S Department of Defence, the other one (GLONASS) by the USSR authorities.

Both systems are pretty well known from the technical literature and will not be described here. We shall limit our discussion to the deficiencies of these systems, as seen by the international civil user community, and we shall describe the main features of an "add-on" which might diminish or eliminate these deficiencies.

GPS and GLONASS will consist, in the end, of a space constellation of 24 satellites each. But these two systems are not compatible one with another, since coding, frequency plan, time and geodetic references are different. Interoperability cannot be achieved unless a bridge is built between them enabling the navigator to use a hybrid GPS/GLONASS constellation.

Moreover the US military operating agency envisages to degrade the precision of the already less precise (C/A) signals by applying what is called "selective availability"

While such degradation will not affect certain user, it will render the system less effective or even unusable to other. (e.g. instrument approaches and landing for aircraft; marine navigation in straits or harbour entries when visibility is low; precise time distribution etc.)

It is true that in some cases this can be circumvented by using differential techniques.

It is not known yet if the USSR intends or not to degrade the GLONASS signals by applying like the USA some kind of "selective availability".

From an operational point of view, satellite navigation can be regarded as "another navigation system" used in parallel to other means, or as "sole means of navigation".

In the first case there is no problem by using a "non guaranteed availability/integrity" military navigation system since traditionally navigators have used such systems in the past and still use them at present (e.g. VLF/OMEGA; LORAN C; TRANSIT etc.).

In the second case when such system is regarded as "sole means of navigation", the situation is very different and raises a number of problems related among others to the safety of navigation and liability in case of accidents. Other problems are of political nature, since, at least for the aeronautical navigator, one of the aims is to eliminate in the medium to long term, the expensive and ineffective terrestrial navigation systems. But this would expose users to the monopolistic situation of a global system run by the military authorities of two countries, and paid for by the taxpayers of these two countries which could raise another number of difficult problems.

Nevertheless, GPS and GLONASS are realities which have created an important industrial momentum, and which can basically be used in an efficient way by the international civil user community, if certain precautions are taken in the frame of an international cooperative effort.

What has to be avoided ?

- lack of integrity warning since neither GPS nor GLONASS can warn user about satellite malfunction in quasi real time
- the impact of degraded signals due to "selective availability"
- the relative incompatibility between GPS and GLONASS
- the monopolistic situation of these two systems
- the lack of legal liability in case of accidents occurring through the usage of these two systems

THE CIVIL OVERLAY

The European Space Agency in close cooperation with INMARSAT and with some national organisations (CNES in France, DLR in the Federal Republic of Germany) has developed the concept of a civil "add-on" or overlay to GPS and GLONASS.

This overlay will have the following functions:

- Broadcast from (at the beginning) geostationary "host" satellites wide band navigation signals fully compatible with both GPS and GLONASS
- These signals will contain integrity information on all GPS and GLONASS satellites. This information will be broadcasted in quasi-real time. (About 10 seconds after occurrence of a malfunction).
- The same signals will contain information permitting to compensate the difference of the time and geodetic references between GPS and GLONASS hence permitting the utilisation of a hybrid GPS/GLONASS satellite constellation.
- This geostationary overlay could form as well a worldwide high precision time distribution network.

If such an overlay is operated by a politically accepted international operational agency, this will diminish or eliminate the problems arising from the monopolistic situation of GPS and GLONASS.

It will as well create the "core" of a civil navigation satellite system capable of being enlarged to global coverage if circumstances will require such an evolution.

The SIGNAL STRUCTURE

From the frequency point of view, taking into account constraints arising from the host communications satellites able to carry a navigation package, as well as constraints dictated by the differences between GPS (CDMA) and GLONASS (FDMA), a frequency plan was devised by placing the civil signals as near as possible (as interferences would permit) to the GPS centre frequency, the centre frequency of this civil overlay signal being a multiple of the GLONASS FDM frequency step. For global geostationary coverage, only two frequencies would be required (Fig. 1). The separation between these two frequencies could be a submultiple of the GLONASS FDM frequency step.

For the signal structure, it was proposed to use two spread spectrum signals modulating a single carrier in phase quadrature. One of these signals will be similar to the GPS format, 1023 length codes, same PRN codes and data frame structure, the other signal will be similar to GLONASS, using 511 length codes (Fig. 2).

The next step is to see how integrity information and health warning can be integrated into the GPS and, respectively, GLONASS-like signals.

What are the requirements?

a) Positive acknowledgement of the health of all satellites

b) Alert of malfunction of a satellite

Both information (a) and (b) will have to have a repetition rate of, at the most, 10 seconds

c) Level of malfunction of a satellite (e.g. for aviation, phase of flight)

d) Optionally, the geographical area for which the alert is valid

Information (c) and (d) does not need to be broadcast every 10 seconds.

All this data will have to be inserted into the GPS and GLONASS signal structure.

The data relevant to GPS satellites will be transmitted in the GPS-like signal only, while the GLONASS one in the GLONASS like signal of the civil overlay.

GPS FRAME STRUCTURE

The GPS data is transmitted at 50 bits/second (BPSK) and is organised in tiered superframe, frames and subframes.

The superframe lasts 12.5 minutes and is made up of 25 frames with a duration of 30 seconds each. Each frame is subdivided into 5 sub-frames of 6 seconds each. A subframe contains 300 bits. This means that in order to comply with the 10 seconds integrity information/health warning requirement, this information will have to be inserted into every subframe.

The subframe data structure is shown in Fig.3. The first two words have a fixed format in every subframe.

Word 1 is the telemetry word (TLM), while words 2 is the handover word (HOW).
The telemetry word (TLM) is not used by standard GPS receivers and, hence, is fully available for other use.

The handover (HOW) contains a time of the week counter which, among other, is transferring time information to the user. This time information is needed for using the ephemeris information. This time of the week counter can be moved somewhere else in the GPS frame structure, thus allowing the HOW to be used for the alert warning.

The structure of the telemetry word is shown in Fig. 4a. From the total of 30 bits the first 8 bits form a preamble, while the last 6 bits are used for parity checking of the word. Sixteen remaining bits are left free but the last two of them might not be available, since they are used to ensure that parity ends in "00" perhaps as an aid to subframe acquisition.

The structure of the handover word (HOW) is shown in Fig. 4b. Since the information required by the user will be transferred somewhere else in the frame structure, the first 17 bits of this word could be used for the alert function.

In order to be able to give positive or negative acknowledgement of satellite health status for a full GPS configuration, it is necessary to transmit 24 bits (one bit per satellite). By eliminating the telemetry word and moving somewhere else in the system structure, the handover word, at least 30 bits remain available. However, it is necessary to transmit this information with a very high degree of integrity.

The 24 health warning bits are protected by 12 bits of error correction. For example, a (63,51) BCH code could be used when shortened to (36,24). This would allow 2 errors to be corrected. Assuming a gaussian channel symbol error rate of 10^{-3} then this code would achieve an error rate in the order of 10^{-12}. The existing parity bits could then be used to detect residual errors in both words. (See Fig. 5).

The option of sending information concerning the phase of the flight to which an alert is applicable, was retained. However, this information does not need to be transmitted (and cannot be transmitted) within a (maximum) 10 seconds delay as the health warning. Some spare bits in subframes 1 and 4 can be used for this purpose. The phase of flight data format is shown in Fig 6.

During discussions held in the frame of the ICAO FANS Committee, various opinions were expressed as to eventually indicate the geographical area for which an alert is applicable. Although it is not sure whether this option will be retained in the end, we have made provision for accommodating it into the signal structure.

GLONASS FRAME STRUCTURE

The GLONASS signal structure is in many aspects similar to the GPS one. The data rate is 50 bits per second (BPSK) and the signal consists of tiered superframe, frames and subframes.

The superframe is 2.5 minutes long and consists of 5 frames of 30 seconds each. Each frame consists of 15 subframes each of 2 seconds duration.

In Fig. 7, it is shown that there are 50 bits spare in GLONASS subframe 4. However, using only these bits for health warning would give an update rate of 30 seconds which is not satisfactory. Subframes 6 to 15 are, however, normally used for almanac data. In the case of the NAVSAT GLONASS-like signal, it is not necessary to transmit GLONASS satellite almanac information, but only information concerning the civil overlay which, in a fully deployed civil system, would amount to a total of 12 (non geostationary) satellites.

These frames will have to be reserved for such a purpose. However, as shown in Fig. 8, it is still possible to include health warning in subframes 9 and 14, in addition to subframe 4.

These health warning bits can be protected by the same code as used in the GPS- like signal. This is a shortened (63,51) BCH code, which would result in an error rate in the order of 10^{-12}. The 8 parity bits existing in the original GLONASS subframe would be retained for residual error-check after error correction of the health warning bits.

As in the GPS-like signal, it is required that "phase of the flight" alert is included in the GLONASS-like signal. It is possible to include this information in subframes 8 and 15 which gives an update rate of 30 seconds for 12 - malfunctioning satellites.

TIME TRANSFER

The GPS and the GLONASS navigation satellite systems have not the same time reference. GPS system time is, in principle, referenced to U.T.C. but the time broadcast by individual satellites may be affected by selective availability (S.A.). GLONASS is referenced within 1 microsecond to Moscow standard time. This accuracy is not good enough to permit single transfer of Moscow time into U.T.C.

It will be necessary to measure off-air, both GPS and GLONASS system time and to broadcast the differences between each of these two systems and NAVSAT U.T.C in the NAVSAT GPS and GLONASS-like signals. This information can be accommodated in the subframes 9 and 14 of the GLONASS-like signal and in subframe 4 of the GPS like signal.

For transferring time differences with a resolution of 10 nanoseconds 44 bits are necessary. The same (36,24) BCH code can be used for encoding (as is done for the health warning function).

The complete GPS-like frame structure is shown in Fig. 9.

SATELLITE NAVPACK REQUIREMENTS

Bearing in mind that navigation packages shall be flown on board "host" geostationary satellites in a first phase of implementation of a civil "overlay" to GPS and GLONASS, the most important requirement is to minimise the power and mass of the navigation transponder.

The link budget valid for geostationary navigation packages gives a required EIRP of 27.5 dBW (EOC).

By translating this EIRP into a DC power estimate for the NAVSAT transponder, we obtain the breakdown below:

ITEM	POWER WATTS
RF Amplifier chain	37.0
Local oscillator and circuits	4.3
ON/OFF TTC	0.2
Regulation circuits	10.36
DC/DC conversion	12.97
Contingency 10%	6.5
TOTAL	71.3

Assuming a C-Band uplink and a completely self-contained NAV-PAC, we obtain the mass breakdown shown below.

ITEM	MASS Kg
C-Band Antenna	0.900
C-Band Receive filter	0.040
Input switch	0.227
Transponder (fully redundant)	5.040
Output switch	0.227
Transmit filter	0.500
L-Band Antenna	2.400
RF & DC Cable Harness	0.260
Contingency 10%	0.960
TOTAL	10.55

CONCLUSION

In this paper, we started from generally accepted opinions concerning problems related to the civilian use of the GPS and GLONASS navigation satellite systems, and we drafted possible solutions in an attempt to solve these. The most important problem appears to be the integrity monitoring and health warning, since it has a number of facets of very different nature.

One of these is the method ued to determine a satellite's health status. In the technical domain, we suggest to measure time differences between satellites, since it is the easiest observable, and to integrate the information obtained into the GPS and GLONASS signal structure.

The method by which integrity is monitored is not necessarily a difficult issue, since in different parts of the world, the methods applied may differ without specific impact on the correct functioning of the overall system.

However, integrating this information into the GPS and GLONASS signal structure needs international acceptance and agreement, since one cannot imagine such a system using different standards in different parts of the world.

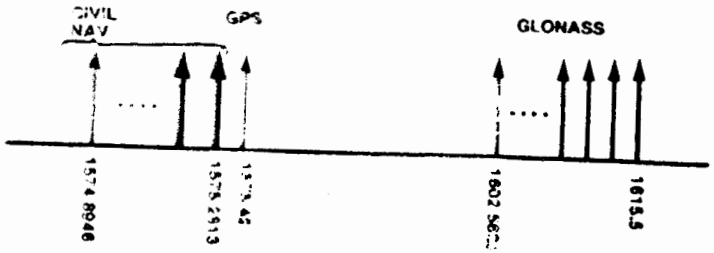


FIG.1
POSSIBLE FREQUENCY PLAN

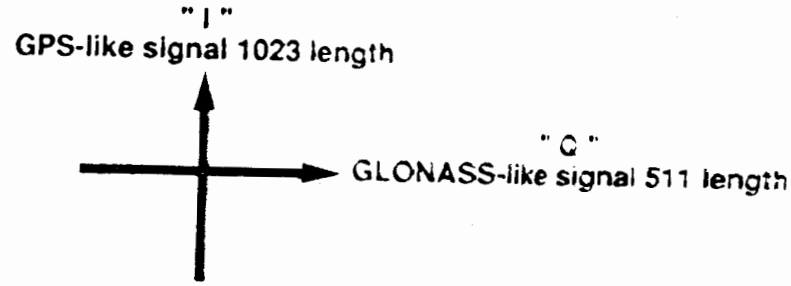


FIG.2
COMPOSITE " OVERLAY" SIGNAL

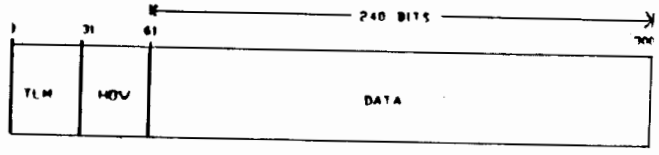


FIG. 3 GPS SUB-FRAME FORMAT

GPS Telemetry and Handover Word Structure

FIG. 4a

TLM FORMAT

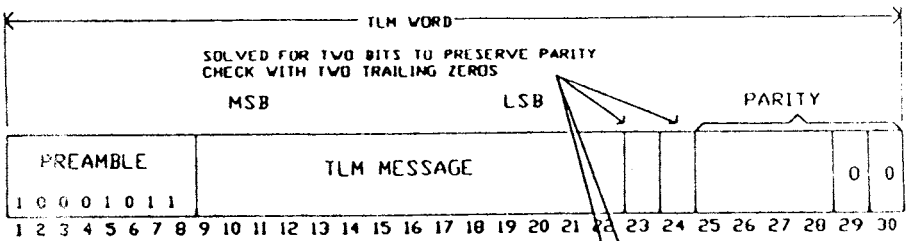
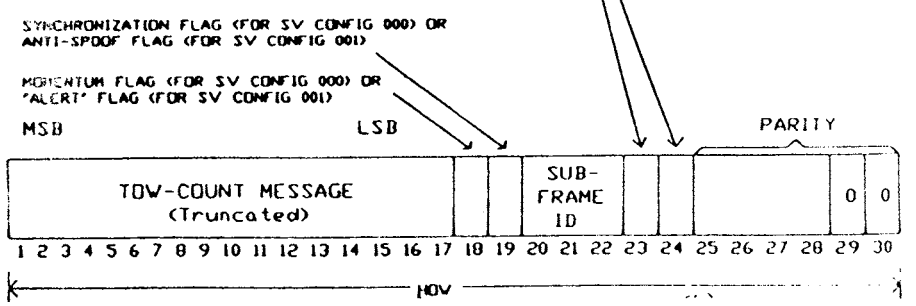


FIG. 4b

HOW FORMAT



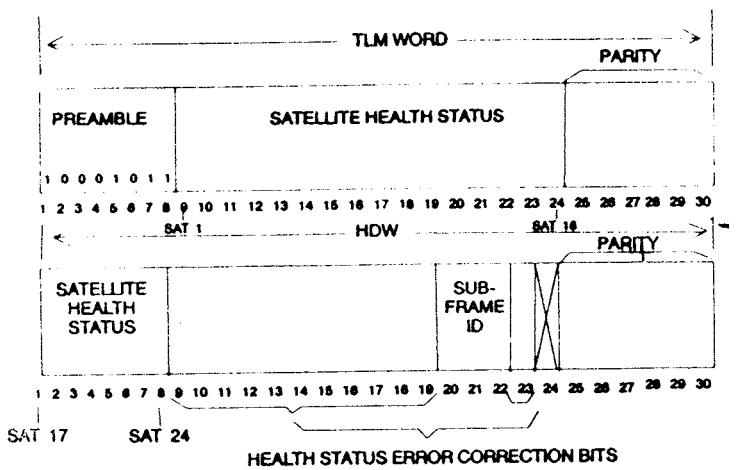


FIG. 5 NAVSAT SIGNAL STRUCTURE (GPS)

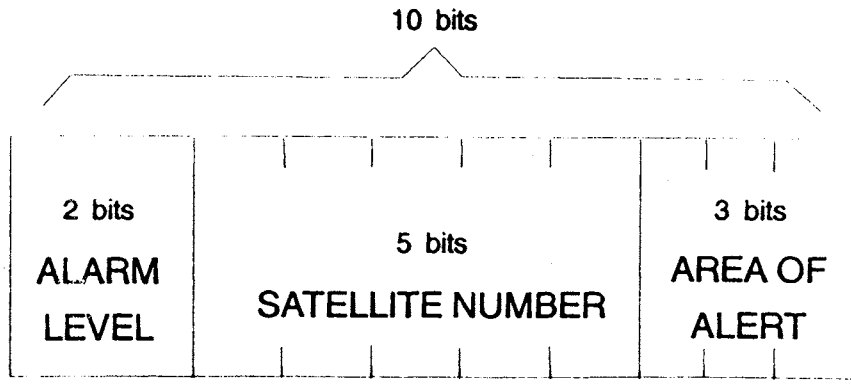


FIG 6 PHASE OF FLIGHT DATA FORMAT

GLONASS Frame Structure

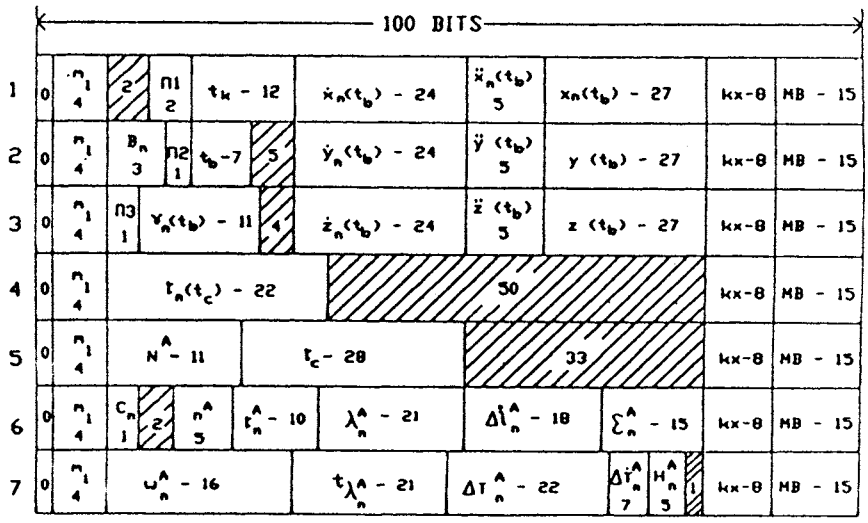


FIGURE 7 GLONASS FRAME STRUCTURE

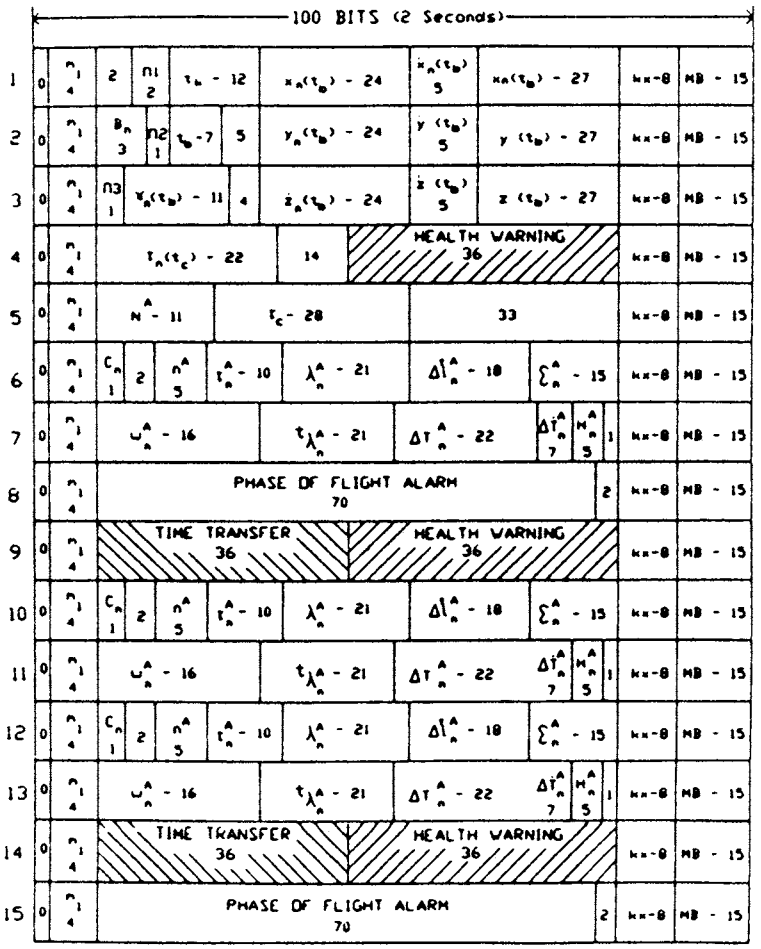


FIG. 8 NEW NAVSAT / GLONASS FRAME STRUCTURE

m_1	Sub-frame number
t_k	Frame beginning time of the current day
B_n	Three bits, 1 indicates satellite health, bits 2 & 3 unused.
Π_1	Two bits,
Π_2	Single bit,
Π_3	Single bit, '1' indicates frames 1-4, '0' frame 5.
$\tau_n(t_b)$	Frequency error at time t_b of transmitting satellite
$t_n(t_c)$	Error between satellite time and Glonass system time
t_c	Ephemeris data time correction
N^A	Day number over four years (starting at leap year)
H^A_n	Carrier character of navigation signal
λ^A_n	Greenwich Longitude of satellite no. n^A
Δi^A_n	Correction to satellite inclination, mean 63°
$t\lambda^A_n$	Time of first ascending node on day N^A
$\Delta\dot{T}^A_n$	Correction to satellite draconic orbit period, (mean 43200sec.)
ΔT^A_n	Satellite orbit period alteration speed
W^A_n	Argument of perigee
ϵ^A_n	Orbit eccentricity
C_n	Satellite Health Status
X_n, Y_n, Z_N	Satellite co-ordinates, (Greenwich rectangular co-ordinates)
$\dot{X}_n, \dot{Y}_n, \dot{Z}_n$	Speed vector components.
$\ddot{X}_n, \ddot{Y}_n, \ddot{Z}_n$	Acceleration components.
kx	Hamming code parity bits
MB	Unique synchronisation word
t_B	Moscow time when data was calculated
N^A	Satellite constellation number
t^A_n	Rough value of error between satellite time and System Time

SYMBOL DEFINITION FOR FIGURE 7 AND 8

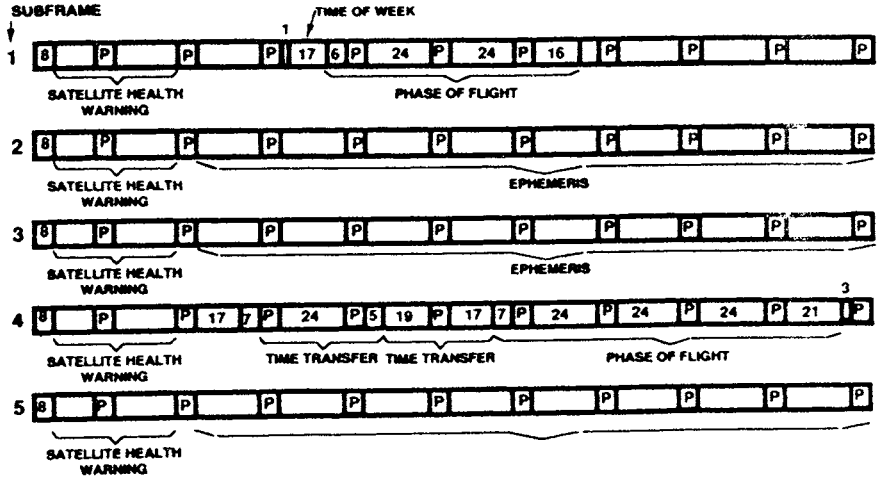


FIG. 9 PROPOSED NAVSAT GPS-like Format

Session APP-01 Number 4

APPLICATIONS OF T/F TO NAVIGATION AND POSITIONING

PH. HARTL
University Stuttgart
Institute of Navigation
D-7000 Stuttgart 1
F.R.G.

ABSTRACT

There exist numerous applications of radio navigation and positioning aids. Most of them are presently still based upon terrestrial systems. But for the future space-based systems will become dominant, because they offer in general better performance. This paper gives a review of the various techniques for radio navigation and positioning, in particular of those for use in space-based systems. All of them require both, good frequency and time information.

INTRODUCTION

Navigation is defined as the process of directing the movement of a vehicle from one place to another. This task encompasses position fixing, velocity and direction determination in a world of motion, telecommunication and remote sensing. Navigation has to model the traffic starting from the present and predicting the future traffic scenario.
Most of the present and future navigation and positioning systems apply radio techniques (NAV´89 1986, PRUNIERAS 1990, VAN BREMEN 1990, COLHOUN 1989, LINKWITZ ET AL. 1989, DURAND 1990, JACOB ET AL. 1989, HARTMAN 1990). Apart from this, the methods of inertial navigation and other on board sensors techniques are in use. The inertial navigation systems and the other techniques require, in general, updating by radio positioning, which determines by this way the accuracy finally achieveable. Radio navigation is still increasing its importance.
There are numerous applications of radio navigation techniques. Ground- based radar can be used for a/c location, for distance and direction measurement, for A/C tracking, and for ground-controlled approach. It can be applied for harbour and river surveillance, or for speed measurement in land traffic and as traffic signal actuator in land traffic. Airborne radio techniques are applied in collision avoidance systems, for Doppler navigation and altimetry as well as for mapping purposes. There are also various hyperbolic positioning systems in use. Finally, many different military applications could be mentioned, such as detection and tracking of any kind of manned vehicle and missile, and use of sidelooking radar as a reconnaissance tool. Many of the mentioned systems are terrestrial or autonomous systems, so far. It is, however, expected that in the near future most of the applications can use space-based systems as alternatives or as replacement of terrestrial and autonomous systems.
Any radio technique requires good frequency - performance, if high quality, good bandwidth economy and good power economy are requested. Precise time and frequency measurements are needed for modern high quality telecommunication and in particular for the precise ranging, position and velocity measurements.
The terrestrial navigation and positioning systems are in most cases less accurate than space -based systems. We shall, therefore, concentrate in the following in particular to these space based techniques.
2. Why space- based systems?

Any precise radio positioning system relies on a very good knowledge about the signal path and the signal velocity. In practically all cases one must assume in very precise systems that transmission occurs along the line of sight and that the influence of the atmosphere and of the earth´s surface is at least a secondary effect, if not negligible at all. These conditions can be only be met in terrestrial systems over very small distances. For longer distances carrier frequencies below UHF must be used, which are either reflected via the ionosphere, or guided along the ionosphere and the earth´s surface, or propagate along the earth´s surface. In all these cases many uncertainties degrade the actual performance. Even in the case of UHF and higher frequencies, where the line of sight conditions are met over modest distances, performance degradations can occur due to multipath and due to the fact that the near horizontal transmission is affected by ray bending.

Space based systems have several advantages. The signals are transmitted or received on earth at relatively high elevation angles. They pass the troposphere under steep angles and its influence is limited, for modest accuracies even negligible . Only the ionosphere requires more careful considerations. Its influence is reduced in very precise systems by two- frequency transmission, as is well known.
The signal paths from and to the satellite(s) are well defined and usually along the line of sight. Only for links close to the horizon problems could occur, such as bending and multipath. Such conditions, which are similar to the regular case of a terrestrial links, are avoided in most cases of practical applications by higher elevation angles. A further advantage of space- based systems is the fact, that measurements are consistent worldwide, as is well known from GPS- applications. Also frequency economy is improved due to the fact that much less transmitters are necessary due to the large coverage. The signal power can also be used more effective and with lower propagation loss.

3. Space - based positioning systems

We have to discriminate between one-way and two-way systems. One-way systems apply at one end a transmitter and at the other end a receiver. It is possible to use many receivers simultaneously. The system will not be saturated by mul-

tiuse, as is well known from TRANSIT/ TSIKADA (STANSELL 1978, ANODINA 1988, GLOBAL POSITIONING SYSTEM 1980, 1984, GUIER ET AL.1963, THOMPSON 1985, WELLS ET AL. 1986) and from GPS / GLONASS. It is also preferable to use the transmitter on board the satellite, as then many users can share the same satellite system simultaneously. Additional advantages are compact, low cost equipment at the user side, no problems to get permission for transmission, better frequency economy and less frequency coordination problems. For the military user it is essential that he can remain passive.

3.1 One- way Systems; instant position determination

There are various one - way systems in use or in introduction. We shall first discriminate among those, which operate with one satellite in contact with the user. Among those TRANSIT and TSIKADA are typical representatives. Before discussing the performance characteristics we have to make clear the following. If a single satellite should allow to determine the 3-dimensional position instantaneously, then it is necessary to know at any time the position of the satellite, as is the case also in the two-way system and in a multi satellite system as well. The special demand for the one- way system is , however: One has to determine at the location of the user´s receiver the polar coordinates with respect to the single satellite. This can be done by two different methods:
The satellite is stabilised in three axis and transmits some characteristic patterns in two orthogonal directions. In addition it transmits signals periodically similar to what is known from the ranging signals of GPS. Under these conditions the receiver must be equipped with a clock at least as accurate as the one on board the satellite.
Alternatively, the satellite transmits the ranging signal and the user has not only the very stable clock, but also a direction measurement system, which allows him to determine the direction where the satellite transmits the signal, he is receiving.

3.2 One- way Systems; integrated Doppler

The TRANSIT system demonstrates that it is also possible to determine the position from one-way doppler counts. Good accuracy can be achieved, if the transmitter and the receiver are equipped with very stable clocks, and if the user is not moving, or if the velocity vector of the vehicle is precisely known.
This Doppler system has the advantage that the accuracy of the absolute time is not as critical for the user equipment as is the case in the one-way ranging. But it can also not be as accurate, because the position determination depends also on the modelling of the various parameters included in the calculation. Details are known from TRANSIT.
As opposed to the TRANSIT system which uses the transmitter on board the satellite a different solution is applied in the DORIS system. This tracking system is used by CNES to determine the satellite orbit, for example TOPEX or SPOT. For this purpose 50 transmitters have been displaced around the globe, each equipped with a very stable clock and with some channel capacity to transmit grounds stations´s identification and some meteorological and technical information from ground. Of course, a very stable clock is also necessary on board as well. The orbit determination possible with this Doppler information depends in its accuracy on the stability of the individual transmitter frequencies. Both components are important, the long term drift and the short term stability. This latter one must be guaranteed for the passover time, i.e. for about 10 minutes. Any drift during that time will lead to some error in the orbit determination (SERVICE ARGOS 1978).

3.3 One- way ranging with many satellites

The GPS- system is well known. It is a one-way system, which avoids the constraints mentioned above. This is possible because of the simultaneous visibility of at least four satellites by any user , i.e. GPS- receiver. GPS- NAVSTAR is a microwave system. The satellites transmit with very good accuracy the frequency, the ranging information and the GPS-time. The synchronization of the individual satellit clocks is very good and is performed by a ground-based network. Ranging to at least 4 satellites allows the three- dimensional position determination. This is obvious, because the satellites indicate their own position during the signal transmission and the receiver can determine the time of arrival of the signals. This assumes a precise clock at the user´s receiver, if 3 satellites are used. With 4 satellites a bias of the user´s clock can be eliminated, as under these circumstances only the delay- differences of the 4 satellites are relevant.
With the measured signal delay at the user´s position one knows the range to the satellite. The Doppler measurements by comparison between the received frequencies and the receivers frequency allow the instanteous velocity vector of the user to be determined.
If the user applies the interferometric mode then angular or attitude measurements are also possible.

3.4 Two-way Systems; satellite laser ranging

One well known technique for orbit determination is laser ranging (NATO 1988). This is the most accurate operational tracking technique, so far. It is in the order of one to some cm ranging accuracy, but is similar in its constraints to the Doppler measurement of transit: With one ground station one can only get sufficient orbit information by measurements over some arc. So, it assumes again tracking from a stationary position. This requirement and the fact that laser ranging is dependent on the weather conditions , prevent its use in operational navigation. So, in other words, this technique is good for gaining the orbital elements of a satellite system and for control of the on board time, but is not of use for the direct navigation of a vehicle.

Laser ranging requires good short time stability. The present systems use very short, but powerful laser signals. In the future we intend cw- semiconductor laser, instead. This makes the laser ground stations small and cheap. It is planned to be introduced as part of the PRAREE- system.

3.5 Two-way Systems; Satellite Microwave Ranging

A two-way microwave ranging system is PRARE, which will be flown for the first time on ERS-1, the first ESA remote sensing satellite. Launch is foreseen end of 1990. This system operates in the space based mode, i.e. transmitter and receiver are on board the satellite. On ground a transponder system is necessary. This means that the system is again designed for orbit determination and for direct vehicle navigation. It applies PN- modulation and uses two frequencies (S-band, X-band) on the down-link for compensation of ionospheric effects and one uplink X-band frequency.
This system is competitive to laser ranging. It can also be used for performing the error analysis of a GPS- satellite, if a PRARE space segment is installed in a GPS, as is under discussion. In the poster session of this symposium the advantages of GPS/PRARE are shortly discussed.

3.6 Two-way Systems; Satellite Microwave and Laser Ranging

As it is necessary to compensate the tropospheric effect very precisely in microwave techniques, it is proposed that PRARE will be used on board the European Polar Platform in an extended version, which we call PRAREE. In this case it is proposed that synchronous to the PN- modulation of the microwave down-link signal an amplitude modulated semiconductor laser signal will be used on board the Platform with a transmitter power of about 0.5 Watt. This signal is received and detected on ground simultanously with the microwave signal and the delay can be used to study the tropospheric effects. As is well known, optical signals are not influenced be the wet component of the tropospheric, whereas microwave signals are. With the combined system we hope to achieve mm and even sub-mm ranging accuracies and to learn how to model the wet componention the tropospheric microwave delay in the future to an accuracy of a few mm or even better.

3.7 Two-way Microwave Ranging from Geosynchronous Satellites

GEOSTAR / LOCSTAR applies a technique, which is based upon the use of two types of transponders, one in each satellite and one in each user´s vehicle (O´NEIL 1984). A central ground station can transmit to two- or in the three dimensional case to 3- geosynchronous satellites. These satellites retransmit the signal to the vehicles, which send the amplified signals back to the satellite´s nearest groundstation. The ground station receives these echos, determines the 2 or 3 satellite-vehicle ranging data sets resp. and calculates the position of the vehicle. This information is sent back via one of the satellites to the user. This user can also transmit messages over one of the satellite to the ground station.

There is a demand for short term stability for the clocks at the satellite ground stations.

4. Time

4.1 Time Transfer

Time transfer by use of satellites is well known from the GPS-system. Another type of time transfer one was used in the NAVSAT- experiment. Under the assumption of a very stable on-board clock and under the assumption that the distance between the transmitting satellite and the receiving station is known, say to about 3 km accuracy, the time transfer could be performed with 10 us accuracy by this way (SCHLÜTER 1988, DFG 1990).

4.2 Time Synchronization

If one requires a time synchronization globally for atomic clocks with accuracies in the nano- and subnanosecond range one has to apply the two-way split transmission / reception. MITREX can perform this via telecommunication satellites. The improved MITREX 2500 Modem in its digital version. It is described in another paper of this symposium.

4.3 Use of time information for telecommunication

Assume that a user intends to transmit very seldom some messages and that the transmitter wants to remain undetected then it is necessary that the messages are sent as bursts during very small time slots. Both, the transmitter and the receiver must know about the time slots. A precise time schedule is required. The more stable the two clocks can be kept the shorter can be the time of transmission for a given message.
A modified version for longer messages could be frequency hopping. Here it is important to know very precisely the discrete frequencies among which the transmission can occur.

4.4 Use of time information for Collison Avoidance

Assume that all vehicles are equipped with very precise time information. It is then possible to equip each participant with a simple instrument transmitting some pulses say every second. If a vehicle receives such a signal and determines the time difference between transmission and reception then it can calculate the range and the range rate of the correspond-

ing transmitter. So, it becomes possible to see, if such vehicle is close and/ or ifit is approaching dangerously.

5. Radar Altimetry

There exist various methods for radar altimetry. The best known technique is the one used by airborne altimeters on the basis of FM- modulated carrier signals. The transmitted signal is swept in its carrier frequency between some lower and upper frequencies f1, f2, the bandwidth of which being B=f2 - f1. The received echo signal is of the same character. It is shifted in time according to the two-way signal propagation time from the A/C to ground and back to the A/C. The frequency difference between transmitted and received signal is a direct measure of the altitude of the A/C.
In the spaceborne radar altimetry the pulse compression method is used. The reason for this is the following: Very short pulses should be used in order to achieve a very high spatial resolution. However, a short impulse requires very high transmission voltages to achieve a large energy per pulse.This is to be avoided. For this reason it is general use to feed the short impulse through a dispersive filter, such as a Surface Acoustic Waves (SAW) filter, before transmission. The different spectral components of the pulse are then delayed differently and the output pulse of the SAW- filter is prolonged. The instantaneous output power of the radar- transmitter is then smaller. The received echoe is again fed into a SAW- filter in such a way that the original delays of the frequency components are now reversed. The detection process can, therefore, act as if the very short pulse would have transmitted. Modern radar altimeters can also use PN- coded signals and apply correlation techniques to realize the fact of pulse extension / compression (KIM ET AL. 1989, SCHÜSSLER 1989).
Anyhow, it is necessary that the delay time of the frequencies and phases of such altimeters are measured very precise. An improvement in this respect reflects immediately in improvement of accuracy.

6. SAR

The Synthetic Aperture Radar systems (SAR), which are used with great success in spaceborne and airborne remote sensing, apply also the pulse compression technique described above. In addition, SAR makes use of the antenna synthesis which relies on excellent Doppler shift measurements of the received vs. the transmitted signal.

7.OUTLOOK

There are many applications of time and frequency in navigation. They will even increase in the future, when the higher frequency ranges can be applied from or in space. Laser systems might come up and the demands to achieve optical signal sources of excellent frequency stability is almost certain. Any improvement of frequency stability can also open new applications. For example, it might be possible in the future to measure the gravity field via the emitted frequency and determine by this way directly the precise orbit of a satellite.

8. REFERENCES

- T.G. Anodina: The Glonass system technical characteristics and performance. FANS/4-WP/70 & WP/75. 6. Mai 1988.
- G. Blewitt, T.P. Yunck, S.M. Lichten, u.a.: GPS Geodesy: A Status Report. In: K. Linkwitz, U. Hangleiter (Eds.): High Pre cision Navigation. Integration of Navigational and Geodetic Me thods. Heidelberg 1989.
- E. van Bremen: Loran C sur les Routes: une Idee Nouvelle en Radionavigation. In: Navigation. Revue Technique de Navigation Maritime, Aerienne, Spatiale et Terrestre. Vol. 38, Janvier 1990, No. 149.
- Ch. Colhoun: The New Market for Vehicle Location Systems; Where do we stand? In:Deutsche Gesellschaft für Ortung und Navigation e.V. (ed.): Land Vehicle Navigation 1989. 4.-7. July 1989 Münster, Federal Republic of Germany. Köln 1989.
- DFG Deutsche Forschungsgemeinschaft: Satellitengeodäsie. Ergebnisse aus dem gleichnamigen Sonderforschungsbereich der Technischen Universität München. Hrsg. von Manfred Schneider. Weinheim 1990.
- J.-M. Durand: Inconvenients du GPS: Etudes Comparative des Soutions Adaptees a l'Aeronautique Civile. In: Navigation. Revue Technique de Navigation Maritime, Aerienne, Spatiale et Terrestre. Vol. 38, Janvier 1990, No. 149.

- Global Positioning System, Vol.I Papers, published in NAVIGATION and reprinted by the INSTITUTE OF NAVIGATION, Washington D.C. 1980.
- Global Positioning System, Vol.II Papers, published in NAVIGATION and reprinted by the INSTITUTE OF NAVIGATION. Washington D.C. 1984.
- W.H. Guier: Navigation Using Artificial Satellites. The Transit System. In: G. Veis (ed.): The Use of Artificial Satellites for Geodesy. Amsterdam 1963.
- R. Hartman: la Navigation Integree: Inertie a Laser/GPS. In: Navigation. Revue Technique de Navigation Maritime, Aerienne, Spatiale et Terrestre. Vol.38, Janvier 1990, No. 149.
- Th. Jacob: Integrated Navigation System for Approach Guidance for Regional Air-Traffic Using GPS. In: K. Linkwitz, U. Hangleiter (Eds.): High Precision Navigation. Integration of Navigational and Geodetic Methods. Heidelberg 1989.
- Young-Hoon Kim: A KU-Band Solid State Radar Altimeter for Topographic Application. In: K. Linkwitz, U. Hangleiter (Eds.): High Precision Navigation. Integration of Navigational and Geodetic Methods. Heidelberg 1989.
- Klaus Linkwitz; Wolfgang Möhlenbrink: Navigational Methods of Measurement in Geodetic-Surveying. In: K. Linkwitz, U. Hangleiter (Eds.): High Precision Navigation. Integration of Navigational and Geodetic Methods. Heidelberg 1989.
- NATO: Introduction to Navstar/GPS user equipment. ANP-2 Edition No.1. September 1988.
- NAV'89: World-Wide Navigation into the 21st Century. Papers presented at the 1986 Conference of the Royal Insti-

tute of Navigation, Brighton 1986.
- G.K. O´Neil: The Geostar system. In: The Global Civil Satellite Navigation System, Collected papers of NAV´84 Conference, Royal Institute of Navigation, U.K., May, 1984, pp. 339-345,
- J. Prunieras: Aides a la Navigation: Signalisation Maritime. In: Navigation. Revue Technique de Navigation Maritime, Aerienne, Spatiale et Terrestre. No. 149, Janvier 1990.
- W. Schlüter: Zeit und Frequenz in Meßverfahren der Geodäsie. Deutsche Geodätische Kommission (DKG) , Reihe C, Heft Nr. 337, Frankfurt 1988.
- H. Schüssler: Radar Altimetry for Land Applications. In: K. Linkwitz, U. Hangleiter (Eds.): High Precision Navigation. Integration of Navigational and Geodetic Methods. Heidelberg 1989.
- Service Argos: Argos user´s guide. Centre National d´Etudes Spatiale. Toulouse, France 1978.
- T.A. Stansell: The Transit Navigation Satellite System. Magnavox Government and Industrial Electronics company, Torrence, Ca, U.S.A. 1978.
- Thompson, B.D.: Everyman´s Guide to Satellite Navigation, ARINC, Research Corp. 1985.
- Wells, David et al.: Guide to GPS Positioning, Fredericton, N.B., Canada 1986.

Quartz, Crystals and Products

QCP-01 Review Papers

Chairman: C.K. Hruska

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session QCP-01 Number 1

INNOVATIVE FREQUENCY STANDARDS : PIEZOELECTRIC RESONATORS AND OTHERS

R.J. BESSON

ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

Nowadays standards, including piezoelectric oscillators, tend to be considered as "black boxes" just delivering frequency. As a consequence, it is not always clear where recent advances come from. In fact they are due to progress in very different fields whose connection is eventually ignored. Resonator design theory and processing, material improvement and supply, oscillator design theory and achievement, are the basic needed fields but improvements can also be obtained from technologies as hermetically sealing or s.m.c. electronics. Anyway, innovation is still badly needed to match the most recent requierements as very low environmental sensitivity.

In this paper, evolution of various concepts in different areas especially resonator design and connected fields is first reviewed. Some recent progress are presented and discussed as monolithic structures for resonators (piezoelectric or not) and advances in material improvement. Influence on oscillators performances is then discussed especially in the case of sensitivity to environment.

In conclusion, possible future trends are presented.

INTRODUCTION

The purpose of this paper is to try and figure out some ideas on innovative frequency standards that could appear in the next five years in the small field of piezoelectric resonators. Even so, it is a very though job, since innovation often comes from the place it was not supposed to. Knowing this, we will first present a brief historical review and attempt to dig out some underlying ideas or to indicate tendancies that could be of interest in near future. We will then ask questions -all sorts of questions even queer ones- because too many problems are supposed to be definitively solved. First target is resonator design and connected fields, then the proper use of resonators will be rapidly adressed.

I - BRIEF HISTORICAL REVIEW

Piezoelectric resonators are solids of given configuration, shape, and dimensions prepared from piezoelectric crystals under precise control of geometry and orientation. Only some vibrations are piezoelectrically driven. Electrodes provide the electric field necessary to cause the desired mechanical resonances (and others, through coupling). The history of piezoelectric resonators is already a long one since it starts in 1918. However development has been slow until world war II. Important changes have been introduced by R.A. Sykes in 1948 with use of coated units and by A.W. Warner in 1952 with use of energy trapping in contoured thickness shear crystals. Another important change has been introduction of doubly rotated SC cuts in 1975 and 1976 by Eernisse (1) but doubly rotated cuts had been somewhat proposed already in 1936 (2) : time constants may be large !

Electromechanical coupling makes piezoelectric resonators very attractive since it provides easy excitation and detection of resonances. Nevertheless, as a first approximation, piezoelectricity does not drastically change the problem of a vibrating body. Moreover, a non piezoelectric body may be very attractive if available technology provides easy excitation and detection of vibrations : technology may reverse priorities ! This is also true if we consider electrode deposition. First resonators were electrodeless : progress in the forty's was to go to electrode deposition. But in the eighty's technology improvements allowed progress by going back to electrodeless resonators.

Running through history of piezoelectric resonators and oscillators tells us several conclusions :

* time constants for change are longer than in many other domains,
* technology may reverse priorities and booste again "old" ideas,

* technology is critical and calls for processes homologation in the most demanding areas (space, defence, ...),
* last fifty years have seen at least 4 orders of magnitude improvements in stability and aging as well,
* ultimate oscillators have seemed and still seem resonator limited.

However new needs (for space applications, under bad environmental conditions ....) are extremely demanding. Improvement of piezoelectric resonator is particulary necessary for lower aging, better thermal stability, lower thermal transients and lower environmental dependence (vibration shock, acceleration, radiation, ...). Basically, improvements come from efforts in many domains (better theory, better or new piezoelectric material, better design, better processing, ...). In fact, late improvements are largely due either to better THEORETICAL UNDERSTANDING or to better consideration and achievement of BOUNDARY CONDITIONS. Both lead to new concepts and new design yielding better short term stability, aging rates, thermal characteristics, g and environmental sensitivities. Now, due to passive methods (3)(4) improvements on resonator alone can be clearly seen. Also, it may be time for still newer concepts in resonators. For somebody who is not too much impressed by the extraordinary amount of available literature on resonators, each step of the resonator design can be reconsidered. This sometimes arises questions which, at first, look almost childish. For example :

Why do crystals have usually circular or rectangular shapes and spherical contours ?

The answer is that quartz does not care but it is usually easier to deal with those shapes and contours. If correctly handled this interrogation can lead to interesting new shapes or contours. Technical feasibility or/and tradition often determines the fabrication method. In addition, it is so delicate to make excellent resonators that the slightest change can be critical. This does not encourage the use of new concepts ; it promotes the advantages of established procedures. But, theoretical understanding has been considerably improved and technology has fast evolved. **In other words, important further advances are still possible and probable.** Part of the problem is that every step of the fabrication process has to be excellent ; any failure anywhere can mask an improvement nevertheless achieved. To begin with, improvement of the material quality is slow and difficult though very impressive effort has been made in that domain ...

II - ARE THERE HOPELESS OR RIDICULOUS QUESTIONS ? DO WE REALLY ASK OURSELVES ALL POSSIBLE QUESTIONS ?

Innovation is somewhat connected to dreams and we do not dream much those days. But we badly need improvements of our standards for new applications (recently I was asked if a resonator with a g sensitivity in the 10^{-12} range was conceivable : there was definitively an application needing such a requirement).

Among possible questions let us list some ones :

1) Have we achieved the ultimate of piezoelectric frequency standards ? Are we far from it ?

2) May Q factors of quartz resonators be somewhat improved ?

3) What kind of results could be obtained with "perfect" super lattice quartz material (properly handled) ?

4) Which is the lowest g sensitivity feasible in the case of quartz Xtal resonators ?

5) Which is the best resonator frequency range ? Is it possible to use quasi electromagnetic waves instead of quasi electromechanical waves ?

6) Have we paid enough attention to possible miniaturizations and use of monolithic structures ?

7) Is energy trapping by spherical contours an optimal design ?

8) Has resonator edge to be circular or rectangular ?

9) Are nanotechnologies (in the 10 $\mathring{A}$ range) of some use in quartz resonator design ?

10) What is the future of non piezoelectric material for resonator design ?

We will now try to see how some of those questions have been handled in recent innovative attempts.

IV - SOME RECENT INNOVATIVE ATTEMPTS IN RESONATOR FIELD

Advances due to resonator theory improvements have already been discussed elsewhere **(5)** and will not be commented anymore here. However they are a very important source of innovation provided proper link is made with engineering. Any of

the further described attempt has needed a carefull theoritical analysis.

a) Obtaining low-g sensitivity : advantages of monolithic structures

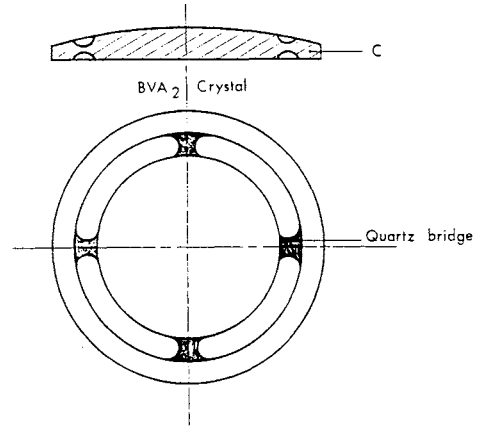


fig 1

BVA_2 resonator structures of fig. 1 are well known, they have been described elsewhere (6) (7). A few comments will however be made about the definitive advantages of the monolithic connection between vibrating and dormant part of the crystal. Properly handled, this technique allows **precision** of the location of bridges (angular position and position by respect of nodal plane). It also allows **repeatability** in machining and **symmetry** if needed. Discontinuities and stresses at fixation points are avoided. Provided the connection to the enclosure is done properly (8) it is possible to obtain maximum g sensitivity of less than 10^{-10}/g with an excellent yield. Of course the results may be obtained with unelectroded crystals (7) or with electroded ones (9) since electrodes cause less than 1 % variation in the g sensitivity.

Let us give the main parameters of a tactical resonator of the BVA_2 design with improved mounting (8) :

Nominal Frequency 9.999983 Hz
Frequency tolerance at T.O.P. : - 10 Hz + 8 Hz
Cut : doubly rotated S.C., 3rd overtone
Enclosure : HC 40 regular
$R_1 \approx 95\ \Omega$, Q factor (typical) : $1.3\ 10^6$
T.O.P. 85 - 5 + 8°C Power level : 80 µW
Aging (typical) : 10^{-11}/day or less
g sensitivity (typical) : $7\ 10^{-11}$/g max
Barometric sensitivity (typical) : $5\ 10^{-11}$/Bar

b) A thickness shear resonator with a non circular edge

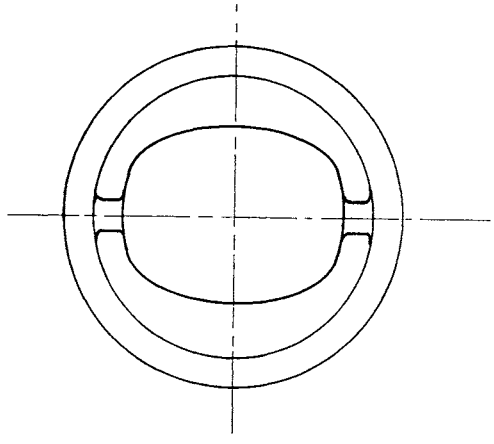


fig 2

This resonator, sometimes called P.V.A., is represented on fig. 2. In principle it is designed so that the shear compression wave (caused by the principal thickness shear) merges in phase with the principal wave after reflexion on the edge (**1 0**). The edge shape, no more circular, is computed and then achieved by ultrasonic machining. This type of resonator has been tested successfully but has not been developed yet.

c) Dual resonator configurations

Using two resonators can yield **insensitivity to accelerations**. For instance, 2 resonators of the BVA2 style (placed one on top of the other) will give reduced g sensitivity provided a proper relative orientation of axis is choosen ("bi-resonateur" French patent 7828728). Use of left and right quartz wafer should, in principle, extremely reduce sensitivity. In fact, this dual configuration is not very handy and it is usually more convenient to use a pair of a regular Xtals and look for proper compensation.

Evaluating two resonators in the same plane wafer has also recently been proposed to obtain a **new planar accelerometer**.

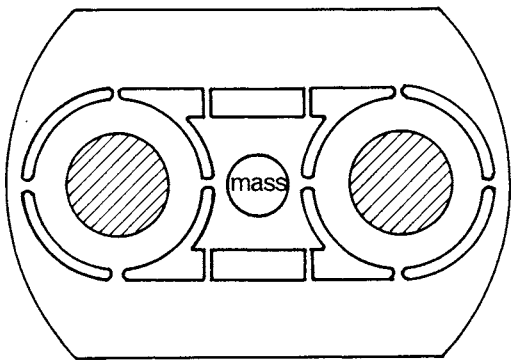
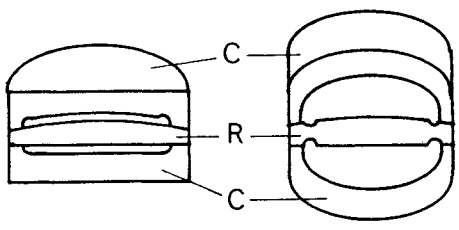


fig 3

Among other problems exhibited by acçelerometers the proper thermal behaviour is very difficult to achieve. This is partly due to the use of flexure modes which don't have third order frequency temperature characteristics (but which are more g sensitive compared to shear modes). For that reason, I was looking for a structure that would use thickness shear modes and allow a differential output. Scheme of Fig. 3 came out. This structure was very promissing because in principle the temperature behaviour of each resonator was to be the same. Also individual frequency deviations under acceleration were adding and the device could have a T.O.P. However we needed more than 5 years to obtain the device described elsewhere in this Forum (**1 1**).

d) Design of a pressure and temperature gauge

Let us start from a classical BVA_2 (**7**) of Fig. 4. By continuity it can be transformed into the resonator of Fig. 5.



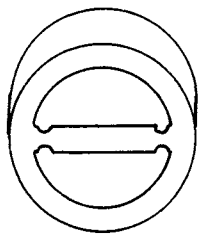


fig 4 fig 5 fig 6

In those two schemes the resonator R is surrounded by condensers C that may eventually play the part of enclosure for instance by proper use of glass fritt sealing. Then obviously the resonator can be used as a pressure sensor and as a pressure and temperature gauge if a doubly rotated is used (the C mode can be made pressure sensitive by use of proper bridges and the B mode is temperature sensitive). The second step of innovation consists in replacing the glass fritt seal by monolithic junction of Fig. 6 using ultrasonic machining. But then space is needed for machining. This leads to the device of Fig. 7 which is a quartz cylinder ultrasonically machined to leave a resonator inside. Final step is the use of two end caps sealed on the main cylinder. The proper cut and shape of cylinder together with orientation and contour of the vibrating plate then provides a pressure and temperature gauge (**1 2**).

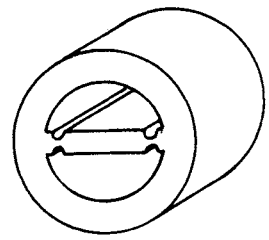


fig 7

e) Non piezoelectric electromechanical resonators

Piezoelectricity is very convenient because it provides an easy way to excitation and detection of mechanical vibrations of a given Q factor. But non piezoelectric material (crystalline or amorphous) may prove better Q and ultimately better $\sigma_y(\tau)$ for the oscillator built with such a resonator. Then several authors are using non piezoelectric resonators either via electrostrictive excitation or via accessory piezoelectric resonators (for instance made of piezoelectric material deposited onto the main device). See for example **(1 3) (1 4) (1 5)**. Optical or thermal excitation of vibrations may also be used **(1 6)**. **Several materials** may be used ranging **from pure fused silica to cristalline sapphire (1 7)**.

V - SHOULD A SUPERLATTICE QUARTZ MATERIAL BE PROPERLY TRIED INTO A RESONATOR ?

The intrinsic Q factor of a quartz resonator is given by the well known formula :

$Q_i = \frac{\overline{C}}{\omega \eta_S}$ where $\overline{C}$ is the elastic coefficient for the cut, η_S the viscosity and $\omega = 2\pi f$.

In fact, the previous formula assumes a perfect resonator and Q_i represents an upper limit since the actual Q factor Q_a is given by

$$\frac{1}{Q_a} = \frac{1}{Q_i} + \frac{1}{Q_1} + \frac{1}{Q_2} + \ldots\ldots + \frac{1}{Q_n}$$

where the Q_n each correspond to a particular phenomenon increasing damping (acoustic losses through fixations, mass loading, losses in the surfaces layers, losses due to imperfections, dislocations, impurities and so on .....).

Consequences are as follows :

- The actual Q_a factor depends on the **material's quality.**
- The material has to be "**properly**" used "properly" meaning Q_n as high as possible.

a) Material's quality

An incredible amount of effort has been accomplished recently to obtain high purity "dislocation free" quartz material, after drastic improvements were shown possible **(1 8) (1 9)**. Recent improvements in resonator understanding and technique, radiation hardening requirements have strongly boosted research and development for "better" piezoelectric material. "Better" may also mean better availability and continuity in the supply of a given quality of material.

However :

- there is a recent tendancy in slowing down the effort on quartz material's quality because the yield of previous efforts is not considered large enough. Also, the total business of quartz industry, worldwide is small compared to silicon industry thus leading to a smaller effort compared to the effort in silicon industry which has yielded exceptionnally good material.

- more efforts are probably needed to overcome present limitations.

In a recent paper, Laudise and Barns have reviewed recent advances dealing with perfection of cultured quartz and its

connection to growth processes (2 0) "Acoustic losses are most likely due to small (a few hundred molecules) number of aggregates of H_2O, allthough lesser quantities of OH, which charge compensate Al^{+++} and Fe^{+++} at Si^{++++} sites are also present" "Dislocation free material can be grown by seed selection and procedures that eliminate inclusions".

It seems that cultured high purity "dislocation free" quartz with lower acoustic losses is potentially in our hands. Under those conditions, is the "old" limit of Q factor (i.e. 3 10^6 for 5 MHz AT cut units) still valid ? Everyone of us has occasionnally already overcomed this limit obtaining 3.2 10^6 or even more but every quartz man knows that the better the material the most difficult the proper technique to take advantage of it.

The potential use of quartz resonators at liquid Helium temperature should not be forgotten. Smagin (2 3) has claimed Q factors corresponding to 8.4 10^8 at 5 MHz. Experiments in our laboratory have been quite disappointing yielding much smaller values. Obviously the quartz mounting was creating stresses on the crystal at helium temperature. Some further experiments using tactical BVA could be of interest.

Let us go now to the acceleration sensitivity problem. The fact that g sensitivity depends upon mode shape and location (2 1) which themselves are not uniform from resonator to resonator or even depends upon temperature (2 2) may mean, among other possibilities, that various imperfections of Xtal interfere in g sensitivity through variation of critical parameters.

b) Proper use of "superlattice" quartz material for quartz resonator evaluation

Not only it seems very cleat that the resonator should be "electrodeless" (6) (7) but also probably asymptotic methods of surface preparation should be used (2 3) (2 4). This should include all precautions to prevent sawing, lapping and polishing from causing a pertubed layer close to the surface. We have reasons to believe that quartz lapping has to be partially reconsidered (2 5) and that moreover quartz crystal nanotechnologies (in the 10^{-9} m range) have to be promoted. A 2 nm Rz (2 10^{-9} **m peak to valley max**) surface of crystal quartz has already been achieved on a one inch diameter sample using a specially dressed diamond wheel in an ultra stiff machine (2 6). A recent study indicates that benefits from shear mode grinding (2 7) can also be obtained.

The resonator should also be monolithic (i.e. connections from dormant to vibrating part should be made out of quartz i.e. B.V.A. style manufacturing should be used) and all precautions possible should avoid any stress in material and at the fixation points (for instance by annealing).

VI - IMPROVEMENT IN OSCILLATOR DESIGN THEORY AND ACHIEVEMENT

One of the problems already pointed out (2 8) is that there is not enough connection between the three trades involved in quartz oscillators achievement (i.e. material field, resonator field, electronics). **Matching crystal to oscillator** (or the reverse) is a very important task for success. Also oscillator design now benefits from computer aided design and various improvements **(29)** including surface mounted components use and several others. Of course influences **humidity** have to be **eliminated** (3 0) through hermetically sealing (this also helps'eliminating pressure variations influence).

VII - CONCLUSION : POSSIBLE FUTURE TRENDS (NEXT 10 YEARS)

If we try to summarize all the data obtained in the last 20 years, it is clear that 2 characteristics, aging and g sensitivity of oscillators, have been in constant progress while $\sigma_y(\tau)$ seemed to somehow reach a limit in the vicinity of 7 $10^{-14}/10$ s at 5 MHz. It is hard to predict what would happen with much better Q factors because it may depend on the type of Xtal (there is some evidence that part of the noise is correlated to processes in adhering electrodes) and obviously we still have to clear up various influences on noise processes (though in principle $\sigma_y(\tau)$ should vary as $1/Q^2$).

Some more confidence can be paid to improving aging and g sensitivity. An aging rate in the of 5 10^{-11}/year to 5 10^{-10}/year is probably in a reaching distance while a g sensitivity in the 10^{-11}/g range or better is probably achievable. Of course, such a low g sensitivity will also influence other parameters since an oscillator is never in a perfect mechanical environment. In other words we should see some more of the competition between high quality quartz standards and atomic rubidium references.

REFERENCES

(1) E.P. EERNISSE, Proc. 29th AFCS, pp 1-4 (1975) and Proc. 30 th AFCS, pp 8-11 (1976)
(2) C.F. BALDWIN, S.A. BOKOVOY, "Piezoelectric quartz element", US Patent, 2212130 (1936)
(3) A.E. WAINRIGHT, F.L. WALLS and al, Proc. 28th AFCS, pp 117-180 (1974)
(4) S.R. STEIN, C.M. MANNEY Jr and al, Proc. 32nd, AFCS, pp 527-530 (1978)
(5) R.J. BESSON, 1rst European Frequency and Time Forum, pp 20-30 (1987)
(6) R.J. BESSON, Proc. 30th AFCS, pp 78-82 (1976) and Proc. 31st AFCS, pp 147-152 (1977)
(7) R.J. BESSON, French patents n° 7601035, 7802261, 7828728
(8) R.J. BESSON and al, French patents n° 8509097 and 8814197 other patents pending
(9) R.J. BESSON, French patent n° 7717309
(1 0) R.J. BESSON, French patente n° 7802261
(1 1) R. BOURQIN, B. DULMET, R. BESSON, these proceedings
(1 2) R. BESSON, J.J. BOY, J.P. VALENTIN, French patent n° 8313745 and M. VALDOIS, P. MAITRE, R.J. BESSON, J.J. BOY, French patent n° 8313746
(1 3) D.V. SCHICK, D.S. STEVENS, H.F. TIERSTEN, J. Appl. Phys. 60, oct. 1984
(1 4) K.M. LAKIN, J.S. WANG, A.R. LANDIN, Proc. 36th AFCS (1982)
(1 5) K.M. LAKIN, IEEE Ultrasonics Symposium, 1983
(1 6) E. DIEULESAINT, D. ROYER, Proc. Ultrasonics Symposium, pp 793-795 (1982) and D. ROYER, E. DIEULESAINT, J. de Phys. C_6 suppl. n° 10, Tome 44 (1983)
(1 7) E. DIEULESAINT, D. MAZEROLLE, D. ROYER, Proc. 1rst EFTF, pp 36-38 (1987)
(1 8) D.F. CROXALL, IRA CHRISTIE and al, 36th AFCS, pp 62-65 (1982)
(1 9) A.F. ARMINGTON, J.F. BALASCIO, 38th AFCS, pp 3-9 (1984), 39th AFCS, pp 230-233 (1985) and 40th AFCS, pp 70-75 (1986)
(2 0) R.A. LAUDISE and R.L. BARNS, IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, vol 35, n° 3, pp 277-287, may 1988
(2 1) E.P. EERNISSE, R.W. WARD and al, Proc. 43rd AFCS, pp 388-395 (1989)
(2 2) W.P. HANSON, T.E. WICKARD, Proc. 43rd AFCS, pp 427-432 (1989)
(2 3) A.G. SMAGIN, Pribory i Tekhnika Eksperimenta, n° 6, pp 143-145 (1974)
(2 4) A.G. SMAGIN, Kristallografia, Vol 4, n° 6, pp 862-866 (1959)
(2 5) J.D.B. DE MELLO, Uberlandia University (Brazil), currently in our laboratory. Private communication
(2 6) YOSHIOKA and al., Ultraprecision grinding technology for brittle materials : application to surface and centerless grinding processes, A.S.M.E. Milton shaw grinding Symposium, PED, Vol 16, 1986
(2 7) N.J. BROWN, B.A. FUCHS, 43rd AFCS, pp 606-609 (1989)
(2 8) R.J. BESSON, Influence du matériau sur les progrès récents des résonateurs et oscillateurs, Journées d'Etudes DRET, SDRG3, October 1986
(2 9) R.J. BESSON, M. MOUREY, 44th AFCS (1990)
(3 0) F.L. WALLS, 42nd AFCS, pp 279-283 (1988).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session QCP-01 Number 2

Current Developments in Crystal Filters

Robert C. Smythe
Piezo Technology Inc.
Orlando, Florida, U.S.A.

Introduction

Applications for crystal filters have significantly expanded during recent years. Higher system operating frequencies and the increased use of digital transmission as well as higher data rates and the use of spread spectrum transmission are a common thread in a number of applications such as cellular radio telephones and GPS (NAVSAT global positioning systems) receivers for position-locating and navigation applications, and call for filters with higher center frequencies, wider bandwidths, and improved group delay combined with good selectivity.

This paper surveys recent developments and current trends in monolithic and discrete-resonator crystal filters in response to these needs. Of particular note are the continued development and expanded use of VHF monolithic and discrete-resonator fundamental mode filters and the more extensive use of multiresonator monolithic devices, particularly in the VHF range. Other developments include monolithic and discrete-resonator SC-cut crystal filters, allowing operation at higher signal levels than AT-cut filters, and filters having extremely good amplitude linearity at low signal levels. The use of lithium tantalate to achieve wider bandwidths than are possible with quartz has continued to develop and has been extended to center frequencies as low as 2 MHz. Filter usage of berlinite is being explored. Another candidate material is lithium tetraborate.

Increasing levels of integration and miniaturization across virtually the entire electronics domain are influencing crystal filter requirements in two ways. First, in many new and projected receiving systems a portion of the required selectivity is achieved in integrated circuits, such as switched-capacitor filters. In these systems, the role of the crystal filter is to provide first i-f selectivity. Additional selectivity is then provided in integrated form at a much lower second i-f frequency. Second, as a result of integration of other components, the crystal filter is often one of the largest system r-f components. Size reduction is being achieved by a variety of means including fundamental mode devices in place of overtones, monolithic dual resonators (MDR's) instead of discrete resonators, multiresonator monolithics instead of MDR's, component size reduction, and hybridization, but further reductions will continue to be needed. Increasingly, there is a need for filters which are not only miniaturized but also surface-mountable.

VHF Applications

Since very early days, an important trend in crystal filter development has been toward the use of higher center frequencies. The stimuli are higher system operating frequencies and the need for wider bandwidths.

Important examples are current and planned cellular radiotelephones in the 800-900 MHz band. Operation of the U.S. system began around 1984, followed closely by a quite similar British system. Voice transmission is analog FM, with digital signalling; although channel allocation is accomplished automatically by the system, each channel accommodates a single user. These systems require about a 30 kHz channel bandwidth; group delay requirements are modest. These requirements can be met using either fundamental mode MDR's, typically at 45 MHz, or third-overtone MDR's, typically at 90 MHz [SM83]. The former offer the highest attenuation and the lowest impedance and are inductorless, while the latter are relatively high impedance and have somewhat less ultimate attenuation, but may allow a simpler front-end filter.

Currently GSM, an advanced European cellular system is under development. The system uses time-division multiplexing with multiple users per channel. It requires a bandwidth of 160 - 200 kHz with low group delay variation -- 500 to 600 nanoseconds is desired in some instances -- and fairly good selectivity. When combined with the need for low cost and high production volumes, these requirements stretch the present state of the art considerably, and may eventually have to be compromised somewhat in order to keep the cost sufficiently low.

Aubry and colleagues have analyzed two possible configurations, both employing four resonators on a single wafer [AU89]. By using either lithium tantalate or berlinite and making a judicious choice of center frequency, inductorless filters can be realized.

With lithium tantalate, the four resonators are acoustically isolated from one another and a ladder circuit is used. Center frequency is around 45 MHz. Experimental results are favorable. For the second configuration, berlinite is proposed; it would consist of two MDR's on a single ion-milled wafer. Center frequency is around 70 MHz, using the fundamental mode. Either configuration could be built in a low-profile TO-8 enclosure.

A more conventional approach uses ring-supported fundamental mode AT-cut quartz resonators in a tandem lattice arrangement. Figure 1 shows passband and stopband response of a four-pole 70 MHz design. A transitional Butterworth-Gaussian response provides a delay variation of less than 600 ns.; 3 dB bandwidth is ±125 kHz with 55 dB attenuation at ±650 kHz. For such small delay variations, component tolerances are extremely tight.

The ring-supported resonator, whose development was pioneered by Berté [BE77a, BE77b], has been the most important development for the expansion of VHF filter capabilities. Berté used argon ion-beam milling to reduce the thickness of the central portion of a circular wafer, figure 2. An advantage of this process is that it is easily adapted, not only to different cuts of quartz but also to different materials. In particular, it is useful for lithium tantalate, which is extremely difficult to etch chemically. Using ion-beam milled resonators, quartz and lithium tantalate multiresonator filters up to 240 MHz have been demonstrated [DA82].

Reactive ion-beam milling [WA84] of ring-supported resonators is also feasible, as are chemical milling [HA83, KA84, HU85, SM85b, SM88b] and plasma etching [ST89]. Currently, chemical milling is the most widely practiced of these. Using fundamental mode chemically-milled ring-supported quartz resonators, bandwidths up to 1% and more have been achieved in the lower portion of the VHF range. Figure 3 shows the response of a three-pole, single-section filter having a bandwidth of slightly more than 1 MHz at 70 MHz, with good unwanted mode suppression and ultimate attenuation.

The availability of fundamental mode devices has also made higher-order filters practical. Figure 4 presents the response of a six-pole, 125 MHz tandem monolithic filter, and having over 100 dB ultimate attenuation [SM85b], while figure 5 shows an eight-pole, 70 MHz filter with a bandwidth of 330 kHz. Both were made using chemical milling.

Figure 6 shows, quite approximately, the current frequency range of fundamental mode and third overtone AT- and SC-cut quartz resonators and ACR's. Dotted lines indicate regions that have been demonstrated on a laboratory basis but not in production. Not included are: 1. low-frequency quartz resonators, which are useful at frequencies down to 10 kHz in certain cases; 2. BT-cut resonators and monolithics, which are in limited use, primarily to extend the fundamental frequency range; 3. lithium tantalate resonators, which are used to achieve wide bandwidths and which are discussed later; 4. resonators or monolithics made with berlinite and lithium tetraborate, which are also discussed later.

Multiresonator ACR's

With the advent of monolithic crystal filters in the late 60's and early 70's, the monolithic dual resonator (MDR) became a common and convenient building block for at least two reasons. First, most filter networks can be subdivided into symmetrical two-pole sections. Second, the MDR is much easier to fabricate than a higher-order acoustically-coupled resonator (ACR); moreover, two MDR's in tandem usually offer better spur performance and higher ultimate attenuation than a single four-pole ACR. Nevertheless, in certain applications, multiresonator ACR's are advantageous for one or more of the following three reasons:

1. Inductorless construction. A multiresonator ACR may allow an inductorless realization which would not be possible with separate MDR's. This most commonly occurs for overtone devices. An excellent recent example is the VHF front-end filter developed by Dworsky and used in an advanced paging receiver [DW85]. Electrode and mounting arrangements and the attenuation characteristic of this fifth-overtone, 175 MHz, four-pole filter are shown in figure 7. In this filter, all four resonators are acoustically coupled. The package is a low-profile TO-8. Another example, figure 8, is a 168 MHz, fifth overtone four-pole with a bandwidth of 25 kHz in an 11.4 mm high HC-18-style package [SM89]. The linear electrode layout shown permits the two-pole sections to be acoustically isolated for improved unwanted mode attenuation and connected in tandem without the use of an inductor [SM82].

2. Miniaturization. A multiresonator ACR allows a size reduction. If, at the same time, it eliminates the need for inductors, as in the preceding paragraph, the decrease in size is dramatic. Figure 9 plots the attenuation of a four-pole ACR having a bandwidth of 3.5 kHz at 21.5 MHz. The design of this filter was motivated primarily by the need for a very small size. It consists of two MDR's on a single 11 mm diameter, third-overtone blank, which is mounted in a custom-tooled "slim HC-12" holder measuring 14.4 x 4.7 x 14.6 mm, which also contains the junction capacitor. To achieve this size, it was necessary to locate the two-poles side-by-side, as in the surface-mounted filter shown later in figure 12a, rather than using

the preferred in-line arrangement. While stopband performance is still good, some coupling can be seen via whole-plate modes in the vicinity of 200 kHz above center frequency.

3. Cost reduction. Cost reduction may result from the use of multiresonator ACR's, although this is not always the case. While the number of devices is reduced, their complexity is increased; hence, a savings in device cost occurs only if production methods are adequate. This is not always possible for filters built in small quantities. If inductors are eliminated, a cost saving more frequently results. Not only is the coil a cost item, its tuning is critical, adding substantially to manufacturing labor costs.

Piezoelectric Materials

Quartz crystal filters are essentially narrow-band networks. The combination of high Q (typically $1x10^4$ to $1x10^6$), high stability, small size, and low aging found in quartz bulk-wave resonators is unequalled in any other electrical technology. While these properties are essential for achieving narrow bandwidths, the modest piezoelectric coupling of alpha quartz places severe limitations on the maximum filter bandwidth. Typical quartz crystal filter bandwidths are less than a few tenths of one percent of center frequency. By incorporating inductors into the filter network, maximum bandwidths of several percent can be achieved [KI87, SM89].

A more attractive way of achieving wider bandwidths is by using a material that has a larger piezoelectric effect than quartz, while still possessing adequate Q and stability. At present, the only suitable commercially available material is lithium tantalate, which is in limited use. However, berlinite and lithium tetraborate are under development and show promise. (Lithium niobate, commonly used for wide-band surface acoustic wave filters possesses no temperature-compensated bulk-wave orientations.)

VHF filters with ion-milled, ring-supported resonators of both quartz and lithium tantalate have been described by d'Albaret and Siffert [DA82], by d'Albaret and Wolcoff [DA85], and by Brauge, Fragneau, and Aubry [BR85]. Bandwidths up to 4% have been achieved using ion-milled fundamental mode resonators of X-cut lithium tantalate.

Lefèvre and co-workers have used laser machining to fabricate extensional mode single resonators and coupled pairs, both with integral suspensions, at frequencies from 150 kHz to 4.2 MHz [LE85, LE87]. High Q's were obtained -- for example, 50×10^3 to 80×10^3 at 2 MHz. High piezoelectric coupling, typically 25%, allows bandwidths up to several percent to be achieved without using inductors. For example, at 455 kHz, a six-pole filter having a bandwidth of 16 kHz, or 3.5%, was realized using three dual-resonator devices.

Vale and Weinert [VA86] achieved a bandwidth of 3 kHz at 2 MHz using four lithium tantalate X-cut strip resonators [FU83] in a crystal-capacitor ladder configuration. The design uses a .05° ripple, linear phase transfer function [ZV67] to provide an impulse response with less than -28 dB ringing. Unwanted modes were attenuated by 84 dB. Resonator Q's of 80×10^3 were obtained. The capacitance ratio of 7.5 indicates the possibility of much wider bandwidths.

Quite recently, Weinert and co-workers [WE89] have demonstrated high Q's and single-mode coupling in -16.5° rotated Y-cut lithium tantalate resonators with lateral-field excitation. While Weinert's objective is high overtone resonators for oscillator applications, the work demonstrates the high Q-f product obtainable. At lower overtones, impedance levels useful for filters can be inferred from the work. Experimental frequency-temperature measurements were not reported, but the calculated first-order temperature coefficient of frequency is zero [BA86].

Berlinite (aluminum metaphosphate, $AlPO_4$) is a very promising material for future filter applications [DE85, DE86, HO87, CA88, DE88, PH89]. It possesses an orientation having a cubic f-T characteristic and about twice the coupling of AT-cut quartz. Resonators, filters, and monolithic dual resonators have been demonstrated.

Another candidate material is lithium tetraborate ($Li_2B_4O_7$). Although filters using lithium tetraborate have not, to the writer's knowledge, been reported, the material promises to allow about one order of magnitude greater inductorless bandwidth than AT-cut quartz [EM83] and a parabolic frequency-temperature characteristic with turnover near 25° C and a second-order f-T coefficient of -0.26 ppm/°C^2 [FU85]. Experimental resonators have been fabricated by Peach, et al [PE83] and Fujiwara, et al [FU85]. Ballato, et al have calculated lateral-field and thickness-field coupling [BA89a] and frequency-temperature behavior [BA89a, BA89b] for this material.

Nonlinearity

The most important nonlinear filter characteristics are intermodulation distortion [HO73, SM74, TI75, HO85], variation of attenuation with input signal level [SM83, SM89],

variation of frequency with drive level [HO85], and excess phase noise [SM85d]. At low signal levels, it is now well-known that these effects can be traced to resonator surface conditions, including contamination. At higher levels, elastic nonlinearity [TI75] dominates. In this range, the SC-cut has been found to have about a 6 dB IM advantage over the AT-cut [HO85].

Figure 10 shows the variation of mid-band insertion loss with input level for a four-pole, 20 MHz monolithic filter used in a high-performance network analyzer. For this application it is necessary to hold the loss variation to ±0.005 dB over a 40 dB input range. Low-level mechanisms are the dominant cause of variation; rigid attention to cleanliness is required to achieve the specified performance.

At higher levels, elastic nonlinearities cause both loss variations and frequency shift. Figure 11 compares the loss versus input power for AT, BT, and SC 25 MHz fundamental mode MDR's for input levels from 0 to +25 dBm [HO85]. As would be expected, both the BT and SC units show considerably less frequency shift than does the AT unit. In addition, the latter shows strong nonlinear mode coupling at the highest input level. The improved power handling of the BT and SC devices must be weighed against their less favorable f-T characteristics, impedance level, and capacitance ratio.

Surface-Mount Packaging

Currently there is considerable pressure to develop surface-mountable filters which are compatible with other surface-mounted components. While almost any filter package can be re-designed to be, in a sense, surface-mountable by measures such as replacing through-hole pins with gull-wing or J-lead pins, and by replacing studs with solderable mounting tabs, such measures are not always adequate, since they reduce neither the height nor the volume of the package.

Achieving height and volume reductions is by no means a minor task. Trends include increased use of MDR's and multiresonator devices, as already noted, use of chip components, and all the previously-mentioned techniques for eliminating inductors. Leadless crystal packages are starting to play a role. Though they are not particularly volume-efficient, they allow very low profiles. An example, figure 12, has a height of only 1.65 mm, with length and width of 11.4 mm. The version shown contains a single-wafer four-pole ACR composed of two, two-pole sections [SM88a]. Another version [HO90] houses two separate 45 MHz MDR's. This unit is used in the surface-mount filter shown in figure 13, whose overall height is 5.6 mm.

Conclusion

Crystal filter technology continues to evolve in response to system requirements. Many filters which were state-of-the-art 10 to 20 years ago are market commodities now. At the same time, the maturing of ring-supported resonator and MDR technology is leading to a greatly expanded role for VHF filters, especially those using the fundamental mode. Filters which were previously realizable, if at all, by using spreading coils are now being realized as inductorless designs. The result is not only improved performance but also reduced size and cost. Further miniaturization is being made possible by increasing use of multiresonator monolithics, usually four-pole devices. Finally, surface-mount packaging techniques are actively being pursued. Further advances in this area are still needed. Also needed are improvements in electrode pattern definition. Conventional aperture plating masks limit the minimum size and dimensional accuracy of electrode features, restricting the practical range of designs at higher frequencies. One solution is photolithography [SH85].

Bibliography

The following bibliography includes, besides the works cited in the text, a number of pertinent recent references; however, no attempt has been made at completeness. A fairly extensive bibliography, covering roughly the period 1968-83, but including a number of significant earlier references as well, can be found in Gerber and Ballato [GE85].

[AU89] J.-P. Aubry, et al, "A Strategy for a Low Cost IF Filter for Digital Radiotelephone," Proc. 3rd European Frequency and Time Forum, pp. 387-393, 1989.

[BA86] A. Ballato, et al, "Lateral-Field Coupling of Rotated BAW Plates with 3m, 4mm, and 43m Symmetries," Proc. IEEE Ultrasonics Symp., pp. 339-342, 1986.

[BA89a] A. Ballato et al, "Lateral- and Thickness-Field Coupling in Lithium Tetraborate," Proc. 43rd Annual Symposium on Frequency Control, pp. 557-562, 1989.

[BA89b] A. Ballato, J. Kosinski, and T. Lukaszek, "Lateral-Field Temperature Behavior of Lithium Tetraborate," Proc. IEEE Ultrasonics Symp., pp. 441-444, 1989.

[BE77a] M. Berté, "Acoustic Bulk Wave Resonators and Filters Operating in the Fundamental Mode at Frequencies Greater than 100 MHz," Electronics Letters, v.13, pp. 248-250, 1977.

[BE77b] M. Berté, "Acoustic Bulk Wave Resonators and Filters Operating in the Fundamental Mode at Frequencies Greater than 100 MHz," Proc. 31st Annual Symposium on Frequency Control, pp. 122-125, 1977.

[BR85] R.J. Brandmayr and J.R. Vig, "Chemical Polishing in Etching Solutions that Contain Surfactants," Proc. 39th Annual Symposium on Frequency Control, pp. 276-281, 1985.

[BR85] J. Brauge, M. Fragneau, & J-P. Aubry, "Monolithic Filters Using Ion Etched Fundamental Mode Resonators Between 60 and 100 MHz," Proc. 39th Annual Symposium on Frequency Control, pp. 504-513, 1985.

[BR86] R.J. Brandmayr and J.R. Vig, "Further Results on the Use of Surfactants in Chemically Polishing Quartz Crystals," Proc. 40th Annual Symposium on Frequency Control, pp. 86-90, 1986.

[CA75] R.N. Castellano, "A Survey of Ion Beam Milling Techniques for Piezoelectric Device Fabrication," Proc. 29th Annual Symposium on Frequency Control, pp. 128-134, 1975.

[CA88] H. Carru & R. Lefévre, "Semi-Analytical Analysis of Coupled Thickness-Shear and Thickness-Twist in Monolithic Crystal Filters," Proc. 42nd Annual Symposium on Frequency Control, pp. 101-105, 1988.

[CL88] R.L. Clark & M.K. Yurtseven, "Spurious Signals Induced by Vibration of Crystal Filters," Proc. IEEE Ultrasonics Symp., pp. 365-368, 1988.

[DA82] B. d'Albaret and P. Siffert, "Recent Advances in UHF Crystal Filters," Proc. 36th Annual Symposium on Frequency Control, pp. 405-418, 1982.

[DA87] B. d'Albaret & B. Wolcoff, "La Montée en Fréquence des Filtres Piézoélectrique a Ondes de Volume," Proc. 1er Forum Européen Temps-Fréquence, pp. 129-135, 1987.

[DE85] J. Detaint et al, "Crystal Growth, Physical Characterization and BAW Devices Applications of Berlinite," Proc. 39th Annual Symposium on Frequency Control, pp. 234-246, 1985.

[DE86] J. Detaint et al, "Berlinite: Characterization of Crystals with a Low Water Concentration and Design of Bulk Wave Resonators," Proc. 40th Annual Symposium on Frequency Control, pp. 101-114, 1986.

[DE88] J. Detaint, et al, "Evaluation and Applications of High Purity Berlinite Crystals," Proc. 2nd European Frequency and Time Forum, pp. 873-889, 1988.

[DW85] L. N. Dworsky and C. S. Shanley, "The Motorola Multi-Pole Monolithic Filter Project," Proc. 39th Annual Symposium on Frequency Control, 1985, pp. 486-490.

[EM83] C.D.J. Emin & J.H.F. Werner, "The Bulk Acoustic Wave Properties of Lithium Tetraborate," Proc. 37th Annual Symposium on Frequency Control, pp. 136-143, 1983.

[FU83] Y. Fujiwara, S. Yamada, & N. Wakatsuki, "Miniaturized $LiTaO_3$ Strip Resonators," Proc. 37th Annual Symposium on Frequency Control, pp. 343-348, 1983.

[FU85] T. Fujiwara, M. Ono, M. Sakai, & Noboru Wakatsuki, "Strip Type Resonator of Lithium Tetraborate," Proc. 39th Annual Symposium on Frequency Control, pp. 351-355, 1985.

[GE85] E.A. Gerber and A.D. Ballato, Eds., "Precision Frequency Control, Vol. I: Acoustic Resonators and Filters," Academic Press, Orlando, 1985.

[GU72] G.K. Guttwein, A.D. Ballato, and T.J. Lukaszek, "VHF-UHF Piezoelectric Resonators," U. S. Pat. 3,694,677, 26 September, 1972.

[HA83] W.P. Hanson, "Chemically Polished High Frequency Resonators, Proc. 37th Annual Symposium on Frequency Control, pp. 261-264, 1983.

[HE78] P.A. Herzig and T.W. Swanson, "A Polylithic Crystal Filter Employing a Rhodes Transfer Function," Proc. 32nd Annual Symposium on Frequency Control, pp. 233-242, 1978.

[HO73] W.H. Horton and R.C. Smythe, "Experimental Investigations of Intermodulation in Monolithic Crystal Filters," Proc. 27th Annual Symposium on Frequency Control, pp. 243-245, 1973.

[HO85] M.D. Howard, R.C. Smythe and P.E. Morley,"Monolithic Crystal Filters Having Improved Intermodulation and Power Handling Capability," Proc. 39th Annual Symposium on Frequency Control, pp. 491-503, 1985.

[HO87] J. Hou & B.H.T. Chai, "BAW and SAW Devices on Berlinite with Low Water Content," Proc. IEEE Ultrasonics Symp., pp. 419-424, 1987.

[HO90] M.D. Howard and R.C. Smythe, "Quartz Crystal Filters: a Review of Current Issues," Proc. RF Expo, March, 1990, in press.

[HU85] J.R. Hunt and R.C. Smythe, "Chemically Milled VHF and UHF Resonators," Proc. 39th Annual Symposium on Frequency Control, pp. 292-300, 1985.

[KA84] P.J. Kavolis and W.P. Hanson, "Filter Applications of High Frequency Chemically Polished Fundamental Mode Bulk Wave Quartz Crystal Resonators," Proc. 38th Annual Symposium on Frequency Control, pp. 251-253, 1984.

[KI87] R. G. Kinsman, "Crystal Filters: Design, Manufacture, and Application," John Wiley and Sons, New York, 1987.

[LE85] R. Lefèvre, L. Jenselme, & D. Servajean, "Laser Processed Miniature $LiTaO_3$ Resonators and Monolithic Filters," Proc. 39th Annual Symposium on Frequency Control, pp. 333-337, 1985.

[LE87] R. Lefèvre & C. Joly, "Dispositifs Miniatures en Tantalate de Lithium Découpes par Laser avec Suspension Intégrées," Proc. 1er Forum Européen Temps-Fréquence, pp. 189-194, 1987.

[MC79] D.I. McLean, A.F. Graziani, and J.J. Royer, "New Discrete Crystal Filters for Bell System Analog Channel Banks," Proc. 33rd Annual Symposium on Frequency Control, pp.166-172, 1979.

[OK83] M. Okzaki & S. Watanabe, "Miniature $LiTaO_3$ X-Cut Strip Resonator," Proc. 37th Annual Symposium on Frequency Control, pp. 337-342, 1983.

[OK88] Y. Okamoto & H. Sekimoto, "New Monolithic Crystal Filter with Wide Tabs," Proc. 42nd Annual Symposium on Frequency Control, pp. 95-100, 1988.

[PE83] R.C. Peach, C.D.J. Emin, J.F. Werner, & S.P. Doherty, "High Coupling Piezoelectric Resonators Using Lithium Tetraborate," Proc. IEEE Ultrasonics Symp., pp. 521-526, 1983.

[PH89] E. Philippot, et al, "Improvement of Berlinite's Crystalline Quality," Proc. 3rd European Frequency and Time Forum, pp. 57-60, 1989.

[SH77] D.F. Sheahan and R.A. Johnson, Eds., "Modern Crystal and Mechanical Filters," IEEE Press, New York, 1977.

[SH85] C.W. Shanley, et al, "Processing of a Five Resonator VHF Crystal Device," Proc. 39th Annual Symposium on Frequency Control, pp. 367-371, 1985.

[SM69] R.C. Smythe, "HF and VHF Inductorless Filters for Microelectronic Systems," Proc. IEEE/EIA Electronic Components Conference, pp. 115-119; 1969. Reprinted in D.F. Sheahan and R.A. Johnson, editors, Modern Crystal and Mechanical Filters, pp. 301-305, IEEE Press, New York; 1977.

[SM72] R.C. Smythe, "Communications Systems Benefit from Monolithic Crystal Filters," Electronics, v. 45, no. 3, 31 Jan., 1972, pp. 48-51.

[SM74] R.C. Smythe, "Intermodulation in Thickness-Shear Resonators," Proc. 28th Annual Symposium on Frequency Control, pp.5-7, 1974.

[SM79] R.C. Smythe, "Some Recent Advances in Integrated Crystal Filters," Proc. IEEE, v. 67, no. 1, Jan., 1979, pp. 119-129.

[SM83] R.C. Smythe and M. D. Howard, "Current Trends in Crystal Filters," Proc. 37th Annual Symposium on Frequency Control, pp. 349-353, 1983.

[SM85a] R.C. Smythe, "Bulk Acoustic Wave Filters," in "Precision Frequency Control, Vol. I", (E. A. Gerber and A. D. Ballato, Eds.), Academic Press, Orlando, 1985, pp. 188-228.

[SM85b] R.C. Smythe, M.D. Howard, and J.R. Hunt, "VHF Monolithic Filters Fabricated by Chemical Milling," Proc. 39th Annual Symposium on Frequency Control, pp. 481-485, 1985.

[SM85c] R.C. Smythe and P.E. Morley, "Experimental Evaluation of the Effective Nonlinear Elastic Constant for Trapped-Energy and Contoured Resonators," Proc. 39th Annual Symposium on Frequency Control pp. 317-324, 1985.

[SM85d] R.C. Smythe, "Phase Noise in Crystal Filters," Proc. 39th Annual Symposium on Frequency Control, pp. 138-139; 1985.

[SM85e] R.C. Smythe, J.R. Hunt, and M.D. Howard, "Chemically-Milled VHF and UHF Quartz Resonators and Monolithic Filters," Proc. IEEE Ultrasonics Symposium, pp. 302-305; 1985.

[SM88a] R.C. Smythe, "An LCC Monolithic Filter," Proc. 2nd European Frequency and Time Forum, pp. 637-644; 1988.

[SM88b] R.C. Smythe and R.B. Angove, "Chemically-Milled UHF SC-Cut Resonators," Proc. 42nd Annual Symposium on Frequency Control, pp. 73-77, 1988.

[SM89] R.C. Smythe, "Crystal Filters," Ch. 7 in Miniature Filters (S.K. Mitra and C.F. Kurth, editors), Wiley-Interscience; 1989.

[ST89] F.M. Stern, J. Dowsett, D.J. Carter, & R.J. Williamson, "The Fabrication of High Frequency Fundamental Crystals by Plasma Etching," Proc. 43rd Annual Symposium on Frequency Control, pp. 634-639, 1989.

[TE73] G.C. Temes, and S.K. Mitra, Eds. "Modern Filter Theory and Design," John Wiley and Sons, New York, 1973.

[TI74] H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Thickness-Shear," Proc. 28th Annual Symposium on Frequency Control, pp. 44-48, 1974.

[TI75] H.F. Tiersten, "An Analysis of Intermodulation in Thickness-Shear and Trapped-Energy Resonators," J. Acoustical Society of America, v. 57, no. 3, pp. 667-681, March, 1975.

[TI77] H.F. Tiersten, "An Analysis of Overtone Modes in Monolithic Crystal Filters," J. Acoustical Society of America, v. 62, no. 6, pp. 1424-1430, Dec. 1977.

[TI79] H.F. Tiersten and R.C. Smythe, "An Analysis of Contoured Crystal Resonators Operating in Overtone Modes of Coupled Thickness-Shear and Thickness Twist," J. Acoustical Society of America, v. 65, no. 6, pp.1455-59, June, 1979.

[VA86] C.R. Vale & R.W. Weinert, "Ladder Filter Using Temperature Stable $LiTaO_3$ Resonators in Strip Form," Proc. IEEE Ultrasonics Symp., pp. 103-107, 1986.

[VI77] J.R. Vig, J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," Proc. 31st Annual Symposium on Frequency Control, pp. 131-143, 1977.

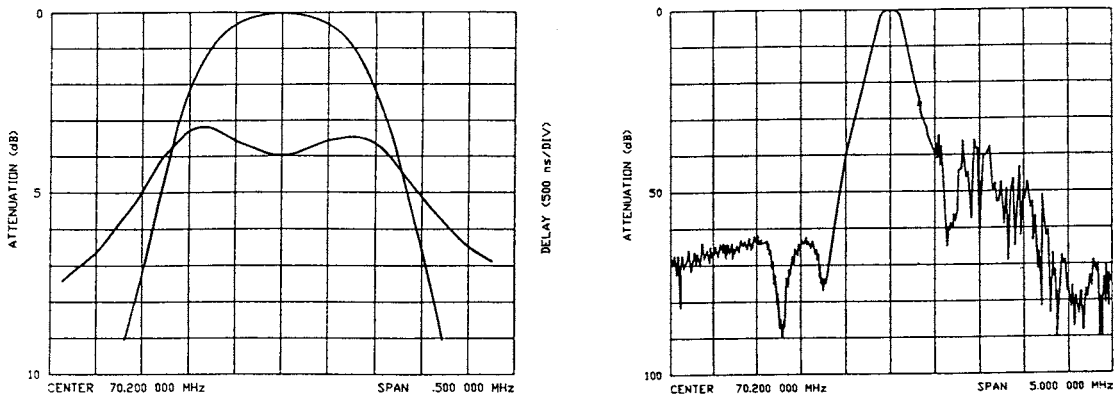
[VI79] J.R. Vig, R.J. Brandmayr, and R.L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd Annual Symposium on Frequency Control, pp. 351-358, 1979.

[VI81] J.R. Vig and R.J. Brandmayr, U. S. Pat. 4,274,907, June, 1981.

[WA84] J.S. Wang, S.K. Watson, and K.F. Lau, "Reactive Ion Beam Etching for VHF Crystal Resonators, Proc. 38th Annual Symposium on Frequency Control, pp. 101-113, 1984.

[WE89] R.W. Weinert, et al, "$LiTaO_3$ Lateral Field Resonators," Proc. IEEE Ultrasonics Symp., pp. 477-480, 1989.

[ZV67] A.I. Zverev, "Handbook of Filter Design," John Wiley and Sons, New York, 1967.



a. b.

Fig. 1. Response of four-pole, 70.2 MHz fundamental mode filter: a. Passband attenuation & group delay; b. Wideband attenuation.

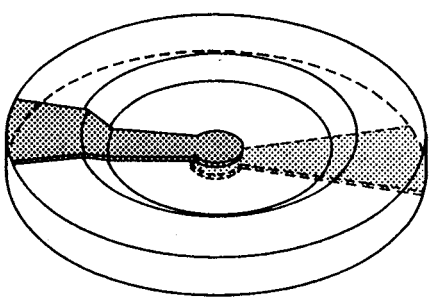


Fig. 2. Ring-supported resonator.

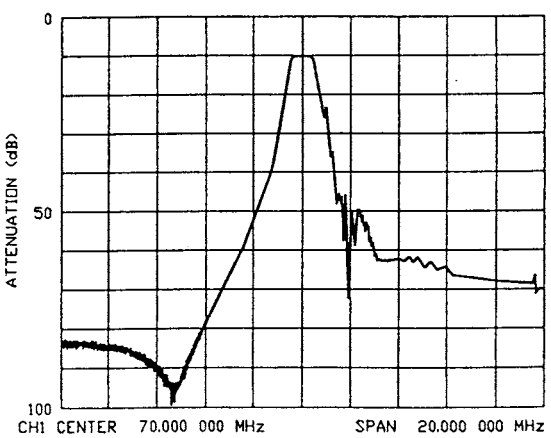


Fig. 3. Response of three-pole, 70MHz, fundamental mode filter; 3 dB bandwidth is 1 MHz.

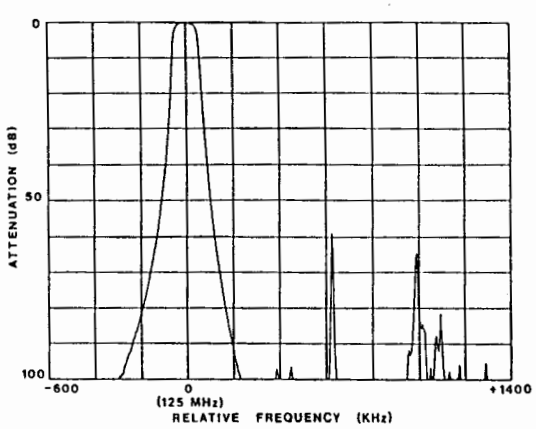


Fig. 4. Response of six-pole, 125 MHz, fundamental mode tandem monolithic filter; 3 dB bandwidth is 100 kHz.

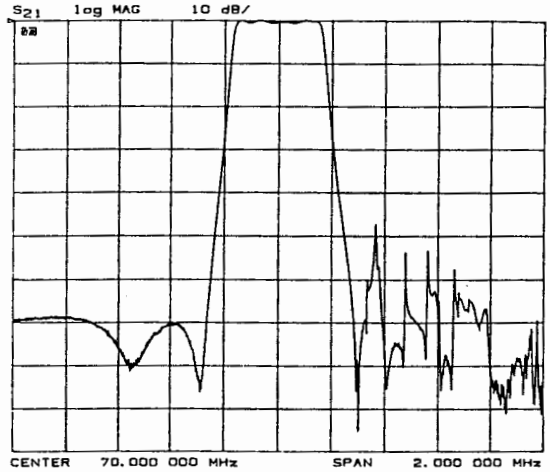


Fig. 5. Response of eight-pole, 70 MHz fundamental mode filter; 3 dB bandwidth is 330 kHz.

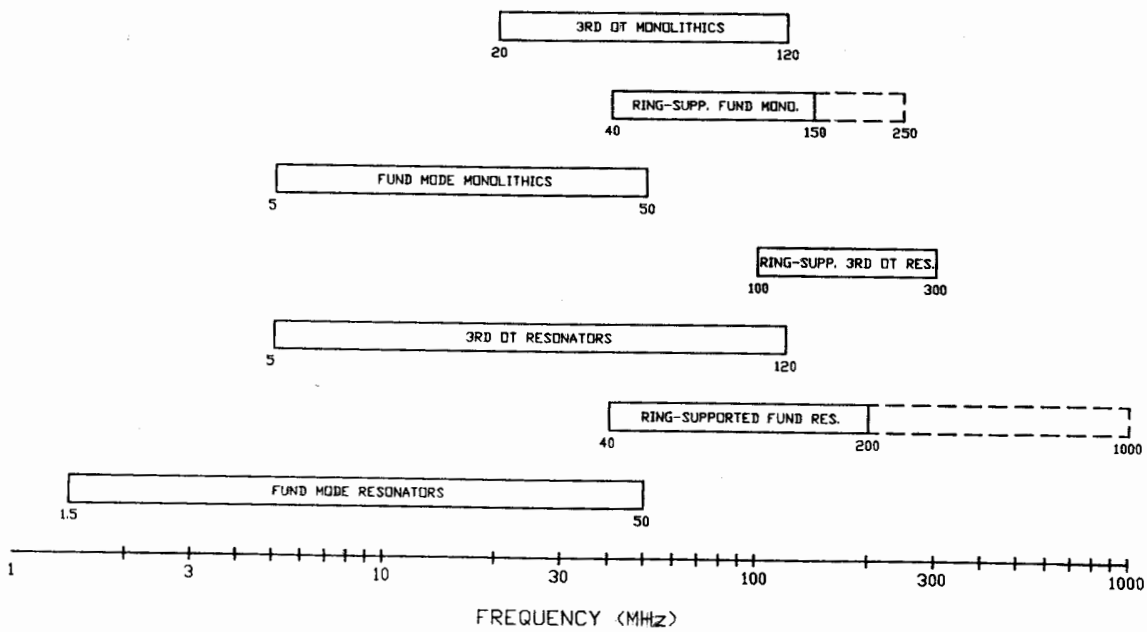
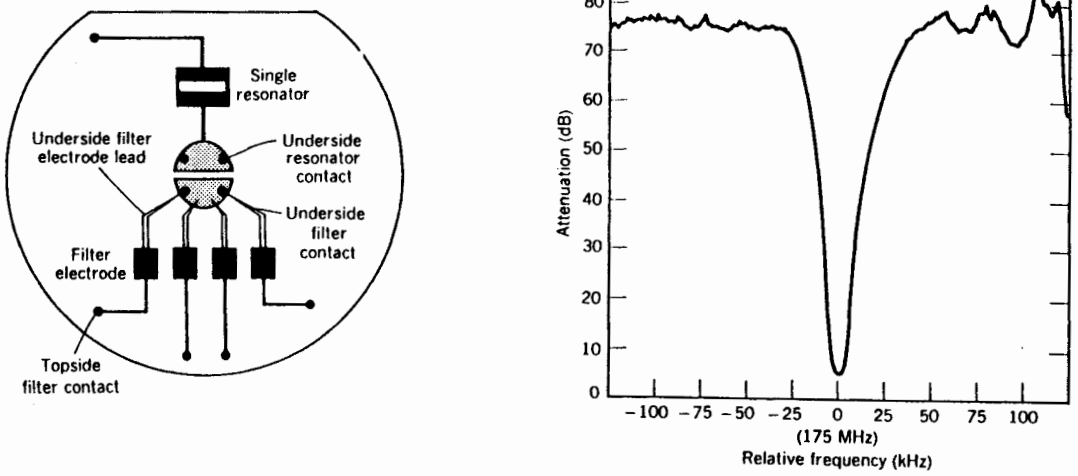
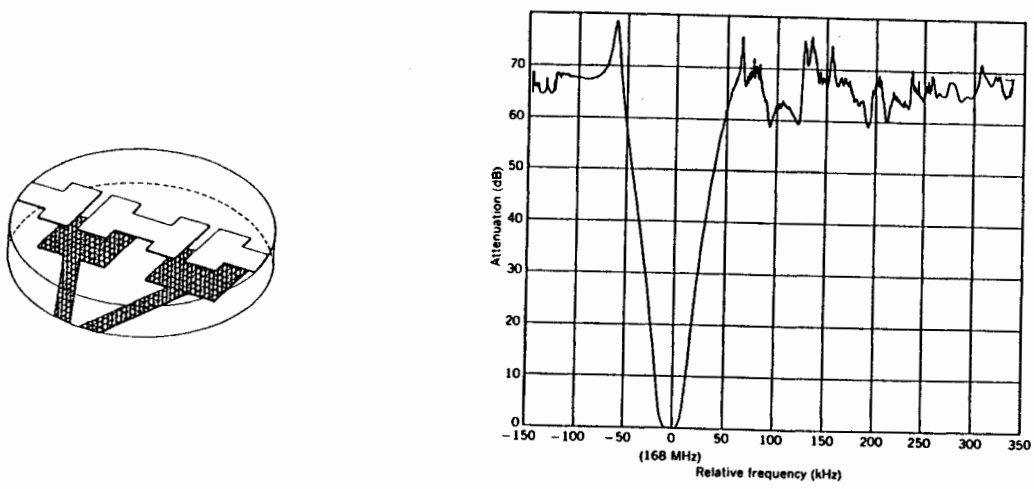


Fig. 6. Approximate frequency range of AT- and SC-cut quartz resonators and monolithics (acoustically-coupled resonators) using current production processes.



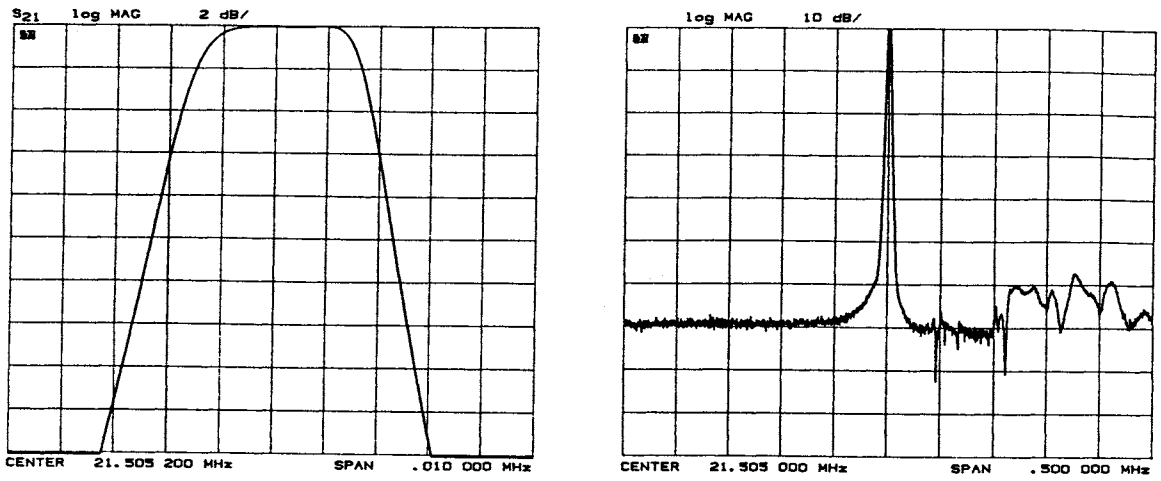
a. b.

Fig. 7. Four-pole VHF ACR. Center frequency is 175 MHz, using the fifth overtone: a. Electrode arrangement, showing four acoustically-couple resonators; grounding is via a novel central mounting pad [DW 85]; b. Attenuation characteristic; 3 dB bandwidth is 12 kHz. (The single resonator shown is used in the first local oscillator of the receiver and is not part of the filter.)



a. b.

Fig. 8. Four-pole VHF ACR. Center frequency is 168 MHz, using the fifth overtone: a. In-line arrangement of two MDR's fabricated on a single 6.4 mm wafer; this topology is used from 30 MHz to 170 MHz; b. Attenuation charcteristic; 3 dB bandwidth is 25 kHz.



a. b.

Fig. 9. Four-pole, 21.5 MHz ACR: a. Passband response; 3 dB bandwidth is 3.5 kHz; b. stopband response; ultimate attenuation approaches 70 dB. This device uses a folded topology similar to that shown in Fig. 12.

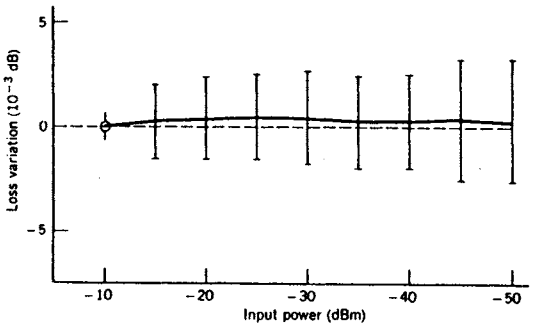


Fig. 10. Variation of mid-band insertion loss with input power level for a four-pole, 20 MHz monolithic filter. Data shown are for a lot of 60 filters; bars denote 2 sigma limits.

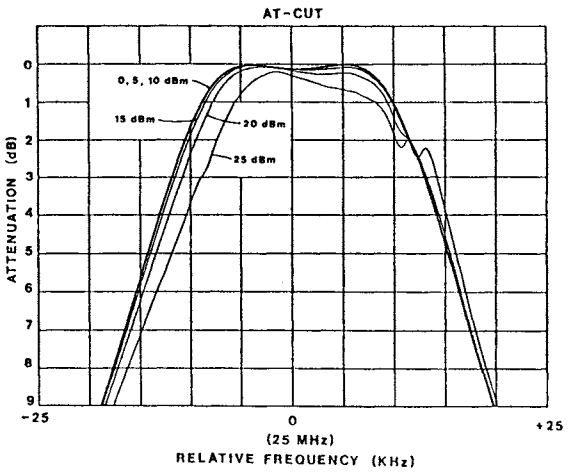


Fig. 11. Passband attenuation as a function of input power level for 25 MHz fundamental mode MDR's using three different cuts of quartz: a. AT-cut; b. BT-cut; c. SC-cut.

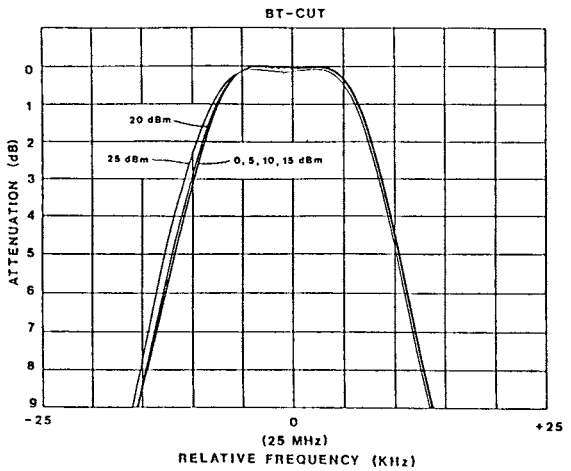


Fig. 11b

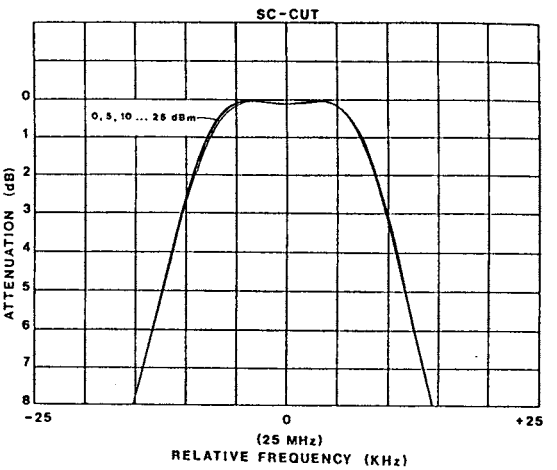
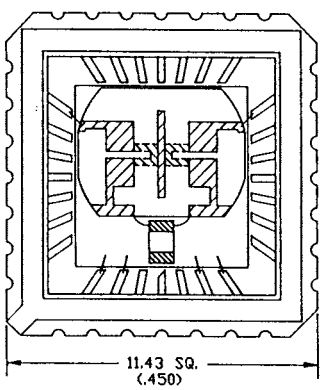
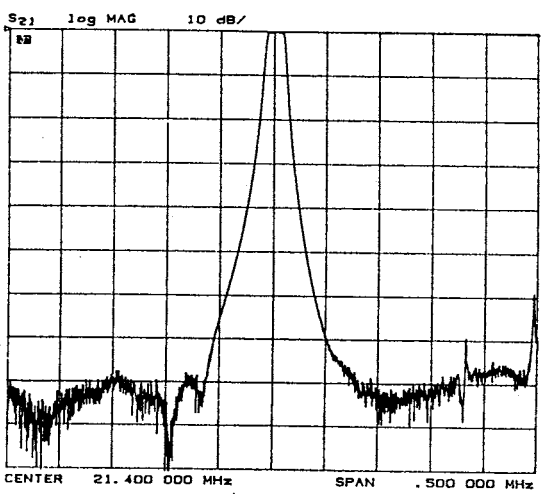


Fig. 11c



a.



b.

Fig. 12. Leadless four-pole, 21.4 MHz ACR: a. A folded topology and central mounting are used; b. Attenuation characteristic; 3 dB bandwidth is ±7.5 kHz.

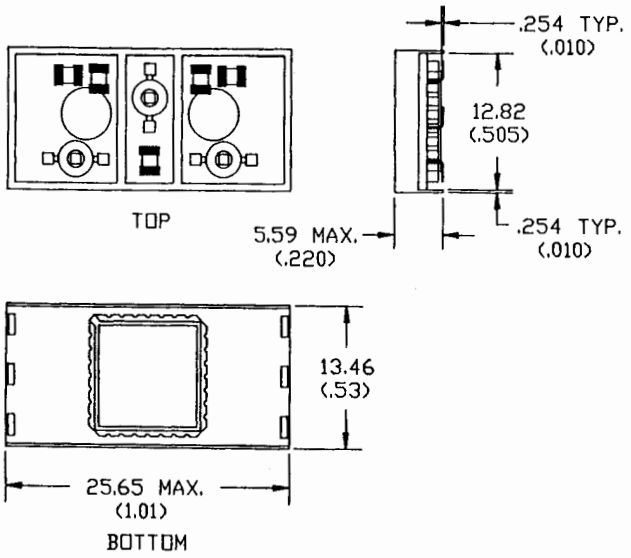


Fig. 13. J-lead surface-mount monolithic filter intended for a hybridized receiver. Terminations for this four-pole filter are 50 ohms. Filters having similar construction are under development at 20 and 45 MHz.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session QCP-01 Number 4

QUARTZ CRYSTAL OSCILLATORS FROM THEIR DESIGN TO THEIR PERFORMANCES

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

Abstract

The paper will review the recent developments in the design and the performances of quartz crystal oscillators.

The design of quartz oscillators can be aided by means of computer programs. The possibilities of these programs, but also their limitations, will be described.

The actual trends in the oscillator developments, in connection with the applications, will be reviewed from high stability ovenized oscillators to low price temperature compensated oscillators. The actual problems encountered in connection with the environment will be considered.

Finally the characteristics and performances of the main quartz oscillator families will be presented.

Introduction

The family of quartz crystal oscillators is composed of a large variety of frequency sources which cover a frequency range from a few 10 kHz (typically 32 kHz for watch crystal oscillators) up to about 200 MHz with regular bulk wave resonators and 1 GHz with resonators made by ionic or chemical etching ; in addition surface acoustic wave devices cover the 100 MHz to 2 GHz range. This corresponds to five decades of frequency (Fig. 1). The variety is also due to the large extend of their performances. The simplest quartz oscillator (XO) (clocks oscillators) have a mean frequency stability (mean including short term fluctuations, aging, frequency shifts due to temperature, etc.) of the order of 10^{-4} to 10^{-5}. Some of the most sophisticated oven controlled oscillators can reach short term stability of 10^{-13} (Fig. 2) ; therefore performances can differ by eight orders of magnitude. In term of volume (Fig. 3), of power consumption (Fig. 4), of price (Fig. 5) these oscillators can also range over several orders of magnitude.

During the last decade many kinds of quartz oscillators were developed following new needs for applications. A classification of these oscillators must involve for example clock oscillators (XOs), temperature compensated oscillators (TCXOs), digital temperature compensated oscillators (DTCXOs) or microcomputer compensated oscillators (MCXOs), voltage controlled oscillators (VCXOs), tactical miniature oscillators (TMXOs), miniature oven controlled oscillators (MOCXOs) and oven controlled oscillators (OCXOs).

Therefore the choice of a quartz oscillator is a compromise between its different characteristics, such as short term stability, spectral purity, aging, stability over the temperature range, consumption, volume, price, etc. According to the applications there can be in addition to be taken into account dynamic temperature characteristic, retrace, output signal, load and voltage supply sensitivities, and related to the environment are the various sensitivities to accelerations, vibrations, pressure, irradiation, and more recently to acoustic noise, magnetic fields, humidity.

Oscillator modeling

Most of the models which have been developed for quartz oscillators were essentially resonators models. There was, since the beginning, a need for the calculation of crystal orientation and resonator structure in order to understand the wave mechanisms, the sensitivity to the environment, the nonlinear behaviors, etc. Since the frequency of the oscillator is given by the resonator (at least at the first order of approximation), and also because there is only a small number of components in an oscillator, there was not such need for modelizing the whole unit ; therefore the design of oscillator was and is still rather empirical. However the behavior of the oscillating loop is basically nonlinear, and this is the condition to oscillate at constant amplitude. Thus the nonlinear differential equations describing the oscillator do not have obvious solutions and only approximated analytical calculations are possible. Even the basic equations cannot be solved directly on a computer because of the high Q-factor of the resonator. High Q means that any transient of the oscillator has a time constant τ which can be defined by

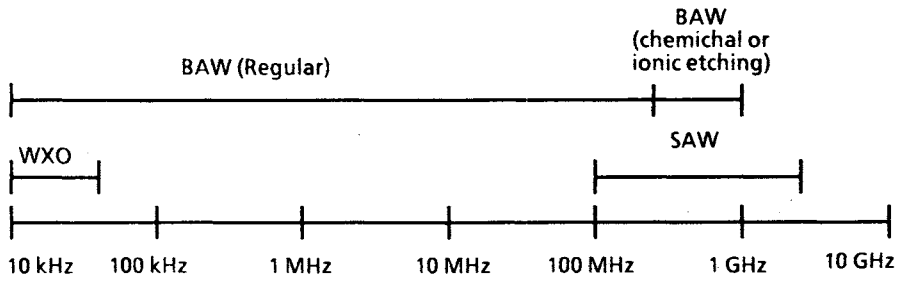


Fig. 1 : Frequency range

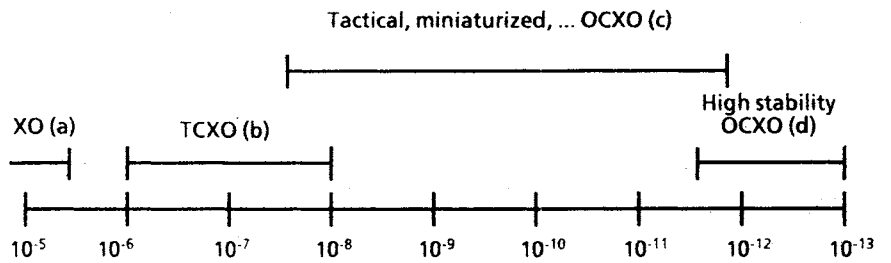


Fig. 2 : Frequency stabilities
a) per day b) over the temperature range c) in various environment conditions d) short term

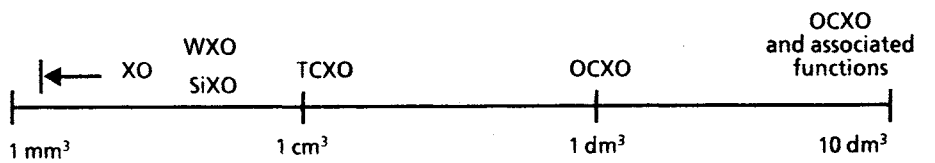


Fig. 3 : Volumes

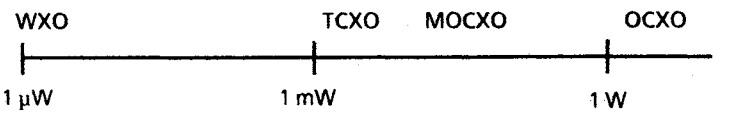


Fig. 4 : Power consumption

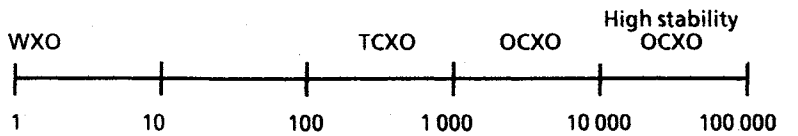


Fig. 5 : Prices (FF)

$$\tau = \frac{Q}{2\omega_o} \quad (1)$$

where Q is the loaded Q factor and ω_0 the angular frequency. The number N of periods of the oscillation during the period τ is

$$N = Q/\pi \quad (2)$$

Therefore a computer simulation will be irrealistic because of the very large number of oscillations involved in each transient. With high Q resonators this number is as high as 10^5 or 10^6.

The solution is to use only a linear model of the oscillator, or in case of nonlinear equations to apply some approximated method known as the method of the average which gives directly the amplitude and the phase (or frequency) of the signal averaged over one period.

The models are of two kinds. The computer program of general use which are electrical models of electronic circuits, like SPICE (and its different versions), ASTEC, IMAG, etc. The specialized programs developed for a special purpose like CODA (Crystal Oscillator Design and Analysis), BPT, LMPO.

In the linear models the conditions of the loop leads to complex equation, which is separated into real and imaginary parts. The imaginary part, which is the zero phase condition gives the frequency, the real part leads to the oscillation, or gain condition, but such a linear model in this form cannot give the oscillation amplitude.

The models specialized in oscillator design are generally hybrid ; they are linear for the frequency and nonlinear for the amplitude.

The fully nonlinear models are not distributed yet. The mathematics for obtaining the phase and amplitude relations is not so easy, but the main problem in fact is in the difficulty of measuring the nonlinear characteristics of the active components, and such informations generally are not available in the data sheets.

Several specialized computer programs have been made, linear and limited to Pierce oscillator [1], [2], hybrid and adaptable to different circuits [3] [4] [5] [6].

In models the complexity of the equations increases with the number of components. A new method, which is in progress, consists in using a formal computer programs for solving the equations [7]. This method is very powerful and enables to modelize oscillators with more than 20 components.

Temperature compensated crystal oscillators (analog and digital)

In analog TCXOs compensation is achieved by means of a network composed of three thermistors and a few resistors. Calibration of the oscillator is made preliminary by recording the voltage versus temperature characteristic, in which the voltage is the one to be applied on the oscillator varactor diode in order to maintain it at constant frequency. Then the thermistor network is calculated to reproduce the same voltage-temperature characteristic. A compensation over a wide temperature range is possible only if using resonators with large electromechanical coupling factors. This requires quartz crystals operated on their fundamental mode. Lithium tantalate has a higher coupling factor, but its temperature coefficients are also higher, therefore this material is more appropriate for VCXOs. Lithium niobate has a much higher temperature sensitivity which prevent to use it. However some new possibility could be offered in the future by berlinite (aluminum phosphate).

In fact most of the price of analog TCXOs is due to the calibration and to the adjustment of the thermistor network. Without calibration the dispersion of the resonator frequency-temperature characteristic limits the performances to several parts per million. With calibration and adjustment analog TCXOs have a typical stability of $\pm\ 1.10^{-6}$ to $\pm\ 5.10^{-7}$ according to the accuracy of the compensation process ; this accuracy largely depends on how powerful the compensation computing program is.

New principles were applied for compensation in the recent past or are still in progress. Compensation can be made by using a polynomial calculation. The frequency-temperature characteristic of a free oscillator is well approximated by the relation

$$(f - f_o)/f_o = \sum_{n=1}^{3} a_n (T - T_o)^n \quad (3)$$

A similar polynomial equation is used for compensation following

$$(V - V_o)/V_o = \sum_{n=1}^{3} b_n (T - T_o)^n \quad (4)$$

where the number N will be limited to N = 4.

The temperature probe is made by using the bandgap of a transistor, and the different terms of equation (4) are obtained by using multipliers in cascade. The all electronic (oscillator and compensation) can be integrated in a ASIC [8].

Another solution consists in building up the curve with linear segments of operational amplifier transfer function. This method is also comparable with integrated circuits. But whatever the analog compensation is, external adjustment of the characteristic by means of resistors is necessary.

Digital temperature compensated oscillators are essentially based on the use of dual-mode oscillators. The first attemps were made with SC-cut resonators operated at the same time on the C-mode (compensated mode) and B-mode (temperature sensitive mode) [9, 10]. The interest of this solution is in fact that the resonator is at the same time the frequency reference and the temperature probe.

The frequency stability over the temperature range of such an oscillator using 5 or 10 MHz SC-cut crystals with fundamental C-mode and 3 overtone B-mode is shown in Fig. 6. But it appeared that the B-mode was not reliable for most resonators, mainly because of activity dips.

Another possible method, which keeps the advantage of the self temperature sensing dual-mode oscillator consists in operating the C-mode at the same time on its fundamental frequency and on one overtone [11]. The difference between the characteristics of harmonic modes is almost linear in temperature and the corresponding beat frequency

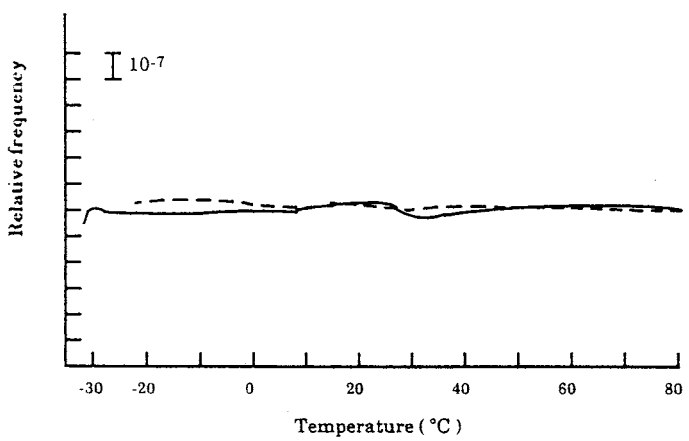


Fig. 6 : DTCXO frequency stability (ref. [10]

can be used for measuring the temperature. Temperature calibration of the oscillator is still necessary, but the digital compensation can be adjusted by introducing the appropriate weighting values in a memory and does not require any external resistor adjustment.

Performances of the different types of TCXOs are summarized in Table I.

	compensation	stability	temperature range
analog	thermistors	$\pm 2.10^{-6}$	-40 °C to + 80°C
	polynomial	$\pm 5.10^{-7}$ to $\pm 10^{-6}$	- 40°C to + 85°C (ref. [8])
digital	B and C- modes (DTCXO)	$\pm 2.10^{-8}$ to $\pm 2.10^{-7}$	- 30°C to + 80°C (ref. [10]
	fund. and 3rd C-modes (MCXO)	5.10^{-8}	- 55°C to + 85°C (ref. [12]
	fund. and 3rd C-modes (MCXO)	$\pm 2.10^{-8}$	- 55°C to + 85°C (ref. [13]

Table I : Performances of analog and digital TCXOs

Voltage controlled oscillators

Voltage controlled oscillators use the possibility of pulling the frequency of a crystal resonator by means of a variable capacitor in series with the resonator. The frequency is given by the well known relation

$$f = f_o \left[1 + \frac{C_1}{2(C_o + \Gamma} \right] \tag{5}$$

where C_1, C_o, and Γ are respectively the motional, parallel, and variable capacitors. Therefore the sensitivity is related to the value of the motional capacitance C_1 which is given by the relation

$$\frac{C_1}{C_o} = \frac{8K^2}{\pi^2(2p+1)^2} \tag{6}$$

where K is the electromechanical coupling factor, and $(2p+1)$ the overtone rank. Therefore the frequency pulling range is maximum for the fundamental mode.

The coupling factor depends on the material, as shown in Table II.

material	electromechanical coupling factor K
quartz	10 %
$LiTaO_3$	44 %
$LiNbO_3$	47 %
$A\ell PO_4$	23 %

Table II : Electromechanical coupling factor

With quartz oscillators a frequency pulling of 10^{-4} to 10^{-3} is possible, with at the same time the low temperature sensitivity of quartz. However the pulling range can be increased by using lithium tantalate resonators [14]. According to their higher electromechanical coupling factor a frequency shift of 1 % can be achieved, with a linearity of the order of 1 %. New possibilities are expected from berlinite crystals ($A\ell PO_4$), which combine a good coupling factor and a low temperature sensitivity.

Miniature ovenized crystal oscillators

When higher performances are required, the influence of temperature must be reduced by using the oscillator in a temperature controlled oven. The advantage is the high temperature stability achievable, the disadvantage is the power consumption, several watts, of such ovens. A compromise can be made by using small size resonators with miniaturized ovens in order to reduce the consumption.

New or reniewed small size quartz resonators were developed during the last decade [15]. A breakthrough was made possible by using ultrasonic machining, chemical etching, ion-beam milling, laser milling, combined with the photolithographic process, which enable to realize crystal plates with lateral sizes of a few millimeters and thicknesses as small as a few microns.

The first small size resonators were made for wrist watch application. Most of them were originally flexural bars and today are 32 kHz tuning forks. At higher frequencies, in fact from 1 MHz to 1 GHz, different kinds of resonators have been studied. Rectangular AT-cut crystals [16 - 20] are a compromise between sizes, electrical characteristics, and manufacturing cost. The first developments of such rectangular resonators again were achieved by the watch industry. More recently similar resonators, with even a smaller size, using a higher electromechanical coupling factor material like lithium tantalate and lithium niobate [21 - 23] were studied. Miniaturized versions of the GT-cut [24 - 25] offered the combination of small size and low temperature sensitivity. Length-extension bars [26] at 1 MHz made by photolithographic process and chemical etching were shown. The reduction of the size of circular AT-cut plates was also undertaken [27] in the 1-5 MHz frequency range. In addition AT and SC-cut resonators at higher frequencies (10 - 30 MHz) will have a smaller size and therefore can be used for the same purpose.

The characteristics of some of these resonators are presented in Table III.

resonator	frequency	Q-factor	frequency stability (flicker floor)	crystal volume
rectangular AT ref. 19 X elongated	10 MHz	3×10^4	1.6×10^{-10}	1.5 mm^3
rectangular AT ref. 3 Z' elongated	5 MHz	5×10^5	2×10^{-11}	4.3 mm^3
GT ref. 13	2 MHz	2.5×10^5	1.5×10^{-11}	.15 mm^3
ZT ref. 17	2 MHz	5×10^5	8×10^{-12}	1.2 mm^3
length ext. Z + 2 ref. 18	2 MHz	10^4	1.4×10^{-10}	0.013 mm^3

Table III : Characteristics of small size resonators

The frequency stabilities shown in Table III, even if they are several orders of magnitude worse compared to the high stability crystals, can satisfy a large part of the oscillator market ; a part which corresponds to stabilities between 10^{-8} and 10^{-11}, which is not covered by TCXOs since their performances are not sufficient, and not by high stability OCXOs because of their price.

Many solutions can be used and it is difficult to give an exhaustive presentation of all the possible configurations ; only two oscillators will be briefly described as examples.

The tactical miniature crystal oscillator (TMXO) was designed for fast warm up and low power consumption [28], and in fact its most original characteristics are in its mechanical design. The crystal is mounted in a ceramic flatpack

enclosure and associated to an hybrid electronic circuit inside an inner container, which is itself in hermetically sealed outer container. Small wire suspension provides the necessary thermal insulation. Short term stability is near 10^{-11} over 1 sec, and the power consumption is around 100 mW at room temperature [29].

A second example is the OCXO developed for the ARGOS and SARSAT programs. The resonator is a 5 MHz rectangular AT-cut bar (of the type presented above). The electronics of the oscillator and of the temperature controlled oven are integrated in one single chip [10]. The stability is 2.10^{-11} over 1 sec. and the power consumption is 150 mW at room temperature.

High stability oven controlled oscillators

High stability OCXOs are designed without to much limitation on their volume and their power consumption. The main goals are short and long term stabilities and spectral purity. High stability requires quartz resonators with Q-factors between 2.5×10^6 and 3×10^6, low noise electronics, and very accurate temperature control. Evaluation of these characteristics can be made by using some of simple equations modelling resonators and oscillators.

The resonator temperature dependence is given by

$$\frac{f - f_o}{f_o} = a_1(T - T_o) + a_2(T - T_o)^2 + a_3(T - T_o)^3 + \tilde{a}\frac{dT}{dt} \qquad (7)$$

Since the crystal cut is such that $a_1 = 0$ and the resonator being operated at its turn over temperature the static behavior depends mainly on the second term in equation (7), and the dynamic behavior is due to the $\tilde{a}$ constant. Since $a_2 \simeq 10^{-9}/°C^2$ and $\tilde{a} \simeq 10^{-7}$ sec/°C for SC-cut crystals, in order to keep the frequency variations at 10^{-13} the temperature fluctuations in the crystal must not exceed 0.01°C for the static part and 1 µ°C/sec for the dynamic one. These numbers, which correspond mainly to laboratory environment, give the requirements for the oven control.

Progress was made by using low noise electronics with high quality conventional AT-cut resonators [30], and then with BVA electrodeless resonators [31] with QAS electroded resonators [32], and more recently with 5th overtone SC-cut crystal [33] and with different types of high stability crystals [34]. A comparison of the short term stability of these different oscillators is made on Fig. 7. It was also shown that surprisingly the humidity can affect the medium and long term stability, and a better control give an improvement by a factor 2 [35].

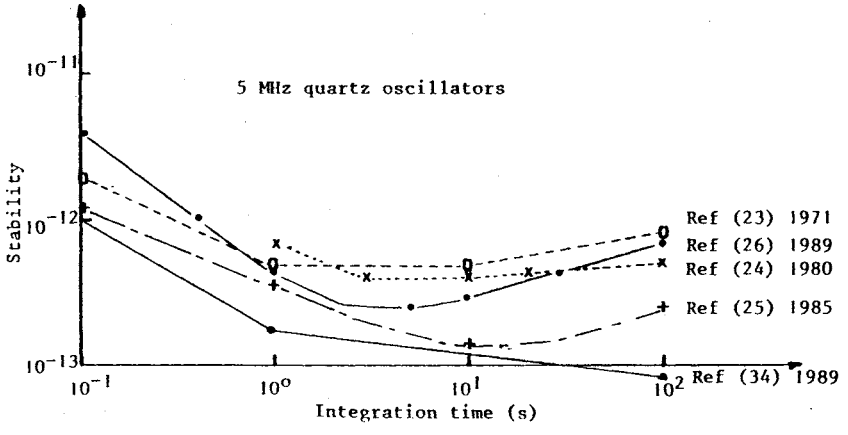


Fig. 7 : Short term stability of high performance 5 MHz quartz oscillators

Possible evolutions and conclusion

The emphasis was put on oscillators and not on resonators, since another paper devoted to resonators can be found in these proceedings.

However the future of quartz crystal oscillators will be largely influenced by the properties of quartz resonators. Major improvements were the doubly rotated cuts, mainly the SC-cut, new principles for the electrodes and the crystal mounting as in the BVA technique, new technologies for the crystal processing like chemical polishing and etching, ion milling, flat-pack enclosures, etc. These new resonators present new characteristics, and in particular the sensitivities to temperature gradients and thermal stresses, to forces, accelerations, pressures, etc. were reduced. A material with higher purity, lower defect contents becomes available for reducing frequency aging or irradiation sensitivity.

The needs for the future in electronic systems can be summarized as oscillators with higher stability and lower power and volume. Higher frequencies are also required for some applications.

Miniaturization oscillators is made possible by the development of integrated circuits and in particular of ASICs, but miniature high stability quartz resonators are not available yet. The feasability of micro-scale devices has been already demonstrated by using the technics of microelectronics. Higher frequencies, 1 GHz for instance, have been achieved the same way. But one important archaism remains the enclosures. Flat-pack, small size, ceramic enclosures are not really developed.

Computor aided design is also at its beginning for oscillators. Computer programs applicable to the different kinds of oscillators circuits, taking into account all the components are necessary. Linear models for calculation of frequency are the first step, but they must be extended to nonlinear models in order to include at the same time frequency and amplitude. This supposes a method for measuring the nonlinear characteristics of the active components of the oscillator. Efforts are in progress today on these different tasks.

Bibliography

[1] R. Nardin, J. Ho, "Computer design and analysis for high precision oscillators", 28th A.F.C.S. (1974).

[2] Y. Onata, "New approach to the design of crystal oscillators", 28th A.F.C.S. (1974).

[3] A. Benjaminson, "The design and analysis of VHF/UHF crystal oscillators", 41st A.F.C.S. (1987).

[4] B.Parzen, A. Ballato, "Design of crystal and other harmonic oscillators", Wiley Intersciences (1983).

[5] B. Parzen, "Classification and a generalized restatement of Leeson's oscillator noise model", 42nd A.F.C.S. (1988).

[6] T. Hall, "Computer aided design and assembly of oscillators", 36th A.F.C.S. (1982).

[7] R. Brendel, LPMO, to be published (1990).

[8] J.S. Wilson, "An improved method of temperature compensation of crystal oscillators", Proc. 37th Ann. Freq. Cont. Symp. (1983).

[9] J.A. Kusters, M.C. Fischer, J.F. Leach, "Dual mode operation of temperature and stress compensated crystals", Proc. 32nd Ann. Freq. Cont. Symp. (1978).

[10] G. Marianneau, R. Prétot, G. Caret, "Compensation numérique de la sensibilité à la température des oscillateurs à quartz", Proc. Congrès Int. Chrono, Besançon (1984).

[11] J.S. Schodowski, "Resonator self-temperature-sensing using a dual-harmonic-mode crystal oscillator", Proc. 43rd Ann. Freq. Cont. Symp. (1989).

[12] M. Block, M. Meiers, J. Ho, "The microcomputer compensated cyrstal oscillator (MCXO)", Proc. 43rd Ann. Freq. Cont. Symp. (1989).

[13] A. Benjaminson, S.C. Stallings "A microcomputer compensated crystal oscillator using a dual-mode resonator", Proc. 43rd Ann. Freq. Cont. Symp. (1989).

[14] L. Bidart, E. Girardet, M. Wolcoff, "Oscillateurs à grande excursion de fréquence", Proc. 1st European Time and Frequency Forum, Besançon (1987).

[15] J.J. Gagnepain, "Recent advances in subminiature/thin membrane resonators", Proc. Ultrasonics Symp. (1985).

[16] J.J. Royer, "Rectangular AT-cut resonators", Proc. 27th Ann. Freq. Cont. Symp. (1973).

[17] M. Onoe, M. Okazaki, "Miniature AT-cut strip resonators with tilted edges", Proc. 29th Ann. Freq. Cont. Symp. (1975).

[18] A. Zumsteg, P. Suda, "Properties of a 4 MHz miniature flat rectangular quartz resonator vibrating in a coupled mode", Proc. 30th Ann. Freq. Cont. Symp. (1976).

[19] A. Zumsteg, P. Suda, W. Zingg, "Energy trapping of coupled modes in rectangular AT-cut resonators", Proc. 32nd Ann. Freq. Cont. Symp. (1978).
[20] M. Okazaki, N. Manabe, "AT-cut strip resonators enclosed in cylindrical package", Proc. 38th Ann. Freq. Cont. Symp. (1984).
[21] M. Okazaki, W. Watanabe, "Miniature $LiTaO_3$ X-cut strip resonator", Proc. 37th Ann. Freq. Cont. Symp. (1983).
[22] Y. Kojima, F. Fujiwara, S. Yamada, N. Wakatsuki, "Chip crystal resonator with load capacitors", Proc. 38th Ann. Freq. Cont. Symp. (1984).
[23] R. Lefevre, J. Jenselme, D. Servajean, "Laser processed miniature LiTaO3 resonators and monolithic filters", Proc. 39th Ann. Freq. Cont. Symp. (1985).
[24] H. Kawashima, H. Sato, O. Ochiai, "New frequency temprature characteristics of miniaturized GT-cut quartz resonators", Proc. 34th Ann. Freq. Cont. Symp. (1980).
[25] J.Hermann, "A novel miniature ZT-cut resonator", Proc. 39th Ann. Freq. Cont. Symp. (1985).
[26] R. Dinger, "A miniature quartz resonator vibrating at 1 MHz", Proc. 35th Ann. Freq. Cont. Symp. (1981).
[27] Y. Oomura, "Frequency temprature behavior of miniaturized circular disk AT-cut crystal resonator", Proc. 34th Ann. Freq. Cont. Symp. (1980).
[28] H.W. Jackson, "Tactical miniature crystal oscillator", Proc. 34th Ann. Freq. Cont. Symp. (1980).
[29] D. Brown, E.Laszlo, R.McGill, P. Stoermer, "Manufacturing methods and technology for tactical miniature crystal oscillator", Proc. 38th Ann. Freq. Cont. Symp. (1984).
[30] H. Brandenberger, F. Hadorn, D. Halford, J. Shoaf, "High quality quartz crystal oscillators : frequency domain and time domain stability", Proc. 25th Ann. Freq. Cont. Symp. (1971).
[31] R. Besson, P. Girardet, E. Graf, D. Emmons, "Performance of new oscillators designed for electrodeless crystals", Proc. 34th Ann. Freq. Cont. Symp. (1980).
[32] J.J. Gagnepain, "A comparison of frequency noise of quartz resonators", Proc. 39th Ann. Freq. Cont. Symp. (1985).
[33] M. Bloch, J. Ho, C. Stone, A. Syed, F. Walls, "Stability of high quality quartz crystal oscillators : an update", Proc. 43rd Ann. Freq. Cont. Symp. (1989).
[34] J.R. Norton, "Ultrastable quartz oscillator for spacecraft", 21st Annual P.T.T.I. (1989).
[35] F. Walls, "The influence of pressure and humidity on the medium and long term frequency stability of quartz oscillators", Proc. 42nd Ann. Freq. Cont. Symp. (1988).

Frequency Standards

FS-01 Review Papers

Chairman: N.F. Ramsey

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-01 Number 1

CLASSICAL FREQUENCY STANDARDS

Claude AUDOIN

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

The background principles of atomic frequency standards are recalled. The cesium, the hydrogen and the rubidium frequency standards are described. Their typical performances are given and the prospects of improvement are mentionned.

1. INTRODUCTION

In this paper, the basic principles of the most widely used atomic frequency standards are summarized. They are the rubidium cell frequency standard, the cesium beam frequency standard and the hydrogen maser, by order of numerical importance. These devices are described at the tutorial level. The sources of frequency stability and of accuracy limitations are discussed and information on the present state of the art is given. Prospects of improvement are mentionned.

For a more detailed insight into the subject, the reader is referred to references [1] to [3].

2. PHYSICAL PRINCIPLES

The principles of atomic frequency standards take root in basic physics. The progress which have led to the present state of the art has been honoured by a number of Nobel prizes. Limiting ourself to the Nobel prize winners who have the most directly contributed to the principles of atomic frequency standards, the list is nevertheless impressive. Max Planck (1918) founded the quantum theory. Albert Einstein (1921) introduced the concept of photon and of stimulated emission, among many other major contributions. Niels Bohr (1922) gave the basis of atomic structure. Louis de Broglie (1929), Werner Heisenberg (1932) and Erwin Schrödinger (1933) created and developed quantum mechanics. Otto Stern (1943) contributed to atomic and molecular beam kinetics. Isaac Rabi (1944) developed the molecular beam magnetic resonance method from which the cesium beam frequency standard is derived. Charles Townes, Nikolaï Basov and Alexándre Prokhorov (1964) implemented the maser effect, in the ammonia maser. Alfred Kastler (1966) gave the principles of optical pumping. Norman Ramsey (1989) introduced the separated oscillatory field method and invented the hydrogen maser. Wolfgang Paul and Hans Dehmelt (1989) invented the ion trap and used it to make very precise measurements on charged particles, such as the electron and the positon, and to implement experimentally the method of radiation cooling.

Let us thus consider an atom of a given element. It has energy levels of well defined energy. Let E_1 and E_2 the energy of two of them, with $E_2 > E_1$. A transition between these two levels may occur under the effect of an electromagnetic radiation of frequency ν_o. Energy conservation determines the value of ν_o. This leads to Bohr's relation :

$$h\nu_o = E_2 - E_1 \quad , \tag{1}$$

where h is Planck's constant. Therefore a resonance may be observed. This resonance has a width $\Delta\nu$. It is given by the Heisenberg's uncertainty relation :

$$\Delta\nu \cdot \Delta t \gtrsim 1 \quad , \tag{2}$$

where Δt is the practical duration of the observation of the transition. Δt ranges from approximately 1 ms in a rubidium cell and in a cesium beam tube to 1 s in a hydrogen maser. With the values of ν_o involved in microwave frequency standards, the atomic line quality factor is comprised between 10^7 to 10^9.

There is presently no reason to doubt that atomic properties do not depend on space and time coordinates (except for known relativistic effects). Therefore, any one can build a device defining the same frequency as others when the same atom and the same transition are used. This frequency standard is based on the microscopic properties of matter. This is in contrast with frequency generators whose frequency is

defined by the resonance properties (electrical or mechanical, or both) of a macroscopic block of matter. Among other parameters, the resonance frequency then depends on the size of the block and thus on temperature, constraints, etc.

The two levels of interest to us are determined by a magnetic interaction in the atoms. It occurs between the magnetic moment of the unpaired electron of alkali or alkali-like atoms and the magnetic moment of the nucleus. It is called the hyperfine interaction. This interaction is weak and it gives a small value of $E_2 - E_1$. Consequently, the resonance frequency ν_o belongs to the microwave frequency range. It is approximately equal to 1.4 GHz, 6.8 GHz and 9.2 GHz for the hydrogen, the rubidium and the cesium atom, respectively. As a practical consequence, the associated electronic system which controls the frequency of a quartz crystal oscillator by that of the atomic transition can be made efficient, small and reliable.

The energy difference $E_2 - E_1$ of the levels that are chosen increases very slightly, in fact quadratically, with the value of the magnetic field applied. It is thus made small, between approximately 10^{-7} to 10^{-5} tesla with the purpose of reducing as much as possible its effects on the value of ν_o.

Another, but annoying, consequence of the smallness of $E_2 - E_1$ is that we have in fact $E_2 - E_1 \ll kT$, where kT is the thermal energy. Therefore, at thermal equilibrium, the two energy levels are almost equally populated. It follows that this thermal equilibrium must be broken in order to be able to observe a change in the atom properties or an exchange of a detectable amount of energy when the atomic transition occurs. Two different methods are implemented to overcome thermal equilibrium. The first one is related to the fact that the atom show opposite values of its effective magnetic moment whether it occupies one or the other of the two energy levels that we are considering. Consequently the deflection of atoms in a beam passing in an inhomogeneous and intense magnet field depends on their internal state and they can be separated. The other method is based on optical pumping. Besides the hyperfine interaction levels 1 and 2 in the ground state, the atoms considered have excited levels of much larger energy, such as level 3 (see Fig. 1). An optical radiation, having an appropriate wavelength, can transfer atoms from level 2 to 3 , for instance. Spontaneous decay to level 1 and 2 occurs. The net result of optical pumping is that the population of one of the ground state level (level 1 here) increases to the detriment of that of the other level (level 2 in the example given).

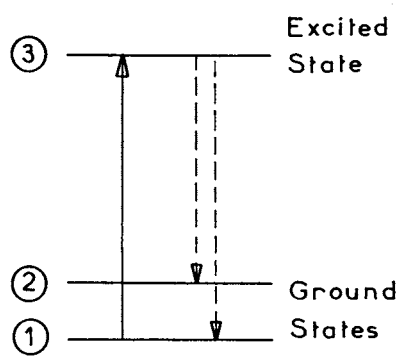


Fig. 1. Principle of optical pumping

3. BRIEF DESCRIPTION OF THE MICROWAVE ATOMIC FREQUENCY STANDARDS

3.1. The cesium beam frequency standard

The first cesium beam frequency standard has been operated by Essen and Parry in 1955 at the National Physical Laboratory (England) [4]. Their industrial development started soon after in the USA and manufactured units were available a very few years later. Laboratory cesium beam frequency standards have been built and operated in various countries with the aim of obtaining the best possible accuracy.

Figure 2 represents schematically a cesium beam tube of classical design. A cesium beam effuses from an oven. The beam passes through a first state selecting magnet. Atoms having the desired internal state are deflected towards the beam axis.

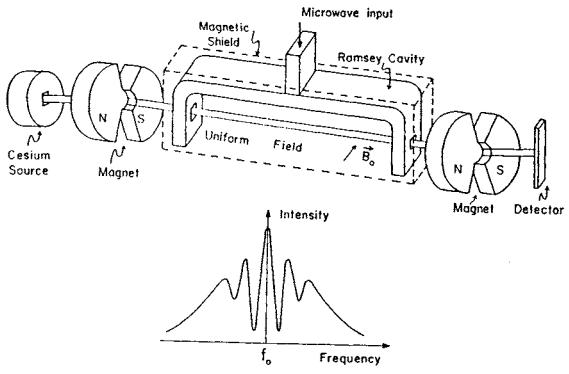


Fig. 2

Schematic representation of a cesium beam tube and shape of the resonance patters. The mass spectrometer and the electron multiplier are not shown.

Afterwards the beam crosses the microwave interaction region. It comprises a microwave cavity with two arms, which enable the Ramsey's two oscillatory field method to be implemented [5]. This method allows in fact to choose the total interaction length at will, without any constraint related to the radiation wavelength. The line quality factor can thus be increased accordingly. The method has the additional advantage of decreasing the size of a number of spurious frequency offsets by the factor ℓ/L, where ℓ is the length of each oscillatory field and L their separation. In the interaction region, atoms change their internal state when the resonance condition given by equation 1 is fulfilled. The value of ν_o is equal to 9 192 631 770 Hz, by definition of the second. The interaction region is shielded against changes of the magnetic field and a magnetic field of approximately 6×10^{-6} tesla is created in that part of the device.

The second state selector deviates the atoms having made the transition towards a hot wire detector. Then, they are evaporated as Cs^+ ions which are selected in a mass spectrometer and accelerated to the first dynode of an electron multiplier. The output signal of the cesium beam tube is thus an electric current.

A resonance pattern, shown in figure 2, is obtained when the frequency of the interrogation signal is swept across the value ν_o. In manufactured devices, the linewidth is of the order of 500 Hz. It can be as small as a few tens Hertz in long laboratory cesium beam frequency standards. The cesium beam tube is a passive device which behaves as a resonator. A block diagram of the electronic system which controls the frequency of a quartz crystal oscillator at a frequency of 5 or 10 MHz is shown in figure 3. A frequency divider may deliver pulses, every second for instance.

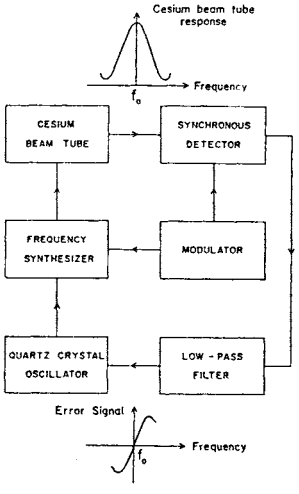


Fig. 3

Block diagram of the frequency control of a quartz crystal oscillator using the atomic resonance pattern.

There are two kinds of cesium beam frequency standards. In manufactured devices a trade-off is made between performances on the one hand and volume and cost on the other hand. Laboratory devices are operated with the aim of realizing at best the definition of the second. Laboratory set-ups are used to steer the frequency of the International Atomic Time scale. They are truly primary cesium beam frequency standards.

Since a few years, a number of laboratories and companies have started the study of optically pumped cesium beam frequency standards [6]. Figure 4 depicts the simplest realization of the device, with a single semi-conductor laser emitting a radiation at 852 nm. In the first region of interaction between the cesium beam and the laser beam, optical pumping occurs, as shown in Fig. 1.

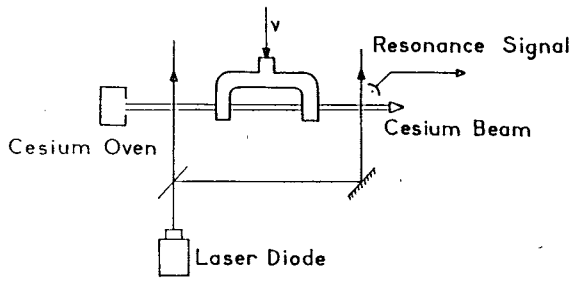


Fig. 4
Simplified representation of an optically pumped cesium beam frequency standard.

The atomic level 2 , for instance, becomes populated to the detriment of level 1 . When the microwave transition occurs, the populations of levels 1 and 2 are exchanged. Therefore, in the second optical interaction region, level 1 in the example given is populated. Consequently, the laser light brings the cesium atoms to the excited state, from which they decay by emitting fluorescence light. This light is detected. Its intensity depends on the exact frequency of the applied microwave field and a resonance pattern, similar to that of Fig. 2 is obtained. There is no significant deflection of the beam and the device is, in principle, very simple to build. Its advantages will be mentioned later.

3.2. The hydrogen maser

The first hydrogen maser has been operated in Professor Ramsey's laboratory, at Harvard University, around 1960 [7]. This device results from attempts to increase the interaction time Δt of equation (2) - and thus to increase the selectivity of the resonance - by storing the atoms in a properly coated bulb. It appeared that the cesium atom is not the most appropriate element for that purpose, due to its large electric polarisability. Hydrogen looked more promising. In other respects, the maser technique was successfully implemented a few years before, around 1955, in the ammonia maser. Then the concept of the hydrogen maser emerged by grouping the storage and the maser techniques.

Atomic hydrogen is produced by dissociating molecular hydrogen in a discharge tube (see figure 5). The source aperture forms a beam which passes through a four or a six pôle magnet. Atoms in the upper energy level are directed to the entrance of a storage bulb. Atoms in the lower energy are deflected off the beam.

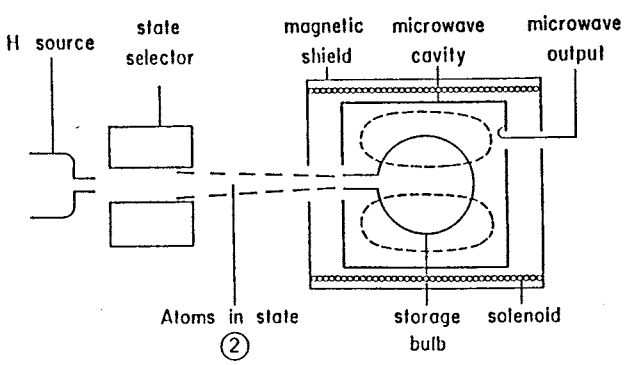


Fig. 5
Schematic representation of the hydrogen maser. The broken lines show magnetic lines of force of the microwave field. The vacuum chamber and the pump are not shown.

The storage bulb is coated with Teflon. Hydrogen atoms do not interact much with this material and the hydrogen atoms can bounce up to approximately 10^5 times on the bulb walls before relaxing to the lower energy state. The hydrogen storage time is thus of 1 s about, which provides a line quality factor of 10^9.

The bulb is placed in a low loss microwave cavity which is tuned to the atomic transition frequency. Atoms bring energy to the resonant cavity and an oscillation can be sustained. The oscillation frequency is mainly determined by that of the atomic transition which is equal to 1 420 405 751,770 ± 0,003 Hz. A magnetic shield protects the interaction region from changes of the magnetic field and a homogeneous field of 10^{-7} T about is created in this region. Hydrogen being a permanent gas, pumps are necessary to evacuate the device.

The hydrogen maser is an active frequency standard. Figure 6 shows a block diagram of the associated electronic system which phase-locks a high quality quartz crystal oscillator to the maser oscillation.

The size of the hydrogen maser is mainly determined by that of its microwave cavity and thus by the wavelength of the transition, i.e. 21 cm. The volume is therefore rather large, of a few hundred cubic decimetres. It is possible to reduce this size by loading the cavity with properly placed dielectric or conducting materials. Small size masers having a volume of a few tens dm^3 have been developed. They are operated either actively or passively.

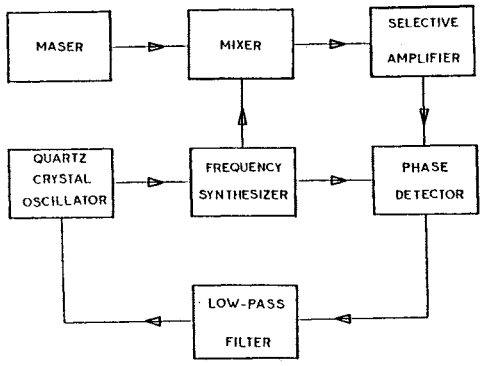


Fig. 6
Block diagram representation of the phase control of a quartz crystal oscillator by the maser oscillation.

3.3. The rubidium cell frequency standard

In the rubidium cell frequency standard, the methods of optical pumping and of optical detection of a microwave resonance are applied in a typical double resonance configuration. This frequency standard has been studied and developed since the end of the fifties [8,9].

The Rb^{87} lamp and the hyperfine filter containing a Rb^{85} vapor, which are shown in figure 7, play the same role as the state selecting magnets in the two atomic frequency standards described above. They enable to create a population difference between the energy levels E_1 and E_2 of the Rb^{87} atoms in the resonance cell. Level E_2 becomes more populated than level E_1, as shown in figure 8.

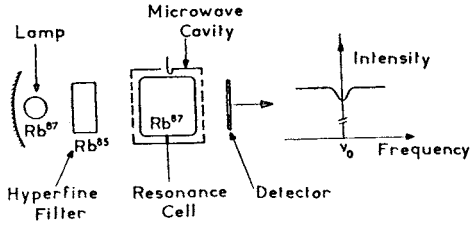


Fig. 7
Schematic representation of the rubidium cell frequency standards.

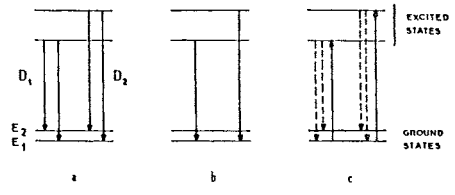


Fig. 8
Optical pumping applied to create a population unbalance in the rubidium cell frequency standard :
a) the lamp emits the D_1 and D_2 lines of the rubidium 87 atom. They show an hyperfine structure in the ground state associated to levels E_1 and E_2 of the text.
b) the hyperfine filter absorbs one of the hyperfine components of the lines D_1 and D_2. The filtered radiation is shown.
c) in the resonance cell, only the rubidium 87 atoms in the lower ground state (level E_1) can absorb the incoming photons.
A population difference is created : level E_2 becomes more populated than E_1.

In the resonance cell, a buffer gas, such as nitrogen, improves the efficiency of optical pumping. Furthermore, it reduces the diffusion velocity of rubidium atoms to a few cm s^{-1}. This eliminates greatly

the Doppler broadening of the microwave resonance line.

When, in addition, a microwave excitation field is produced in the microwave cavity, at the atomic resonance frequency ν_0 (about 6834,58 MHz), a transition occurs between levels E_1 and E_2 and this tends to balance their population. Consequently, the population of level E_1 increases and the resonance cell absorbs more the incident light when the resonance condition of equation (1) is fulfilled. Then a small dip appears in the intensity of the light transmitted by the resonance cell when the frequency of the microwave excitation is swept around ν_0. This is shown in figure 7. The linewidth is of the order of 500 Hz.

In very compact units, the hyperfine filter is integrated into the resonance cell which then contains natural rubidium, which is a mixture of Rb^{85} and Rb^{87}.

As in the other devices a magnetic shield surrounds the interaction region. A magnetic field of about 2.5×10^{-5} T is created.

The Rb cell frequency standard is a passive frequency standard. The principle of the electronic system which probes the atomic transition and which frequency controls a quartz crystal oscillator is the same as shown in figure 3.

5. FREQUENCY STABILITY

4.1. Short term frequency stability

In both categories of frequency standards, active and passive, the short term frequency stability is that of the quartz crystal oscillator controlled by the atomic resonance. This occurs for sampling times τ smaller than the control loop time constant, which is equal to 1 s in order of magnitude.

4.2. Shot noise or thermal noise limited frequency stability

For τ approximately comprised between 1 s and 1 day, the frequency stability is determined by the properties of the atomic resonance i.e. its line quality factor, Q_ℓ, and the signal to noise ratio $S/B(\tau)$ (in amplitude) of its observation during the sampling time τ.

4.2.1. Passive frequency standards

The frequency stability is directly related to the precision of the measurement of the extrema of the atomic resonance pattern. It can be shown that the frequency stability measure $\sigma_y(\tau)$ is given by :

$$\sigma_y(\tau) \simeq \frac{1}{Q_\ell \; S/B(\tau)} \quad . \tag{3}$$

In the cesium beam and the rubidium cell frequency standards, the noise is the shot noise of the flux of particles detected, atoms or photons respectively. We have :

$$S/B(\tau) \simeq \frac{I_s}{I_t^{1/2}} \; \tau^{1/2} \quad , \tag{4}$$

where I_s is the particle flux measuring the line amplitude and I_t is the total flux of particles detected. We thus obtain :

$$\sigma_y(\tau) \simeq \frac{I_t^{1/2}}{Q_\ell I_s} \; \tau^{-1/2} \quad . \tag{5}$$

Therefore the frequency stability varies as $\tau^{-1/2}$, which is related to a white frequency noise.

In the rubidium cell frequency standard, we have $I_s \ll I_t$ when a lamp is used. I_s can be greatly increased by replacing the lamp by a semi-conductor laser and therefore the frequency stability can be improved [10]. This is one of the motivations of the present work in this field.

In the cesium beam frequency standard, I_s is almost equal to I_t and equation (5) gives :

$$\sigma_y(\tau) \simeq \frac{1}{Q_\ell I_s^{1/2}} \; \tau^{-1/2} \quad . \tag{6}$$

The frequency stability is thus determined by the quantity $Q_\ell \, I_s^{1/2}$. It has been shown than it can be increased significantly, for a given interaction length, in an optically pumped cesium beam frequency standard [11].

Figures 9 and 10 show typical values of the frequency stability measure $\sigma_y(\tau)$ for the cesium beam and the rubidium cell frequency standards of classical design.

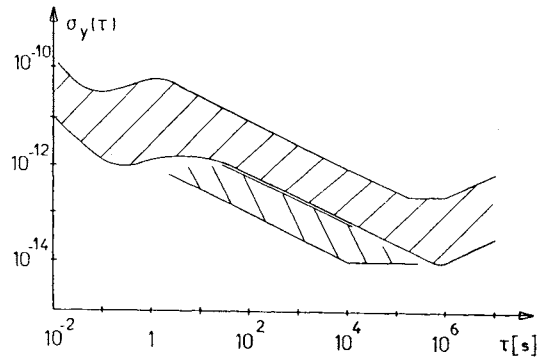


Fig. 9
Typical values of the frequency stability of cesium beam frequency standards. Upper strip : manufactured devices. Lower strip : laboratory devices.

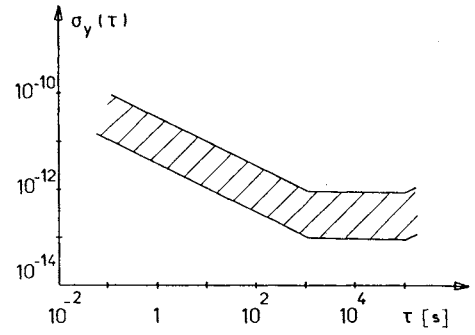


Fig. 10
Typical value of the frequency stability of rubidium cell frequency standards (drift removed).

4.2.2. Active frequency standard

In any oscillator, noise is present in the feedback loop where the signal is generated. Consequently, the frequency of the signal produced is perturbed by white frequency noise. In the hydrogen maser, the perturbing noise is the microwave cavity thermal noise. The related frequency stability is given by :

$$\sigma_y(\tau) = \frac{1}{Q_\ell} \left(\frac{kT}{2P}\right)^{1/2} \tau^{-1/2} \quad , \tag{7}$$

where kT is the thermal energy in the cavity mode and P the power delivered by the atoms to the cavity.

Furthermore, noise adds to the signal produced. This gives white phase noise and :

$$\sigma_y(\tau) = K \left(\frac{kT}{P}\right)^{1/2} \tau^{-1} \quad , \tag{8}$$

where K is a constant depending on the experimental arrangement.

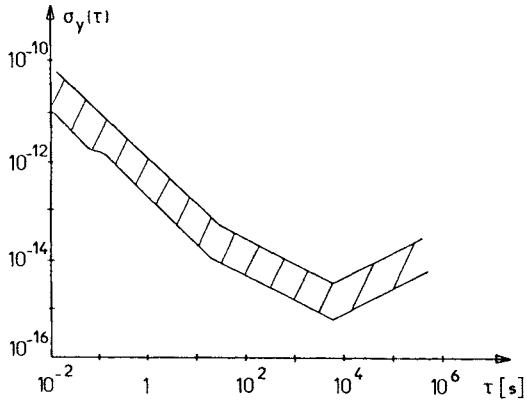


Fig. 11
Typical values of the frequency stability of hydrogen masers.

Figure 11 shows the typical frequency stability of hydrogen masers of classical design. In principle, the frequency stability can be improved by decreasing the temperature. Attempts in this direction have been made since a long time. It is only a few years ago that it has been possible to find a most appropriate coating of the hydrogen storage cell [12-14]. This occurs at a fraction of a Kelvin and gives promise of a much improved frequency stability [15].

4.3. Long term frequency stability

4.3.1. Status

For $\tau \gtrsim 1$ day, the frequency stability measure, $\sigma_y(\tau)$, stops to decrease when τ increases. It even becomes an increasing function of τ .

The most documented information on the long term frequency stability concerns the manufactured cesium beam frequency standards, thanks to their use for time keeping. It is now recognized that the long term frequency stability varies as $\tau^{1/2}$, which denotes a random walk of frequency. We have, for $\tau \gtrsim 5$ days : $\sigma_y(\tau) \simeq 4 \times 10^{-17} \tau^{1/2}$, corresponding to $\sigma_y(\tau) \simeq 2.2 \times 10^{-13}$ for τ = 1 year. However, these

frequency standards may also show a frequency drift with an order of magnitude of 10^{-15} per day. This systematic frequency deviation may dominate the random frequency fluctuations for τ larger than several hundred days.

Laboratory cesium beam frequency standards have a better long term stability since their systematic frequency offsets are measured from time to time. It follows that the frequency stability of TAI is 2×10^{-14} about over a period of several years [16].

The long term frequency stability of the rubidium cell frequency standard is limited by a frequency drift comprised between 1 and a few 10^{-11} per month. It is inherent to the structure of the device, with the reserve of rubidium attached to the resonance cell, the buffer gas is this cell and with the implementation of the double resonance method.

In the hydrogen maser, the long term frequency stability depends whether the cavity pulling effect is eliminated by a cavity auto-tuning system or not. Units have shown the same level of long term frequency stability as laboratory primary cesium beam frequency standards, i.e. a few 10^{-14} over several years.

4.3.2. Origin of long term frequency changes

There is no theory which relates the long term frequency changes to a fundamental physical effect such as shot noise or thermal noise. Instead, they are the result of slow changes of a number of frequency offsets which perturb the idealized resonance frequency defined by equation (1). Most of these frequency offsets depends on the way the atomic transition is observed and thus on the given type of frequency standard. Their change is determined by ageing effects and by variations of the environmental conditions. The rate of ageing and the sensitivity to external perturbations are directly related to the design choices made.

The physical origin of some of these frequency offsets is common to all types of frequency standards, e.g., the second order Doppler effect. Others are specific to a given type, e.g., the buffet gas frequency shift and the light shift in a rubidium cell. These frequency offsets have been listed and considered in detail elsewhere [3]. In brief, the most significant of them are related to :

i) existence of additional energy levels very close to levels E_1 and E_2. Due to their internal structure, the hydrogen atom and the alkali-like atoms require an applied magnetic field in order to separate a number of otherwise degenerated levels. The magnetic field is thus necessary to approximate the two level system assumed above. But, it shifts quadratically the resonance frequency and this is the cause of the sensitivity to changes of this field. In addition, the amplitude of the neighbouring lines may not be balanced, as in the classical cesium beam frequency standard and this is the origin of the Rabi pulling [17].

ii) atomic motion. It gives rise to a possible residual first order Doppler effect and to an unvoidable, relativistic, second order Doppler effect.

iii) distorsion of the atom wave function. Collisions between alkali or alkali-like atoms as well as collisions with a buffer gas disturb the energy levels and cause a frequency shift. Similarly, collisions with the wall of a containing vessel perturb the atom hyperfine interaction and also give a frequency shift.

iv) state selection. A frequency shift occurs when the atoms are submitted simultaneously to the optical pumping radiation and to the microwave field. This is the light shift, which is specific to the rubidium cell frequency standard.

v) coupling to the interrogation microwave field. As a consequence of this coupling, a sufficiently accurate tuning of the microwave cavity is required. Otherwise, a cavity frequency pulling arises.

vi) interrogation process. In passive frequency standards, the measured resonance frequency is sensitive to the quality of the frequency or phase modulated signal which is necessary to probe the transition. Furthermore, in the cesium beam frequency standard, some of the frequency offsets depend on the amplitude of the microwave field.

A part of the frequency offsets are a decreasing function of the line quality factor, which is thus made as large as possible. Furthermore, in the Ramsey's two oscillatory field technique, a number of frequency offsets are reduced, since they are multiplied by the ratio τ_o/T_o, where τ_o is the duration of each interaction with an oscillatory field and T_o is the time delay between the two successive interactions.

These frequency offsets introduce a degree of coupling between the atoms and the macroscopic world. Although weakly, the output frequency of atomic frequency standards thus depends on ageing effects, on temperature, humidity, ambiant magnetic field...

The degree of coupling can be reduced by improving the design of the classical atomic frequency standards, taking into account progress in the analysis of the sources of frequency offsets and past experience [18].

Changes in the structure of these devices can also be introduced. Such is the case in presently studied optically pumped cesium beam frequency standards. Optical state selection and optical detection ameliorate the beam geometry and the knowledge of the velocity distribution. Moreover, it balances the effect of the neighbouring transitions, which suppresses the Rabi pulling, and it enables to decrease the value of the magnetic field applied and, consequently, the sensitivity to changes of its values.

5. ACCURACY. REPRODUCIBILITY

The accuracy of an atomic frequency standard characterizes the degree to which its output frequency agrees with the value of the unperturbed transition frequency. It is expressed as the cumulative fractional uncertainty of the output frequency relative to that of the unperturbed atom.

The accuracy figure depends on the uncertainty of the measurement or of the estimate of the various frequency offsets affecting a given atomic frequency standard. Typical values are the following :

- laboratory primary cesium beam frequency standards : 10^{-14} to 10^{-13}
- manufactured cesium beam frequency standards : 2×10^{-12} to 3×10^{-11} depending on the model and on the specified range of environmental conditions
- hydrogen maser : 2×10^{-12}.

The concept of accuracy does not apply to the rubidium cell frequency standard. Due to the size and the uncertainty of the buffer gas frequency shift and of the light shift, its frequency is always calibrated against that of another atomic frequency standard.

The reproducibility is the degree to which the frequency standard reproduces its normal output frequency without the need of a calibration against another frequency standard, while the environmental conditions are set to any value of the specified ranges. The reproducibility is equal to, or a little better than the accuracy.

The settability is the degree to which the frequency can be adjusted when setting two units to the same frequency or to a desired frequency offset. Typical values are 10^{-12} for rubidium, 10^{-13} for cesium and 10^{-14} for hydrogen.

Improvement of the accuracy and of the reproducibility is intimately related to a better understanding or to a decrease of the frequency offsets.

6. IMPROVEMENT OF CLASSICAL ATOMIC FREQUENCY STANDARDS

Besides the general interest to lower the cost and to increase the reliability, two different requirements emerge to day. The first one is that of improved performances in the present size of the atomic frequency standards. The second one is a demand of the presently achieved performances, but in a smaller volume. Both medium term and long term frequency stabilities are concerned.

Technical improvement of the present designs of the physical package and of the associated electronic system is still possible, in order to fulfil these goals in all types of frequency standards considered here. Moreover, use of semiconductor lasers, at 780 nm or 795 nm for the rubidium cell frequency standard and at 852 nm or 894 nm for the optically pumped cesium beam frequency standard are able to improve the shot noise limited frequency stability. Improvement of the long term frequency stability has already been obtained from a better understanding of frequency shifts such as the Rabi pulling effect. Furthermore, results obtained with optically pumped cesium beams give promise of a further improvement of the long term frequency stability. All these progress will give the opportunity to achieve to day frequency stability accuracy and reproducibility in smaller devices.

7. CONCLUSION

The analysis of the limitations of the classical frequency standards has motivated attempts to realize as closely as possible the ideal conditions of observation of the atoms. They are :

- large value of the line quality factor. This enhances the shot noise or the thermal noise limited frequency stability and reduces some of the frequency offsets,
- large value of the signal to noise ratio. This is beneficial to the medium term frequency stability,
- slow motion. It reduces the Doppler effect accordingly. It may increase the interaction time and it may reduce the frequency of collisions. The limiting case is that of an atom or an ion almost at rest at the center of an appropriate trap. Slowing down can be achieved either by thermodynamical means or the

method of radiation cooling,

- observation of a single atom or ion. This avoids any collective effect and the relevant line broadening and frequency shifts,

- use of atoms or ions having the simplest possible structure, in order to approximate as well as possible a two-level system. Eventually, transitions where frequency is much less sensitive to the value of the magnetic field, such as in magnesium, can be used.

These conditions may not be all compatible. They are experienced to various degrees in the advanced frequency standards.

REFERENCES

(1) H. Hellwig, Proc. of the IEEE 63 (1975), 212

(2) C. Audoin and J. Vanier, Journal of Physics E : Scientific Instruments 9 (1976), 697

(3) J. Vanier and C. Audoin in "The Quantum Physics of Atomic Frequency Standards". 2 volumes, Adam Hilger, Bristol (1989)

(4) L. Essen and J.V.L. Parry, Nature 176 (1955), 280

(5) N.F. Ramsey, Physical Review 78 (1956), 695

(6) M. Arditi and J.L. Picqué, Journal de Physique-Lettres 41 (1980), L 379

(7) H.M. Goldenberg, D. Kleppner and N.F. Ramsey, Physical Review 123 (1961), 530

(8) R.J. Carpenter, E.C. Beaty, P.L. Bender, S. Saito and R.O. Stone, IRE Transactions on Instrumentation I-9 (1960), 132

(9) M. Arditi, IRE Transactions on Military Electronics MIL-4 (1960), 25

(10) J.C. Camparo and R.F. Frueholz, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control UFFC-34 (1987), 607

(11) V. Candelier, V. Giordano, A. Hamel, G. Théobald, P. Cérez and C. Audoin, Applied Physics B, 49 (1989), 365

(12) H.F. Hess, G.P. Kochanski, J.M. Doyle, T.J. Greytack and D. Kleppner, Physical Review A 34 (1986), 1602

(13) M.D. Hürliman, W.N. Hardy, A.J. Berlinsky and R.W. Cline, Physical Review A 34 (1986), 1605

(14) R.F.C. Vessot, E.M. Mattison, R.L. Walsworth, I.F. Silvera, H.P. Godfried and C.C. Agosta, Proc. of the 40th Annual Symposium on Frequency Control (1986), 413

(15) R.F.C. Vessot, E.M. Mattison, R.L. Walsworth, I.F. Silvera and M.L. Dingus, this issue

(16) C. Thomas (1990), private communication

(17) A. de Marchi, Proc. of the 41th Annual Frequency Control Symposium (1987), 53

(18) A. de Marchi (1990), this issue

Session FS 01 Number 2

Advanced Atomic Frequency Standards: Where are we heading?

Jacques Vanier
Laboratory for Basic Standards
National Research Council of Canada
Ottawa, Canada
K1A OR6

Improvements in the accuracy and frequency stability of atomic frequency standards are partly related to enhancement in line-quality factors and signal-to-noise ratio. On the other hand the resonance frequency of interest in atoms or molecules is a function of both the internal variables which are fixed by nature, and the external degrees of freedom which are fixed by the environmental factors. In practice the bulk of the work in this field is aimed at making the frequency independent of the environment as much as possible and at the same time in finding practical systems which will lead to a larger line-quality factor and a greater signal-to-noise ratio.

Further work on so-called classical frequency standards will certainly lead to some improvements in these parameters but it appears that orders of magnitude improvements although possible will be difficult to achieve. Consequently some researchers have turned towards daring new approaches. Among these are laser cooling of atoms and ions, trapping, use of the Ramsey fringes approach in the optical region and the use of single ions. Frequency multiplication chains from the microwave to the visible will permit use of optical frequency standards for practical purposes, such as time-keeping.

The presentation will examine some of the more promising concepts and approaches used or proposed and will attempt to outline from present knowledge what can be expected in the not too distant future in the field of precise and accurate frequency standards.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-01 Number 3

CHARACTERIZATION AND MEASUREMENT OF FREQUENCY STABILITY : TWENTY FIVE YEARS OF PROGRESS

J. Rutman* and J. Groslambert**

* Fédération des Industries Electriques et Electroniques (F.I.E.E.) - 11, rue Hamelin - 75783 Paris Cedex 16 - France

**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

Precision time and frequency sources have known tremendous developments since the early 50'. Improvements of several orders of magnitude have been achieved both in frequency accuracy and stability. This is true for atomic frequency standards (cesium beam, rubidium cell, hydrogen maser) but the progress of quartz crystal oscillators have also been very impressive with the advent of the BVA. Besides the "classical" RF and microwaves regions, frequency standards in the infrared and visible regions have been developed with the advent of saturated absorption stabilized lasers (CO_2, HeNe, ...).

Therefore, a great "metrological challenge" has been to develop both mathematical (but physically realistic) parameters for the characterization of their frequency and phase instabilities and related experimental test-sets for the measurement of these parameters.

These developments were absolutely needed in order to compare unambiguously the performance between the various types of standards, but also in order to assess the performance. The diverse applications (such as fundamental metrology, radiocommunications, radars, space, requirements in terms of oscillator performance.

This paper will review the parallel progress of theoretical characterization and of experimental test-sets (the latter being also influenced by the progress of electronics technology). As far as parameters are concerned emphasis will be put on the evolution of ideas and on key milestones rather than on mathematical developments that can be found in many references.

The first part of the paper is devoted to mathematical parameters. Starting with an oscillator output signal model including random phase and frequency noises, emphasis will first be put on the dichotomy between characterization in the (Fourier) frequency domain and in the (sample) time domain. Several kinds of spectral densities andof variances will be introduced respectively. In the frequency domain, the usefulness of the power law model will be demonstrated with examples of well-known noises incountered in real oscillators (e.g. flicker frequency noise, white phase noise, etc.). In the time domain, the many variances that have been proposed (and more or less used) during the last 25-30 years will be reviewed with their respective advantages and limitations : the variance, sample variance with special emphasis on the two-adjacent-sample "Allan" variance (1966), Hadamard variance (1971), time stability (1971), high-pass and band-pass variances (1974), modified variance (1976) and modified Allan variance (1981). Structure functions introduced in 1976 give a unified picture of the problem. A specific measurement sequence in the time domain looking like a stepwise function is associated to each variance.

A key point is the integral structure of the mathematical relationship that allows one to calculate any of the above variances from the spectral densities. A transfer function which is the Fourier transform of the above stepwise functions appears in the integrals. It will be shown that its general shape (more or less selective) and its behavior for very low Fourier frequencies give a good understanding of all the specific properties of the related variance. Examples will be given.

Application to the frequency domain power laws (f-2, f-1, fo, etc.) will show power laws also in the time domain ($\tau^{-3/2}$, τ^{-1}, $\tau^{-1/2}$, τ^{0}, etc.) with "more or less convergent" variances (i.e. convergent for spectral densities with f^{-n} and increasing n).

In conclusion of the first part of the paper, it will be emphasized that the transfer function concept has played a key role in the general understanding of the problem but has also lead to demonstrate new possibilities of experimental test-sets, e.g. the Hadamard variance measured with a counter in the time domain allows one to estimate (frequency domain) spectral densities at very low Fourier frequencies, and the high-pass variance measured by high-pass filtering of demodulated phase noise gives easy estimates of the well-known and widely used time domain Allan variance.

The contributions from many individuals in USA, Canada, Europe and Japan must be recognized together with the incitating and coordinating role of organizations such as NASA, CCIR and the IEEE which has issued in 1989 an "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology (IEEE Std 1139).

Some important steps of technological progress have been realized in the last years in frequency and time domain measurements. Experimental test-sets that have been developed by several laboratories (especially in USA and France) showing the past few decades, will now be described in frequency and time domain.

I - FREQUENCY DOMAIN MEASUREMENTS

In frequency domain measurements before 1967 only microwaves sources was measured. In this period the characterization was not made in spectral density terms but in deviation frequency terms. In fact the used test sets was adapted to characterize microwaves sources but did not work correctly in the radiofrequency domain. This test set is given on Fig. 1.

The signal of the source to be tested attacks a bridge mount consisting in two arms : a direct arm and a second one which feed through a resonant circuit. The resonance frequency of this circuit is tuned at the frequency of the source to be tested. This resonant circuit transforms the frequency fluctuations of the source to phase fluctuations which are detected by a mixer. The phase of the signals issued from the two arms are adjust at 90°, in this case the mixer works like a phasemeter.

The output voltage of the mixer after filtering is proportional to the frequency fluctuations of the source for Fourier frequencies inside the bandwidth of the resonant circuit. Using a resonant circuit with a quality factor of 10 000 the deviation frequencies of a 10 GHz source can be measured up to 500 kHz. Today the same test set is used but the results are given in $S\Delta\nu(f)$ the spectral density of frequency fluctuations or in $Sy(f)$ the spectral density of relative frequency fluctuations. This test set working at 5 MHz for the same demodulated frequency range requires a quality factor which is very low. The sensibility, which is proportional to the quality factor, is very poor that explains this test set is not used in radiofrequency to measure sources.

In 1972 appears a new method to measure the phase and frequency fluctuations of an oscillator. This method was developed at the NBS [1] : this is the phase lock loop demodulator (Fig. 2).

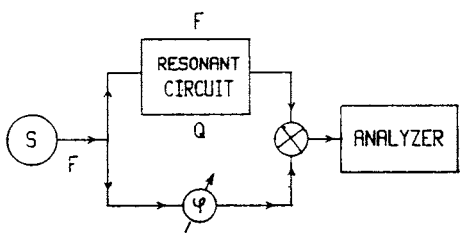


Fig. 1 : Test set adapted to the measurement of microwave source

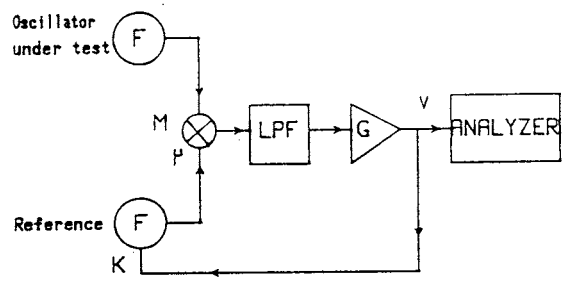


Fig. 2 : Phase lock loop demodulator

This system needs a reference oscillator at the same frequency than the oscillator to be measured. The signals issued from the two oscillators are sent into a mixer which works like a phasemeter with a μ sensibility in V/rad. The output signal of the mixer is low pass filtered to obtain only the low frequency components proportional to the phase difference between the two oscillators. After a G amplification this signal is sent back to the electronic tuning input (K sensitivity in Hz/V) of the reference oscillator.

The mathematical study of this phase lock loop reveals that :
- for Fourier frequencies lower than the product KGμ the error signal V is proportional to the frequency fluctuations of the two oscillators.
- for Fourier frequencies upper than the product KGμ this error signal is proportional to the phase fluctuations of the two oscillators.

Any oscillators at any frequencies can be characterize with this method with only one condition : the needed reference oscillator must be better than the oscillator to be tested. At the limit two identical oscillators can be measured. In this case the spectrum of each oscillator is obtained by reducing the result of the measurement by 3 dB. The performances of this system is especially limited at low frequencies by the flicker noise produces by the mixer and at high frequencies by the thermal noise voltage of the amplifier. These noises are converted in phase fluctuations taking into account the sensitivity of the mixer in rad/V. The noise of the phase lock loop measurement system is given on Fig. 3

In 1981, a dual phase lock loop system was developed at the LPMO [2]. The schematic diagram of this test set is given on Fig. 4.

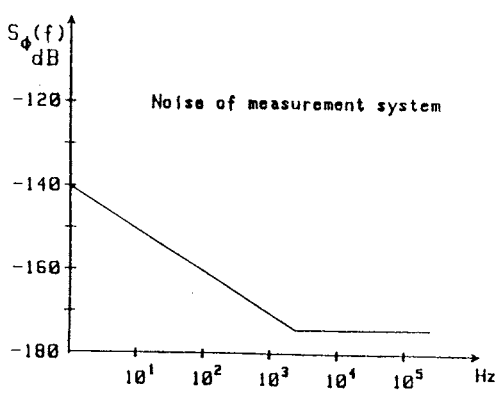


Fig. 3 : Noise of the phase lock loop measurement system

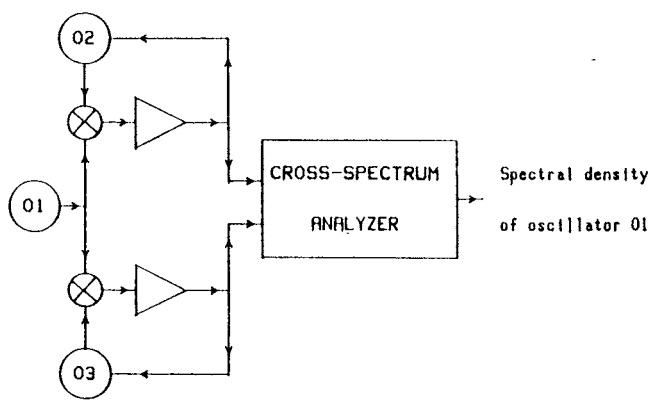


Fig. 4 : Dual phase lock loop system

This system uses three oscillators O_1, O_2, O_3, at the same frequency and two phase lock loops to lock the two oscillators O_2 and O_3 on the oscillator O_1.

The cross-spectrum of the two error signals gives the spectral density of the oscillator O_1. The noises of oscillators O_2 and O_3 appear at the output as non-correlated and are therefore rejected. A rejection factor of the order of 20 dB can be obtained. Consequently, this method gives the possibility to test one oscillator using two reference oscillators which have the same noise spectral quality or even worse than the oscillator under test. The noise of the system that is mixer and amplifier is also eliminated in the decorrelation process.

This method improves the performances of the single phase lock loop system by 10 dB.

II - TIME DOMAIN MEASUREMENTS

Before 1967 the second was defined by the 1/86400 part of the mean solar day. It was then related to the period of the earth rotation. Timekeepers constructed around quartz oscillators were scrupulously keeped far away from external perturbations in observatory cellars. These oscillators can detect the fluctuations of the earth rotation period.

The first cesium clock came out in 1955 with a stability of 10^{-9}. It took ten years for the technology to mature to the point that the performances of this type of atomic clock reach the stability of 10^{-12}, which is the stability of the commercial cesium now.

Since 1967 the second is defined by the duration of 9192631770 periods of the radiation corresponding to the transition between two hyperfine levels of the fundamental state of Cs133 atom.

To protect against a failure of an unique atomic clock standard a mean atomic time had been determinated from several atomic clocks.

In 1965 TOLMAN founds a method to compare distant atomic clocks by using a synchronisation pulse carried by a TV signal. The first comparisons using this method have been done between PRAGUE and POSTDAM. The time measurement resolution was one µs which gives a relative resolution of 10^{-11} per day.

Now with optimized receivers this method gives a possibility of a 20 ns measurement resolution corresponding to a relative resolution of 2.10^{-13} per day.

The device which is used in short term stability measurements is the phase lock loop demodulator with some extra circuits (Fig. 5).

As it was shown in frequency domain measurements section in the phase lock loop system the error voltage is proportional to the frequency fluctuations of the oscillator under test for Fourier frequency inside the bandwidth of the servoloop. Three methods can be used to treat this signal and to obtain the Allan variance which characterize the properties of an oscillator in time domain :

a) In the first method the error signal of the servoloop is converted in frequency by a voltage to frequency converter. By measuring the converter output signal with a counter, the Allan variance can be calculated on the samples.

b) In the second method the error voltage of the servoloop is high pass filtered and a rms voltmeter measures the output of the filter. The value given by the voltmeter is an estimation of the Allan variance.

c) In the third method, the Fast Fourier Transform algorithm found by Cooley and Tukey in 1965 [3, 4] gives the spectral density of the frequency fluctuations. By using mathematical relations between frequency domain and time domain, the Allan variance can be deduced by calculation. The Fast Fourier Transform makes obsolete the Hadamard variance which is no more used.

Two clocks which cannot be locked one to the other can be tested by using the Mixer System developed by the NBS. A schematic diagram of this Dual Mixer System is given on Fig. 6.

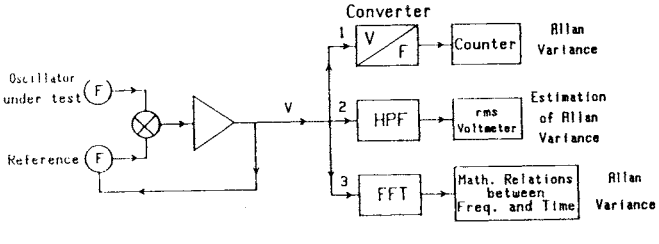


Fig. 5 : Schematic test set for short term stability measurements

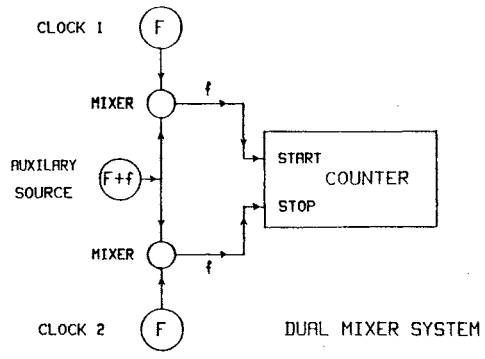


Fig. 6 : Dual mixer system

An auxiliary source slightly shifted from the frequency of the clocks is used to obtain a beat at the output of each mixer. A counter measures the time between the zero crossing of the two beat signals sent to the start and stop inputs of a counter. The Allan variance can be calculated from the samples furnishes by the counter.

The noise of this system is given by $10^{-13}/\tau$.

CONCLUSION

In these twenty five last years important progresses have been made on the resolution of the test sets used in the measurements of spectral densities and frequency stabilities. Nowadays the normalisation of the use of Allan variance to characterize frequency stability in time domain allows one to compare different measurements made by anyone with the same concept. In frequency domain the phase lock loop systems have permitted to measure the radiofrequency sources.

An improvement on the performances of this device has been obtained with the dual phase lock loop systems which have given the possibility to test an oscillator without the limitations of the reference oscillator performances. In general the continuous progress of the electronics technology which permits to obtain better sources needs the parallel continuous improvements of the performances of the test sets.

REFERENCES

[1] J.H. Shoaf, D. Halford, A.S. Risley, "Frequency stability specifications and measurement : high frequency and microwave signals", NBS Technical note 632, january 1973.

[2] D. Fest, J. Groslambert and J.J. Gagnepain, "Individual characterization of an oscillator by means of crosscorrelation or cross-variance method", IEEE Transactions on Instrumentation and Measurements, vol. IM-32, n° 3, september 1983.

[3] W.T. Cochram, J.W. Cooley et al., "What is the Fast Fourier Transform ?", Proceedings of the IEEE, vol. 55, n° 10, october 1967.

[4] J.W. Cooley et al., "Historical notes on theFast Fourier Transform", Proceedings of the IEEE, vol. 55, n° 10, october 1967.

[5] D.W. Allan, "Report on NBS Dual Mixer Time Difference System (DMTD) built for time-domain measurements", NBSIR - 75 - 827, january 1976.

Frequency Standards

FS-02 Review Papers

Chairman: J. Vanier

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS 02 Number 1

PRECISE DOPPLER TRACKING OF INTERPLANETARY SPACECRAFT

Bruno BERTOTTI
Dipartimento di Fisica Nucleare e Teorica
Universita Degli Studi di Pavia
Via Bassi 6
I - 27100 PAVIA (Italia)

The hydrogen masers available at the radio stations for interplanetary communications allow a variety of experiment to measure the structure of space-time and its gravitational field: for example, a microwave phase-locked loop between the ground station and a distant spacecraft is affected, besides the obvious changes in the Doppler effect produces by near-by bodies, by impinging gravitational waves with frequencies of the order of the reciprocal of the round trip light-time. This method is the only one at present available for detection of low frequency gravitational waves, in the mHz band. The state of this research program and the prospects which may be provided by the space mission CASSINI (to be lanched in 1996) are discussed. CASSINI may have, in addition to the baseline X-band coherent radio system, a K_a-band, which will allow elimination (by means of a five ways link), or great reduction, of the noise due to the plasma along the beam. This will make it possible gravitational experiments not near opposition periods and a good measurement of the mass of an asteroid. The project aims at reaching an overall frequency stability better than a part in 10^{15} over integration times of the order of 1000 seconds.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS 02 Number 2

TIME AND RELATIVITY:

EXPERIMENTAL AND APPLIED CURVED SPACE-TIME CHRONOMETRY

Carroll O. ALLEY
Department of Physics and Astronomy
University of Maryland
College Park, Maryland 20742

The fundamental physical insight of Albert Einstein about the concept of time: that the time of an event is not absolute, but is relative to the observer who measures it (first fully recognized by him in Bern in 1905), has been central to the development of relativity theory. It allows the measured velocity of light to be isotropic with the same speed c for all inertial observers, thus making possible the measurement of distances by clock and light pulses. This common measure for time and distance is expressed by the invariance (for different such observers) of the square of the line element (increment of proper time for a clock) of the space-time metric, as first formulated by Herman Minkowski in 1909. Starting from the "Principle of Equivalence" between physical phenomena in an accelerated elevator and those in a gravitational field, Einstein developed a curved space-time theory of gravity. The metric coefficients become functions of position and time and are to be determined as solutions of field equations which relate curvature to mass-energy. Time continues to be central and the theory is best characterized as "Riemannian Space-Time Chronometry". The recognition (and correct implementation!) of the predicted effects of this theory is vital for the proper functioning of modern engineering systems relying on very stable atomic clocks, such as the Global Positioning System.

There exists a viable alternative new metric theory of gravity as curved space-time developed by Hüseyin Yilmaz. It treats satisfactorily all of the known "general relativistic" effects, and has features which are superior to the old theory. Some of these are the following.

Even in accelerated frames of reference, like the surface of the rotating Earth, there is a local kinematic correspondence to the metric of special relativity. There is a second order correspondence to Newtonian gravitation theory which allows simple N-body relativistic equations of motion. In addition the new theory is compatible with the principles of quantum mechanics.

Kinematical experiments with laser light pulses and transported atomic clocks may provide crucial tests of the two theories. Measurements comparing the one-way speed of light between the East-West and West-East directions are in progress, and other experiments are possible.

Posters

PO-01 Short presentation of each poster by the chairman

Chairman: D.J.E. Knight

Session PO-01 Number 1

Studies of a laser diode pumped Rb frequency standard

J.T. Liu, P. Thomann, Oscilloquartz S.A. Neuchâtel
L. Zhang, G. Busca, Observatoire de Neuchâtel

Introduction: The development of laser diode pumped atomic frequency standards is extremely attractive because the laser diode combines single mode, high power density, narrow linewidth and tunability. It is expected that use of the laser diode to replace the r.f. lamp will not only improve the short term stability but also improve the long term stability. Our experiments were only concerned with the basic performance measurements of the laser diode and the laser diode pumped Rb frequency standard in order to make a practical clock. This paper presents experimental results.

1. Experimental Set-up: Fig.1. shows an experimental set-up of the laser pumped 87Rb atomic frequency standard. The laser beam is split in two beams. One beam passes through a natural Rb cell ("reference cell"). The light transmitted by the reference cell is used to lock the laser frequency to the D_2 resonance line of Rb 5S1/2 - 5P3/2 by an electronic servo system. The natural Rb cell is temperature controlled. The other beam pumps an isotopic 87Rb absorption cell contained in the cavity ("clock cell"). With the electronic offset of the laser diode servo system, one can set the laser frequency at any point of the absorption curve of the isotopic 87Rb cell. The oscilloscope is used to observe the absorption curve. PM6652C counter, HP9826 calculator and SR530 lock-in amplifier are used to measure noise, signal beat frequency and short term stability.

2. Transmission coefficient: By sweeping the injection current of the laser, the absorption spectra of the clock cell were measured for different cavity temperatures and different laser power densities. Fig.2 shows three curves for different power densities at the same cavity temperature. Curve 1 is at high power and curve 3 lower. The transmission coefficient is the ratio of the transmitted power on resonance to the transmitted power off resonance. From fig.2, the transmission coefficients are 39%, 31,5% and 17% respectively for the three curves at F=2 and 65°C indicating saturation of the optical pumping at high laser power density. The transmission coefficient also changes with cavity temperature. According to transmission coefficient, the cavity temperature and laser power can be chosen in order to get the largest r.f resonance signal. Because the absorption of the isotopic 87Rb cell is more efficient especially at low temperature, we chose the isotopic 87Rb absorption cell for our experiments.

3. Temperature coefficient: The temperature coefficient of the Rb clock frequency has been extensively measured [1] with r.f. discharge pumping. In this case, the light shift is very serious. The following experiments were performed in order to evaluate the temperature coefficient in a laser-pumped clock. A laser diode simply replaces the r.f. discharge lamp of the Rb clock. The transmitted light from the isotopic 87Rb cell contained in the cavity of the Rb clock is fed into the laser servo system to lock laser frequency to the $5S_{1/2}$ (F=1)-$5P_{3/2}$ transition. The beat frequency between Rb clock and a reference Cs clock (at 5MHz) is measured as a function of clock cell temperature. Hence the temperature coefficient of fractional frequency is obtained. The results, called present "1" (see table 1), are quite different from the values reported in ref.[1]. The same measurements were then repeated with the laser locked to the temperature stabilised reference cell instead of the clock cell. Thus the laser frequency could not be affected by temperature changes of the clock cell. The results appear in table 1 (present measurement "2") and are in close agreement with the values found in ref. [1]. The results (present "1" and "2") differ by a nearly constant number, independent of gas mixture. The difference is 3.6×10^{-10}/°C for the isotopic 87Rb cell contained in the cavity and is independent of the N2/CH4 ratio in the buffer gas mixture. We call this number the extra temperature coefficient. In order to check the origin of this extra temperature coefficient, we compared the laser frequency, while locked to the clock cell, with the reference cell frequency, and found that the laser frequency changed with clock cell temperature by -9MHz/K for the 5 S1/2 (F=1) -5P 3/2 transition. Lightshift measurements presented in the next section of this paper show that the microwave frequency change with laser detuning is - 0.26 Hz/MHz, i.e. -3.8×10^{-11}/MHz. Thus the laser detuning contribution to the extra temperature coefficient can be calculated as:

$$\frac{\partial \nu_{Rb\,clock}}{\partial T_{cavity}} = \frac{\partial \nu_{Rb}}{\partial \nu_{laser}} \cdot \frac{\partial \nu_{laser}}{\partial T_{cavity}} = -3.8\times10^{-11}\text{/MHzx}(-9\text{MHz/K}) \quad = 3{,}4 \times 10^{-10}\text{/K}$$

It is very close to the measured values 3.6×10^{-10}/K. So we believe the extra temperature coefficient is essentially caused by the laser frequency change with the cavity temperature.

4. Light shift measurement With laser diode replacing the r.f. discharge lamp, it might be expected to estimate light shift more precisely because of the narrower spectral linewdith and the tunability of the laser. Fig.1 shows the experimental set up of the light shift measurement. The laser frequency is locked to D2 resonance of the natural Rb reference cell 5S1/2 - 5P3/2 by the electronic servo system. The offset of the laser diode servo system is used to set the laser frequency at any point of the absorption curve of the clock cell. For 5S1/2 (F=1) - 5P3/2 transition, the resonance frequency of the isotopic 87Rb cell is 75 MHz higher than the resonance frequency of the natural Rb cell, for F=2 the former is less than 5 MHz away from the latter. Fig. 3 and Fig. 4 show the results of the light shift measurements. The behavior of each curve is very similar to the results of C.H. Volk et al [2]. At lower power density, as expected, light shift is linear and its magnitude and sign depend on the laser power density and frequency detuning. If the laser power continuously increases, however, the light shift goes up to an extremum, then down to a nearly constant level. Mr. C.H. Volk [3] has suggested an explanation for this behavior. The range of laser power density in which the light shift is nearly constant (so-called "saturation range") is 230-575μw/cm² for F=1, 277-740μw/cm² for F=2 respectively. At saturation range the light shift induced by laser power density is the smallest for each laser frequency detuning. Obviously if the laser diode operates in the saturation range, the long term stability of the Rb clock should improve. The value of the light shift coefficient (fractional shift of clock frequency for 1% change in laser intensity) is given in fig 3 and 4 (average coefficient over saturation range). The values are smaller than the ones quoted in ref [2], possibly because, in the present experiment, an isotopic cell was used instead of a natural Rb cell.

5. Signal to noise and short term stability Using the SR530 lock-in amplifier, the signal, noise, linewidth of the microwave resonance were measured. The frequency stability can be expressed as

$$\sigma_y(\tau) = \frac{0{,}2}{Q(S/N)}\,\tau^{-1/2} \quad [4]$$

$$Q = \nu_{Rb}/\delta = 6.835.10^9/1040 = 6.6.10^6$$

($\delta = \sqrt{3}x$ spacing between the two extrema of the microwave discriminator pattern, S= rms error signal photocurrent at modulation frequency and at maximum of discriminator pattern = .18μA; N= rms photocurrent noise density at modulation frequency= 133pA.Hz$^{1/2}$

These values lead to a calculated square root Allan variance of $2.2.10^{-11}\,\tau^{-1/2}$. It should be noted that the measured intensity noise of the laser is the most important noise source limiting the stability of the clock at the present stage.

6. Conclusion In the present work, we have studied the laser pumped Rb frequency standard. These results are useful to make a practical laser pumped Rb clock. Let the laser diode operate in the saturation range, the smaller light shift can be gotten. The long term stability and the frequency accuracy of the Rb clock could be improved.

Acknowledgement. The authors would like to express their gratitude to Mr. F. Hadorn for his laser diode servo system.

References.

[1] J. Vanier and E. Audoin: the Quantum Physics of Atomic Frequency Standards Vol1, p. 349 and references therein (Hilger. 1989)
[2] C.H.Volk, J.C. Comparo and R.P. Frueholz: Proceedings of the Thirteenth Annual Precise Time and Time Interval (PTTI) Application and Planning Meeting p. 631.(1981).
[3] Ref 2 p. 554
[4] J. Vanier, L.G. Bernier, IEEE Transactions on Instr. and Meas. IM-30, no4, 277-282 (1981).

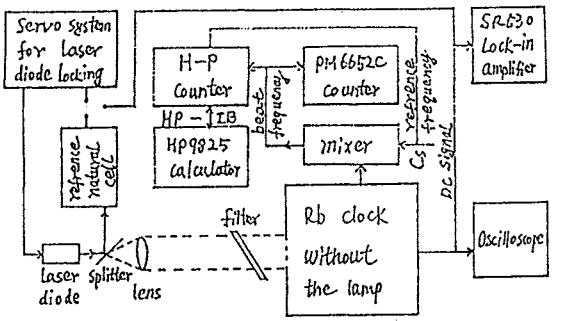


Fig. 1 Experimental setup diagram of the laser diode pumped Rb frequency standard

Table 1 Temperature coefficient of fractional frequency

Temperature coefficient	Pure N_2 k^{-1} mb^{-1}	pure CH_4 k^{-1} mb^{-1}	mixture 50% k^{-1} mb^{-1}
From ref. 1	6.0×10^{-11}	-6.6×10^{-11}	-3.3×10^{-12}
present "1" measurement	1.11×10^{-10}	-1.88×10^{-11}	4.49×10^{-11}
present "2" measurement	6.8×10^{-11}	-5.07×10^{-11}	8.55×10^{-12}

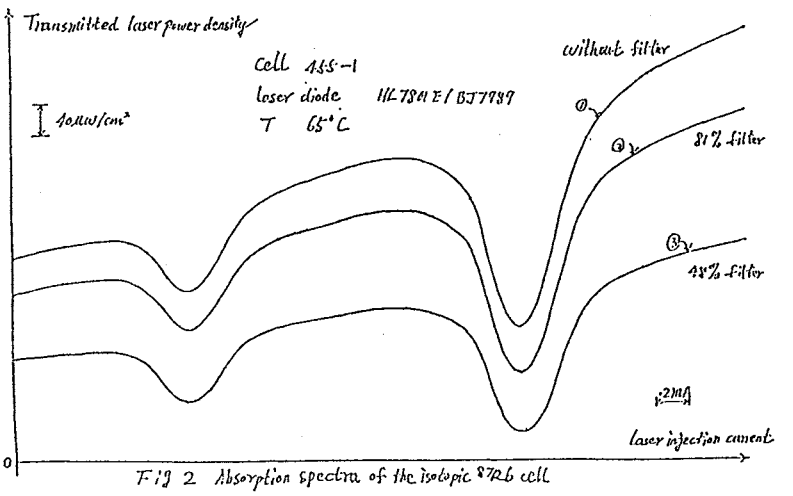


Fig 2 Absorption spectra of the isotopic ^{87}Rb cell

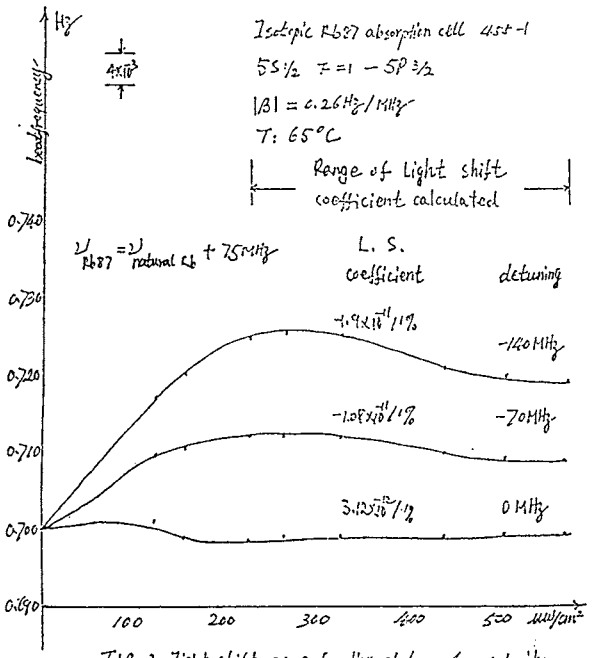


Fig.3 Light shift as a function of laser power density for various laser frequency detunings

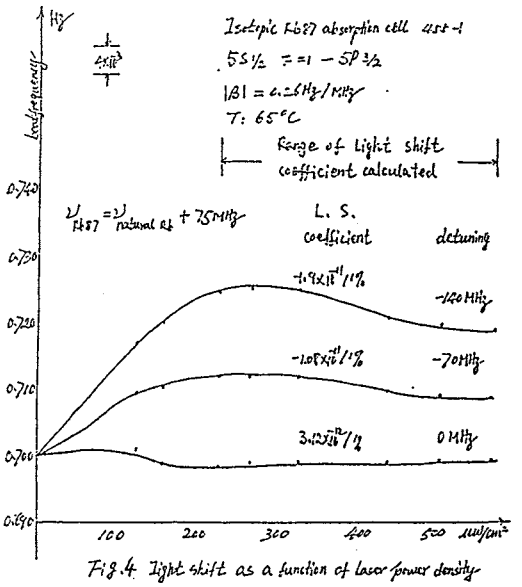


Fig.4 Light shift as a function of laser power density for various laser frequency detunings

Session PO-01 Number 2

ESTIMATION OF SOME FREQUENCY SHIFTS IN AN OPTICALLY PUMPED CLOCK TAKING INTO ACCOUNT MODULATION PARAMETERS

G.D. Rovera*, A. de Marchi+ and E. de Clercq*

* Laboratoire Primaire du Temps et des Fréquences, Observatoire de Paris
61, avenue de l'Observatoire, 75014 Paris, FRANCE.

+ Dipartimento di Elettronica e Automatica
Universita di Ancona, Ancona, ITALY.

It has been shown that not only the short term stability of a standard but also its frequency shifts depends on the type and parameters of the modulation used in the servo-loop [1,2]. In this paper such effects are analyzed for primary laboratory standards using optical pumping and a slow square wave frequency modulation.

Only the most relevant systematic effects in an optically pumped standard, i.e. : 2nd order Doppler effect (D_2), Cavity pulling (C) and phase shift (ϕ) are considered here, together with the short term stability. The formulas used and the notations are the same as in Ref. 1. The relative shifts and the short term Allan variance are given by :

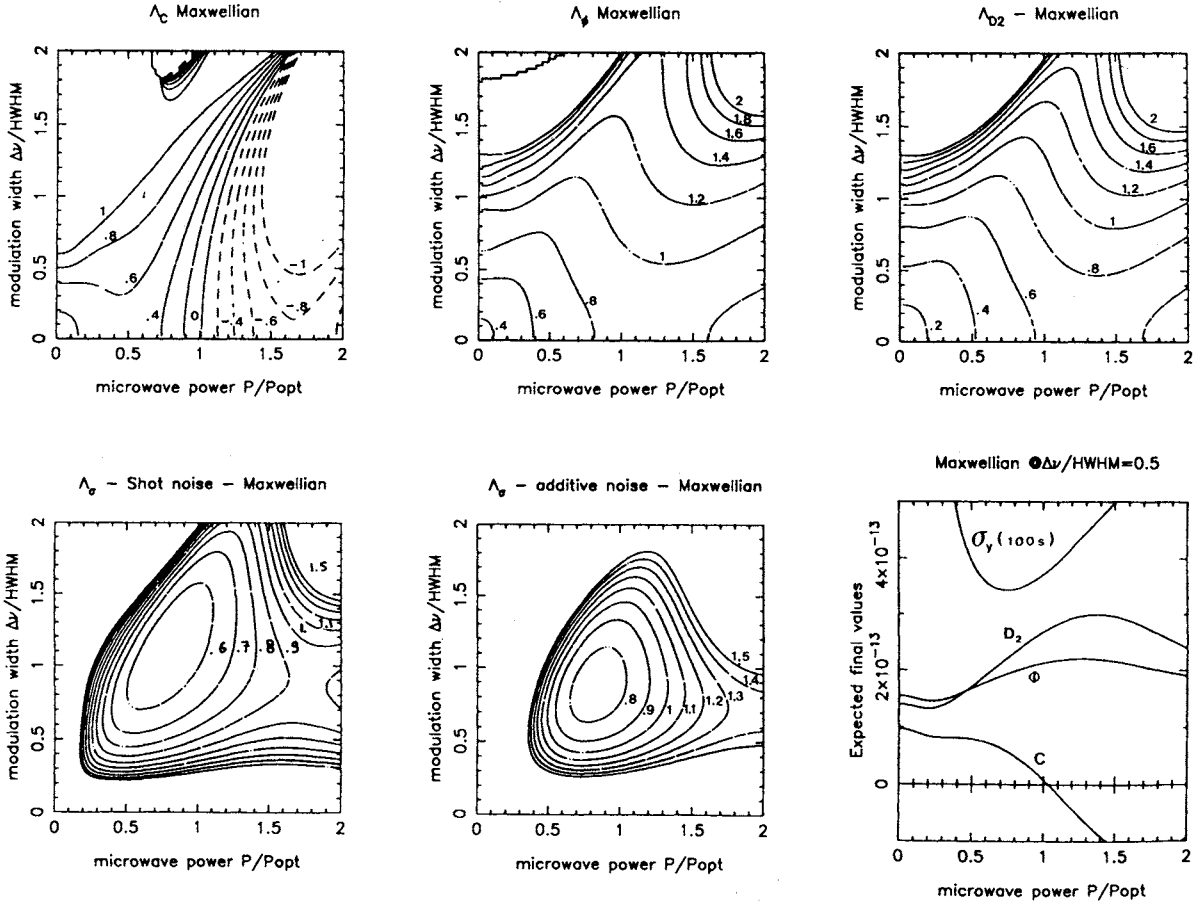
$$\left(\frac{\delta\upsilon}{\upsilon}\right)_{D_2} = -\Lambda_{D_2}\frac{\langle v^2\rangle}{2.c^2} \quad (1)$$

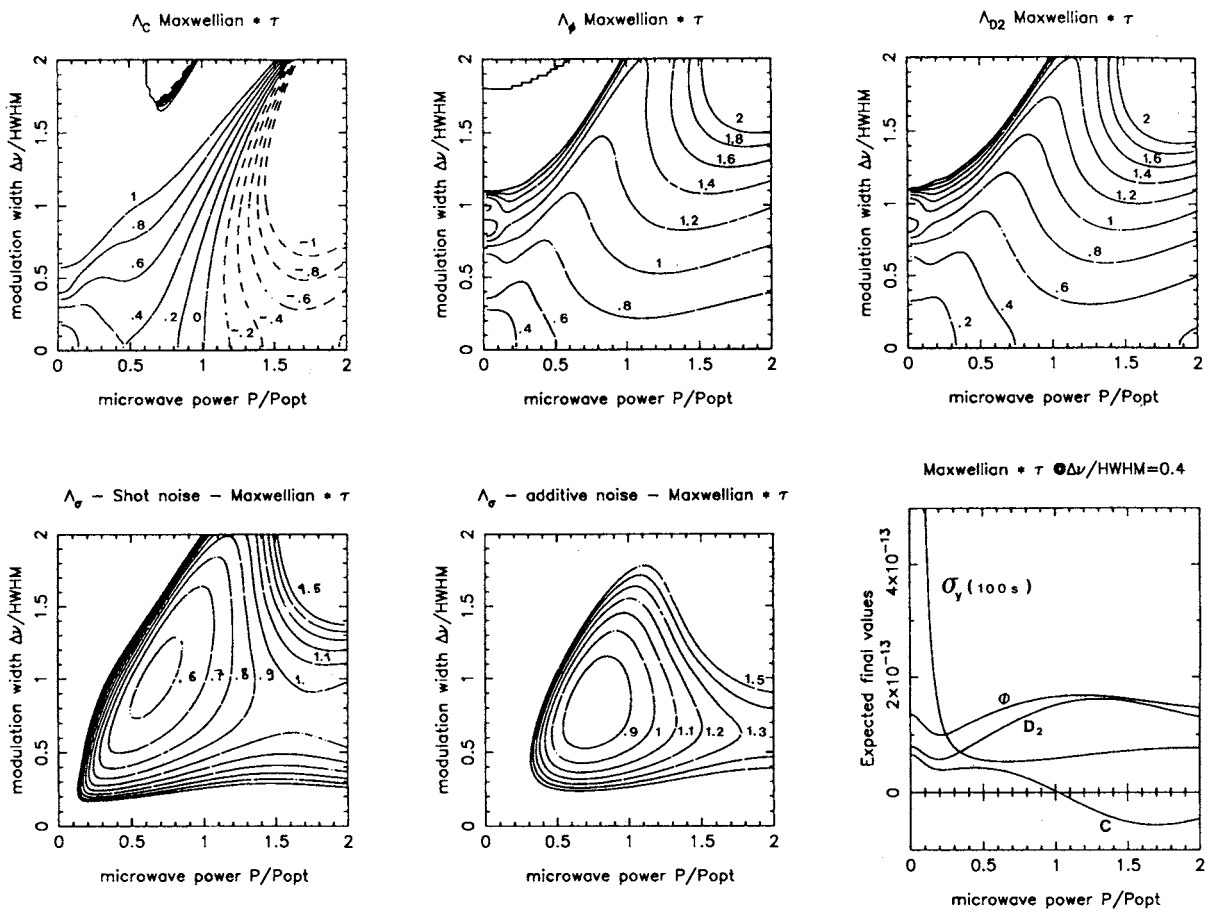
$$\left(\frac{\delta\upsilon}{\upsilon}\right)_{C} = \Lambda_{C}\left(\frac{Q_c}{Q}\right)^2\frac{\delta\upsilon_c}{\upsilon} \quad (2)$$

$$\left(\frac{\delta\upsilon}{\upsilon}\right)_{\phi} = \Lambda_{\phi}\frac{\langle\delta\phi\rangle}{\pi Q} \quad (3)$$

$$\sigma_y(\tau) = \Lambda_{\sigma}\frac{1}{2Q\,S/N\sqrt{\tau}} \quad (4)$$

where $\langle\ \rangle$ means average on the velocity distribution, Q_C is the loaded cavity quality factor, Q is the line quality factor measured at optimum Power P_{opt} in CW operation, $\delta\upsilon_C$ is the cavity detuning, $\delta\phi$ is the phase difference between the two interaction zones of the Ramsey cavity and S/N is the signal-to-noise ratio measured in 1Hz bandwidth at P_{opt} on center frequency in CW operation.





The Λ parameters are calculated as a function of the microwave Power P and of the modulation width. P is normalized to the optimum power P_{opt} which maximizes the CW detected signal at resonance and the modulation width is normalized to the width of the Ramsey Peak (HWHM) at P_{opt} in CW operation.

With optical pumping two kinds of Time Of Flight (TOF) distribution $\rho(\tau)$ are expected : Maxwellian distribution for a single laser scheme and Maxwellian weighted by τ for a two lasers scheme using a cycling transition at the detection [3]. Λ parameters are calculated in both cases. The noise type has also to be taken into account in the calculation of $\sigma_y(\tau)$, which is carried out in the two cases of shot noise limitation (noise proportional to the root of the signal) and additive noise limitation (noise independent of the signal). Both cases, and intermediate ones, can be encountered in an optically pumped standard depending on laser scheme, atomic beam intensity and static magnetic field amplitude.

Calculations are performed with the geometrical configuration of the LPTF laboratory standard. The interaction length in each arm of the Ramsey cavity is 10.16 mm, the distance between the centers of the two arms is 1021 mm and the distance between the center of the first arm and the detection region is 1261 mm. These values are close to those of the optically pumped NRLM clock [4]. The TOF distributions are calculated with an oven temperature of 90°C. Results are shown on the figures above.

As we can see the Λ values, except Λ_σ which presents a minimum, decrease continuously when the modulation width decreases. The chosen modulation width, 0.5 for the Maxwellian distribution and 0.4 for the Maxwellian*τ distribution result from a compromise between accuracy and short term stability. Then we have estimated the final shift values as a function of the microwave power with the following parameters: Q_C=3000, $\delta\upsilon_C$=600KHz, $\delta\phi$=50μrad.The stability is given for τ=100s and a constant additive noise. The signal-to-noise ratio is 1800 for the Maxwellian distribution (1 laser scheme) and 10 600 for the Maxwellian*τ distribution(2 lasers scheme).

In short with a slow square wave modulation, a relative modulation width of about 0.5 and 0.4 and a microwave power near P_{opt} we can cancel the cavity pulling while getting a moderate slope of the second order Doppler shift versus the microwave power (less than 10^{-14} for a 10% power variation) and stabilities for 100 s of 4.10^{-13} and 6.10^{-14} depending on the laser scheme.

We acknowledge A. Premoli for first programm developing.

Work supported by BNM and CEE.

REFERENCES

[1] A. de Marchi, G.D. Rovera and A. Premoli, IEEE Trans. Ultras. Ferro. Freq. Control, UFFC-34, 582 (1987).
[2] J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Adam Hilger ed. (1989).
[3] G. Avila, E. de Clercq, M. de Labachellerie and P. Cerez, IEEE Trans. Ins. Meas. IM34, 139 (1985).
[4] S. Ohshima, Y. Nakadan and Y. Koga, IEEE Trans. Ins. Meas., IM37, 409 (1988).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-01 Number 3

March 10, 1990

Optical pumping of cesium in applied external magnetic field: effect of changes in hyperfine coupling

Claude JACQUES

Division de Physique, CNRC, Ottawa, CANADA K1A 0R6

Pierre TREMBLAY, and Normand CYR

Département de génie électrique, Univ. Laval, Québec, CANADA G1K 7P4

Earlier studies have demonstrated that use of narrow linewidth lasers in optical pumping creates coherences. This effect results in less efficient pumping. Broad linewidth lasers excite neighboring levels not desired in the pumping scheme, which also leads to lower efficiency. An applied external magnetic field destroys coherences between Zeeman hyperfine levels but the resulting changes in hyperfine coupling modify the transition probabilities, and may even allow forbidden transitions, which are the key to efficient optical pumping. In this paper, we evaluate the effect of changes in hyperfine coupling in cesium caused by an applied external magnetic field and the consequence in the efficiency of optical pumping.

I. Introduction

Preparation of a cesium atomic beam by laser optical pumping has been proposed to increase the performance of cesium beam frequency standards which currently operate with magnetic state selection.[1] Different pumping schemes have been introduced using one or two lasers.[2-4] We have chosen to study schemes involving two lasers because they show great promise for cesium preparation since they can act as a nearly perfect state selector.

We have shown in a previous theoretical study[5-7] that it is impossible to achieve perfect optical pumping in cesium by adjusting the linewidth of the lasers. We are faced either with coherences between atomic levels or with excitation of neighboring transitions which in both cases results in less efficient optical pumping. In that study, the magnetic field was assumed to be zero.

It was shown by Théobald *et al.*[8] for one-laser schemes using a σ-polarized laser, that a small magnetic perturbation is important in destroying Zeeman hyperfine coherences. It would be expected that a low applied magnetic field, less than 1 G, would not change the optical pumping efficiency very much since the perturbation in the energy levels is low. However, the magnetic field changes the hyperfine coupling. Thus, the (F, m_F) manifold is no longer a set of pure states, and this results in allowing transitions normally forbidden in zero magnetic field such as $\Delta F = 2$ or $\Delta F = 0$ $\Delta m_F = 0$. The latter is the key to efficient optical pumping as long as it stays forbidden.

In this paper, we evaluate the effect of changes in hyperfine coupling in cesium caused by an external applied magnetic field and the consequence in the efficiency of two-laser optical pumping.

II. Theory

The theoretical model is based on the semiclassical approach using the density operator formalism and fluctuating laser field.[7] This model represents more closely the optical pumping experiment on

the cesium beam since it takes into account the effect of atomic coherences, the effect of neighboring transitions, and the fact that the semiconductor lasers used for pumping have lorentzian profiles with linewidths of the order of the Cs natural linewidth.

In this model, we are interested in the mean value of the population of each Zeeman sublevel. In order to find the expected value of the population difference between the $F = 4$ $m_F = 0$ and the $F = 3$ $m_F = 0$ sublevels of the ground state, we must also evaluate the mean value of the Zeeman coherences, the correlations between optical coherences and field fluctuations of the corresponding laser, and the correlations between ground state hyperfine coherences and field fluctuations of both lasers.

In the presence of an applied external magnetic field **B**, we observe the Zeeman splitting of the hyperfine levels. The energy levels can be determined from the Hamiltonian[9]:

$$\mathcal{H} = A\mathbf{I}\cdot\mathbf{J} + \frac{B[3(\mathbf{I}\cdot\mathbf{J})^2 + \frac{3}{2}\mathbf{I}\cdot\mathbf{J} - \mathbf{I}^2\mathbf{J}^2]}{2I(2I-1)J(2J-1)} + \frac{\mu_B}{h}g_J\mathbf{J}\cdot\mathbf{B} + \frac{\mu_B}{h}g_I\mathbf{I}\cdot\mathbf{B} \tag{1}$$

where, for the cesium ground state $^2S_{1/2}$, $A = 2298.1579425$ MHz,[10] $B = 0.0$ MHz, and for the $^2P^o_{3/2}$ state, $A = 50.275(3)$ MHz, $B = -0.53(2)$ MHz[11]; $g_I = -3.9885395 \times 10^{-4}$.[10]

The evaluation of the eigenvalues and the eigenvectors of the Hamiltonian gives respectively the energy levels and the new state vectors in a particular magnetic field. The eigenvectors are expressed as

$$|\psi(F_e m_e)\rangle = \sum_{F'_e} c_{F_e m_e F'_e} |F'_e m_e\rangle \tag{2}$$

for an excited state, and

$$|\psi(F_g m_g)\rangle = \sum_{F'_g} c_{F_g m_g F'_g} \left|F'_g m_g\right\rangle \tag{3}$$

for the ground state. The sums are done only on eigenvectors with the same m_F in zero magnetic field since the perturbation introduced by the static magnetic field couples only states with the same m_F.

The electric dipole matrix elements become

$$\langle\psi(F_e m_e)| D_q |\psi(F_g m_g)\rangle = a[\psi(F_e m_e), \psi(F_g m_g), q] \left(\frac{3\pi\epsilon_o\hbar c^3}{\omega_{eg}^3}\Gamma\right)^{1/2}, \tag{4}$$

and

$$a[\psi(F_e m_e), \psi(F_g m_g), q] = \sum_{F'_e F'_g} c_{F_e m_e F'_e}\, a(F'_e m_e; F'_g m_g; q)\, c_{F_g m_g F'_g} \tag{5}$$

where

$$a(F_e m_e; F_g m_g; q) = (-1)^{1+I+J_e+F_e+F_g-m_e}\sqrt{2F_g+1}\sqrt{2F_e+1}\sqrt{2J_e+1} \times \begin{pmatrix} F_e & 1 & F_g \\ -m_e & q & m_g \end{pmatrix} \begin{Bmatrix} F_e & 1 & F_g \\ J_g & I & J_e \end{Bmatrix} \tag{6}$$

The last two equations illustrate the fact that although $a(F_e m_e; F_g m_g; q)$ may be zero for some transitions in zero magnetic field, $a[\psi(F_e m_e), \psi(F_g m_g), q]$ will in general be non-zero in a magnetic field, thus allowing usually forbidden transitions.

III. Results

We have calculated the optical pumping efficiency, defined as the difference between the populations of the levels $F = 4$ $m_F = 0$ and $F = 3$ $m_F = 0$ of the cesium ground state $^2S_{1/2}$, and the fluorescence, defined as the number of transitions per second for one atom of cesium, for various magnetic fields **B** from 0 to 0.5 G. One laser, of 30 MHz linewidth, is assumed to be tuned to the transition $^2S_{1/2}$ $F = 3$ $m_F = 0$ - $^2P^o_{3/2}$ $F = 4$ $m_F = 0$, denoted (3–4+$\epsilon\pi$) or (3–4+$\epsilon\sigma$), depending on its polarization, π or σ; $+\epsilon$ stands for the small offset due to the change in the transition frequency caused by the displacement

of the levels in a magnetic field. The second laser, of 2 MHz linewidth, is tuned around the transition $^2S_{1/2}\, F = 4 - {}^2P^o_{3/2}\, F = 4$ denoted (4–4+$\Delta\pi$); it is π-polarized and swept by Δ Hz around the zero field center frequency of the same transition. The radiation level from both lasers is taken as 10 mW cm^{-2}. Samples of such calculations are shown in Fig. 1 and 2 for magnetic fields of 0 G and 0.5 G.

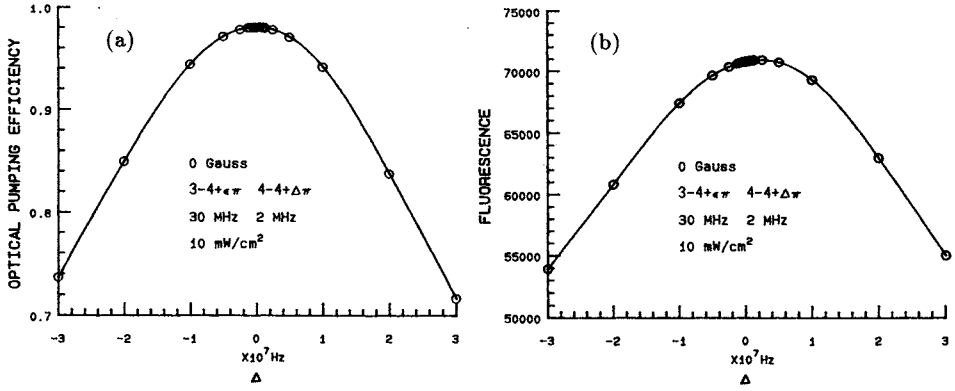


Fig. 1. Optical pumping efficiency (a) and fluorescence (b) at 0 G as a function of the frequency offset Δ of the second laser between -3 and 30 MHz.

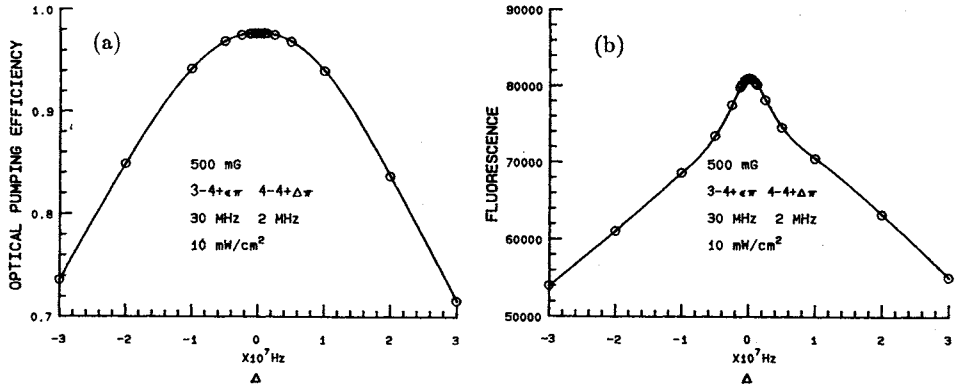


Fig. 2. Optical pumping efficiency (a) and fluorescence (b) at 0.5 G as a function of the frequency offset Δ of the second laser between -3 and 30 MHz.

We observe that, when the first laser is π-polarized, i.e. (3–4+$\epsilon\pi$), the maximum of fluorescence does not correspond, at low magnetic field, to the maximum of optical pumping efficiency, which is centered on the transition (4–4) (Fig. 3). This is not the case when the first laser is σ-polarized, (3–4+$\epsilon\sigma$). The shift of fluorescence peak is large, 2 MHz, at zero magnetic field and smaller above 0.3 G. We still have

to investigate the causes of this effect which was not easily predictable. The optical pumping efficiency peaks are sufficiently broad at the top so that the influence of the fluorescence shift is small if we tune the second laser on the fluorescence maximum.

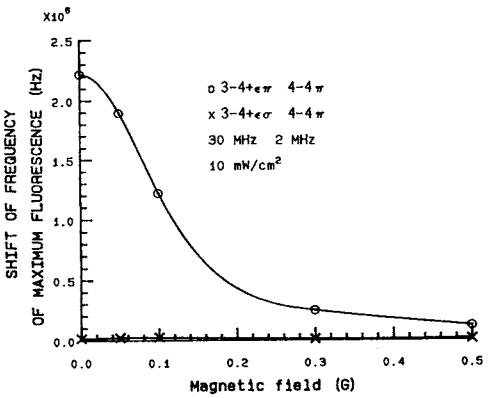


Fig. 3. Shift of frequency at the maximum of fluorescence in relation to the center frequency of the transition (4–4) in zero magnetic field.

Figures 4 and 5 show respectively the optical pumping efficiency at maximum fluorescence and the maximum fluorescence when B varies from 0 to 0.5 G. The results show no major influence of the magnetic field for the two-laser pumping schemes studied here. They can be compared with those from Théobald *et al.*[8] where a large magnetic field produces an increase both in optical pumping efficiency and in fluorescence for one-laser pumping schemes. Figure 5 tells us also that, though the optical pumping efficiency is slightly better for the (3–4+$\epsilon\pi$) (4–4π) scheme than for the (3–4+$\epsilon\sigma$) (4–4π) scheme, the corresponding fluorescence is more than three time smaller. If the lasers frequencies are to be locked from the fluorescence signal, a larger signal would provide better laser stability.

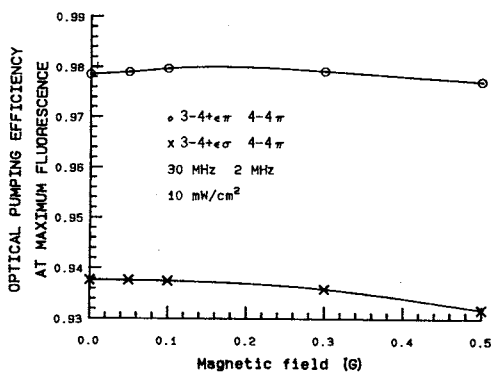


Fig. 4. Optical pumping efficiency as a function of the magnetic field.

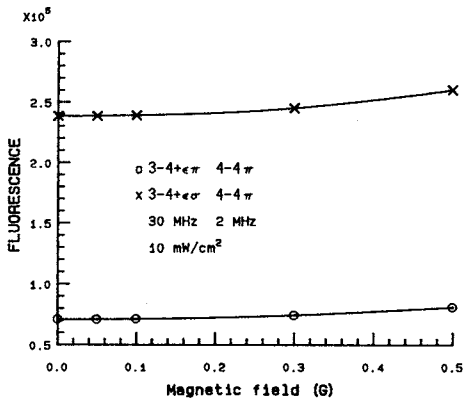


Fig. 5. Fluorescence as a function of the magnetic field.

IV. Conclusion

We have studied the effect of small applied external magnetic field on the efficiency of two-laser optical pumping in cesium by evaluating the change in coupling of atomic states and the resulting changes in atomic level energies and transition probabilities.

The magnitude of the magnetic field is low, less than 1 G. There is no major influence of the magnetic field at these values on the optical pumping efficiency. We observe also that, at magnetic field less than 0.3 G, the frequency of maximum fluorescence is shifted by up to 2 MHz from the center of the transition if the two lasers are π-polarized. Finally, the calculations show that, although the optical pumping efficiency is larger when the two lasers are π-polarized, the fluorescence is three time smaller than when the laser tuned to the transition (3–4+ϵ) is σ-polarized. The latter optical pumping scheme, (3–4 + $\epsilon\sigma$)(4–4π), is preferable if the lasers are to be stabilized from the fluorescence signal.

References

[1]M. Arditi and J.-L. Picqué, J. Phys. — Lett. (Paris) **41**, L379 (1980).

[2]G. Avila, É. de Clercq, M. de Labachelerie, and P. Cérez, IEEE Trans. Instrum. Meas. **IM-34**, 139 (1985).

[3]G. Avila *et al.*, Phys. Rev. A **36**, 3719 (1987).

[4]S.-I. Ohshima, Y. Nakadan, and Y. Koga, IEEE Trans. Instrum. Meas. **IM-37**, 409 (1988).

[5]C. Jacques and P. Tremblay, in *Proceedings of the 42nd Annual Symposium on Frequency Control*, Philadelphia, 1988 (IEEE Press, Piscataway, 1988), pp. 505–509.

[6]P. Tremblay and C. Jacques, in *Fourth Symposium on Frequency Standards and Metrology*, edited by A. de Marchi (Springer-Verlag, Berlin, 1989), pp. 382–385.

[7]P. Tremblay and C. Jacques, Phys. Rev. A (Accepted for publication).

[8]G. Théobald, N. Dimarcq, V. Giordano, and P. Cérez, Opt. Commun. **71**, 256 (1989).

[9]N. F. Ramsey, *Molecular Beams* (Oxford University Press, London, 1956).

[10]E. Arimondo, M. Inguscio, and P. Violino, Rev. Mod. Phys. **49**, 31 (1977)

[11]C. E. Tanner and C. Wieman, Phys. Rev. A **38**, 1616 (1988)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-01 Number 4

VELOCITY DISTRIBUTION MEASUREMENT IN AN OPTICALLY PUMPED CESIUM BEAM RESONATOR

A. HAMEL, V. GIORDANO, P. PETIT, C. AUDOIN
Laboratoire de l'Horloge Atomique
Unité Propre de Recherche du CNRS
associée à l'Université de Paris-Sud
Bât.221 - Université Paris-Sud
91405 ORSAY Cedex - France

ABSTRACT:

The knowledge of the velocity distribution in the cesium beam is necessary to evaluate the performances of a cesium beam frequency standard. The velocity distribution in the optically pumped cesium beam resonator operating at Laboratoire de l'Horloge Atomique has been measured by two methods.

The first method uses the Fourier transform of the clock signal at resonance versus the microwave power. It is a general method which was applied for the first time to a magnetic state selection frequency standard by J.S Boulanger. We demonstrate that, when a one laser configuration is considered, the velocity distribution remains Maxwellian. On the other hand, when the optical detection is performed with a cycling transition (two lasers scheme), the experimental distribution is well represented by a 1/v weighted Maxwellian distribution.

The second method is specific to an optically pumped resonator. Atomic populations are modulated in the pumping region. We analyse the amplitude and phase of modulated fluorescence light emitted in the detection region. The transfer function obtained is the Fourier transform of the time of flight distribution between these two regions. We present here preliminary results relative to this second method.

I. INTRODUCTION:

Optical pumping technics should lead to significant improvements of frequency performances of cesium beam frequency standard [1]. A complete evaluation of the accuracy of this standard needs a precise knowledge of the effective velocity distribution of the atoms in the beam[2].

When atomic preparation and clock signal detection are performed by optical pumping, some processes can lead to a weighting of the maxwellian initial velocity distribution. Noncomplete atomic preparation or detection by cycling transition favour low velocity atoms. Thus it is important to know how the optical pumping affects the velocity distribution of the detected atoms.

Two distribution velocity measurement technics have been applied at the Laboratoire de l'Horloge Atomique to the cesium beam resonator CSIII described in a previous publication [3].

The first method developed by J.S Boulanger in NRC [4] uses the Fourier transform of the clock signal at resonance versus the microwave power. Principles and results of this technics are presented in section II.

In the second method the transfer of a population modulation between the two optical interaction regions of the resonator is analysed. The transfer function obtained is the Fourier transform of the time of flight distribution. Preliminary results are presented in section III.

II.BOULANGER METHOD:

At resonance the transition probability $P(b,\omega_0)$ between the two ground state sublevels interesting the clock transition is directly related to the cosine Fourier transform of the interaction time distribution in the Ramsey cavity $f(\tau)$ [2]:

$$F(b) = 1 - 2\,P(b,\omega=\omega_0) = \int_0^{\infty} f(\tau)\,\cos 2b\tau \;\; d\tau \qquad (I.1)$$

where ω_0 is the clock transition frequency, τ the interaction time and b the amplitude of the microwave perturbation expressed in angular frequency units.

The reverse transform gives:

$$f(\tau) = \frac{4}{\pi} \int_0^{\infty} F(b)\,\cos 2b\tau\, db \qquad (I.2)$$

The microwave frequency is set to the resonance frequency ω_0 and the evolution of F(b) is experimentaly recorded in a large range of variation of the microwave power. Figure 1 shows typical experimental record obtained with one laser configuration.

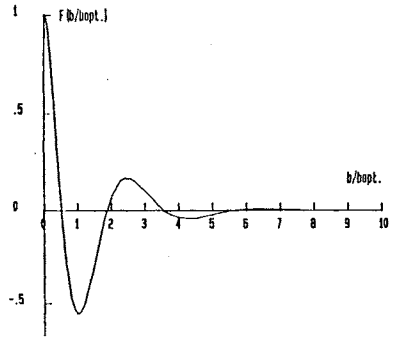


Figure 1. F(b) variation versus Rabi frequency. b_{opt} is the value of b that optimize the amplitude of the clock signal

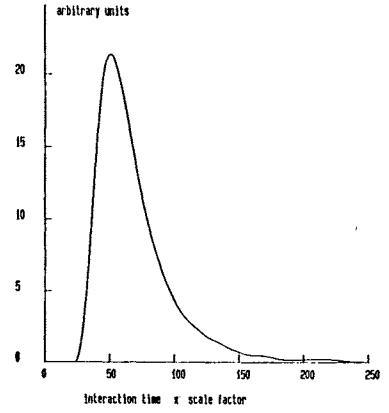


Figure 2. Fast Fourier Transform of F(b)

Then the Fourier transform of F(b) is obtained by Fast Fourier Transform. It is given in figure 2. The precise value of the microwave power seen by the atoms is not known. We do not know the precise value of the cavity Q. This introduce a scale factor in the interaction time scale that we have now to determine.

From the approximated interaction time distribution we calculate the Ramsey pattern of the resonance line that we compare with an experimental record. After some iterations the calculated Ramsey fringe fits with the experimental one, see Figure 3. The interaction time scale factor is thus determined. Then this interaction time distribution is compared with a theoretical distribution in order to determine the weighting caused by optical pumping.

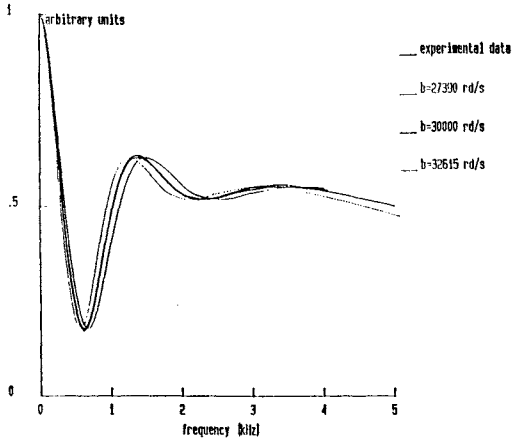


Figure 3. Comparison between the calculated Ramsey pattern and the experimental one

Figure 4 and 5 shows the interaction time distribution computed as described above in the cases of one laser and two laser configurations respectively.

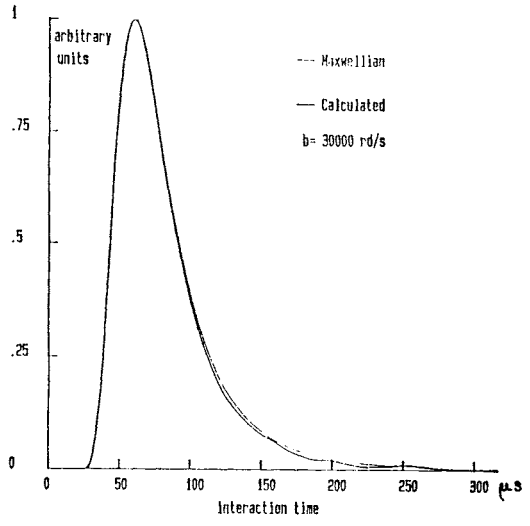


Figure 4. One Laser configuration
3 - 4' π (D1 line)
Comparison between computed and Maxwellian distributions

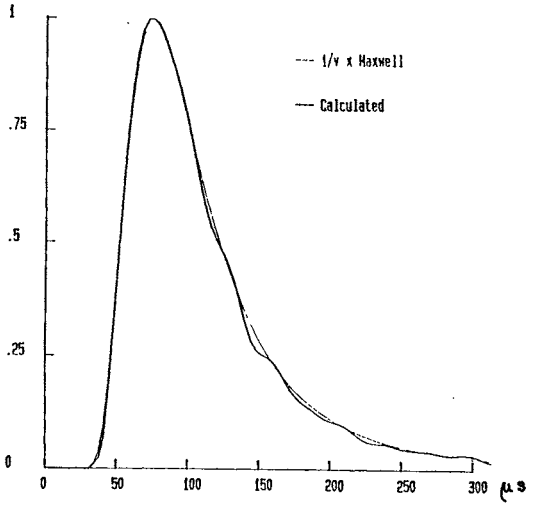


Figure 5. Two laser configuration:
atomic preparation 4 - 4' σ (D2 line)
optical detection 4 - 5' σ (D2 line)
Comparison between computed and 1/v weighted Maxwellian distributions.

One can seen that in the case of one laser configuration the velocity distribution of the detected atoms remains Maxwellian. If the detection is performed by a cycling transition the experimental distribution is well represented by a 1/v weighted Maxwellian distribution. Indeed in such configuration the number of photons emitted in the detection region is proportional to the time of flight of the detected atoms in the laser beam.

The Boulanger method allows us to determine the precise value of the Rabi frequency b, which is also of great interest since a great number of frequency shifts depend on the microwave level.

It should be noted that this technique in only valid if the Ramsey probability can be expressed by the equation I.1. It means that the the static magnetic field in the microwave interaction region has to be set to a large value in order to make negligible the contribution of the tails of the transitions that surround the clock transition. Indeed, at high microwave level, the width of the microwave resonance pattern is notably increased.

III TRANSFER FUNCTION METHOD:

This technique is specific to an optically pumped cesium beam resonator. It is described in Figure 6.

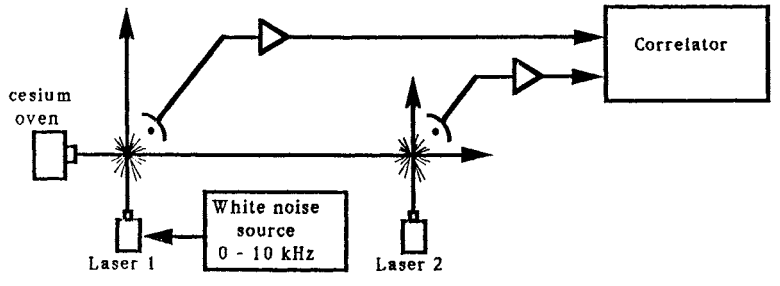


Figure 6. The tranfer funtion method

The mean frequency of the pumping laser (Laser 1) is set on the side of the 3-3'σ optical transition. The injection current of this laser diode is modulated by a white noise. leading to a modulation of the pumping light frequency.This gives a modulated F=3 ground state sublevel population N(t), which is detected by observing the emitted fluorescence light. The laser power is set to a low value (typically 1mW) in order to avoid saturation. The atoms freely evolve between the two optical interaction regions. In the detection region a time dependent atomics flux N'(t) is detected. N'(t) is related to N(t) and to the time of flight distribution f(T) by the convolution relation:

$$N'(t) = \int_0^\infty f(T)\ N(t-T)\ dT$$

where T= L/v is the time of flight between the two optical interaction regions.

We can expresse this relation in the frequency domain as:

$$N'(\omega) = f(\omega) \times N(\omega)$$

$f(\omega)$ appears as the transfer function of the modulation along the atomic beam trajectory. The amplitude and the phase of the detected signal are analysed with an multichannel spectrum analyser Scientific Atlanta SD380 used as a correlator. This apparatus directly delivers the tranfer function $f(\omega)$ and the impulse response f(T) which is the time of flight distribution. Figure 7 and 8 shows typical data obtained with this technique.

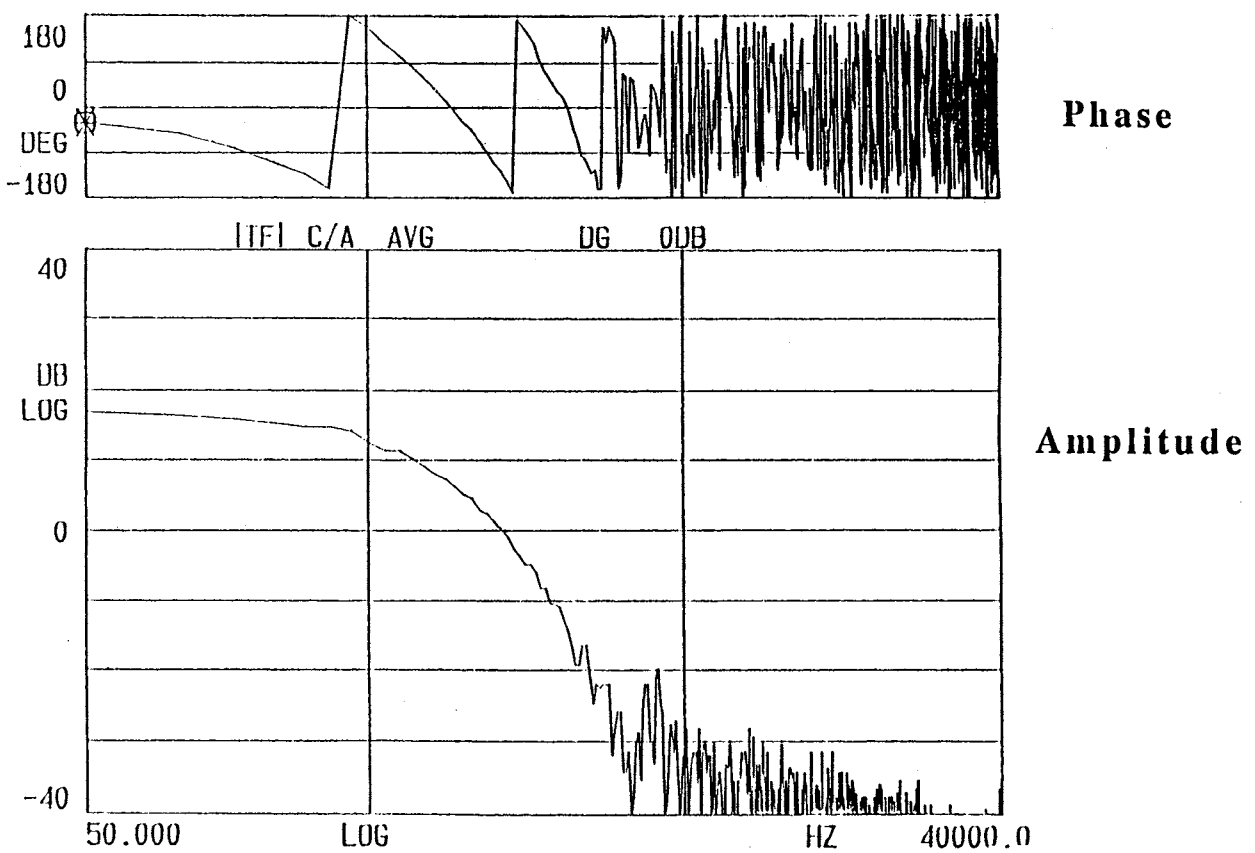


Figure 7. Tranfert function $f(\omega)$. One laser configuration 3-3'σ

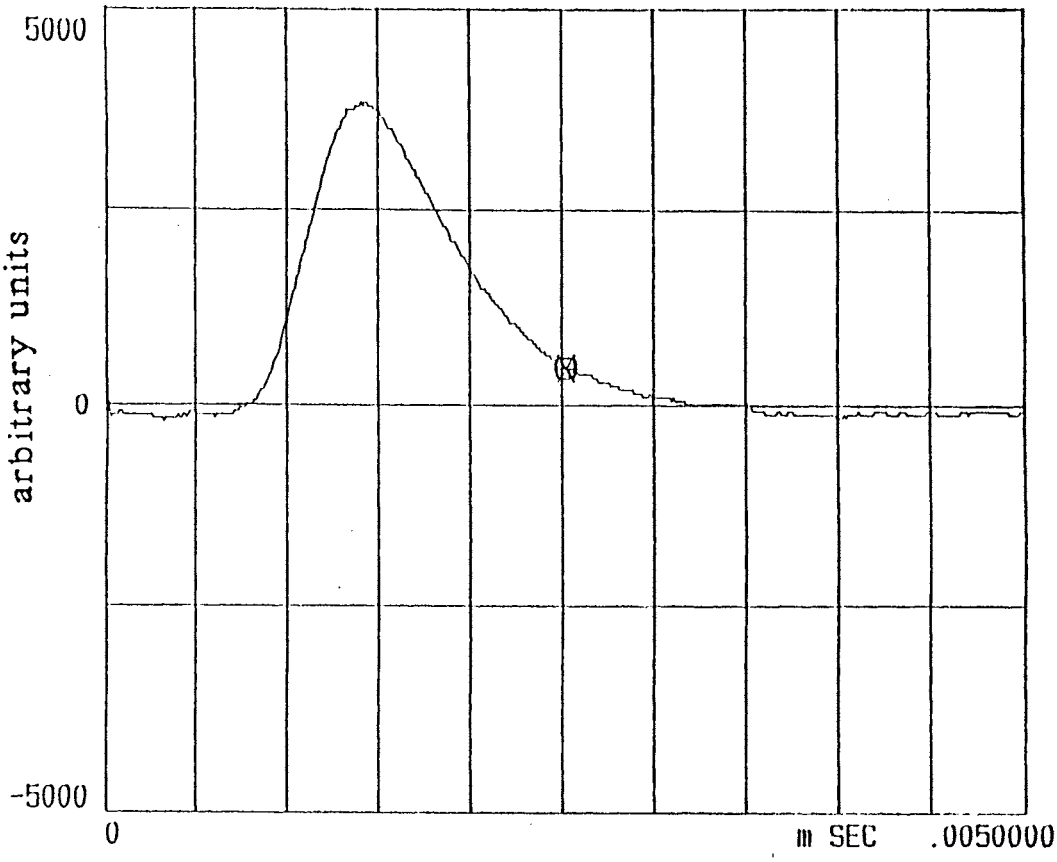


Figure 8 Fourier transform of the function shown in fig 7.
Time of flight distribution

We observe a good agreement between the two methods. However, the second method is less accurate than the first one. Furthermore we detect the total population of the ground state sublevel: the Zeeman degeneracy is not resolved. Nevertheless this technique is very simple to implement and does not need the modification of the static magnetic field value. We think that this technique could be use to analyze the influence of spurious effects on the velocity distribution as Hanle effect.

CONCLUSION:

Two velocity distribution measurement methods have been presented. They lead to the same conclusions. In one laser configuration the laser power is sufficient to achieve a complete atomic preparation and thus the velocity distribution remains Maxwellian. When a cycling transition is used to perform the clock signal detection, the Maxwellian initial velocity distribution of the atoms is weighted by a 1/v factor.

REFERENCES:

[1] : *"Frequency Stability of an Optically Pumped Cesium Beam Frequency Standard: Préliminary Results."*
V. Candelier, V. Giordano, A. Hamel, G. Théobald, P. Cérez et C. Audoin.
Actes du 2eme Forum Européen Temps-Fréquence, Neûchatel, Mars 1988,pp.483-498

[2] : *"The Quantum Physics of Atomic Frequency Standards."*
J. Vanier and C. Audoin.
Adam Hilger, Bristol and Philadelphia. Vol.2, Chapter 5 , part 2.

[3] : *"New Design for an Efficient Optically Pumped Cesium Beam Tube."*
V. Giordano, A. Hamel, G. Théobald, N. Dimarcq, P. Cérez and C. Audoin.
Proceedings of the 43nd Annual Symposium on Frequency Control,
Denver (USA). June 1989, pp 130-134

[4] : *"On the accuracy of Cs Beam Primary Frequency Standards. "*
J.S Boulanger, R.J Douglas, J. Vanier, A.G Mungall, Y.S Li, C. Jacques.
Proceedings of the 16th Annual Precise Time and Time interval Measurements Meeting
Washington(USA). 1984.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO 01 Number 5

GPS SYNCHRONIZED DISCIPLINED RUBIDIUM FREQUENCY STANDARD

J. S. McNabb

TRAK Systems
Division of TRAK Microwave Corp.

ABSTRACT:

A disciplined rubidium frequency standard steered by the corrected 1 PPS output from a GPS timing receiver or other stable 1 PPS source can provide a low phase noise, modest cost, frequency standard a with long term frequency stability of a cesium standard and a short term stability of better than $1x10^{-11}$. Phase correlation to the 1 PPS input is maintained within 50 nanoseconds during disciplining, and within 100 nanoseconds during a five hour coasting period with no disciplining input.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-01 Number 6

A COMPUTERIZED TIME AND FREQUENCY STANDARDS MONITORING SYSTEM

Kalevi Kalliomäki (1,2), Tapio Mansten (1), Jari Mannermaa (2)
(1) Technical Research Centre of Finland, Metrology Research Institute,
Otakaari 7 B, SF-02150 Espoo
(2) University of Oulu, Department of Electrical Engineering, SF-90570 Oulu

1.Introduction

It seem to be quite normal that time and frequency standards are monitored only a few times a day or more seldom. However several things like temperature, pressure, propagation conditions, phase jumps in radio links and so on, which affect accuracy, may vary quite rapidly. Therefore one has to collect relevant data effectively. In our solution the basic data collection cycle is 10 minutes. Every two hour the twelve measuring results from each channel are statistically treated to form the cleaned result which is stored to a floppy disk.

2. Instrumentation

In Metrology Research Institute of Finland traceability of time and frequency is based on several sources like GPS, Loran-C (Sylt) and Sovjet loran RNS. One GPS receiver (Kinemetrics truetime) and three Loran-C phase locked clocks (own construction) serve as a link to UTC. Three rubidium oscillators work as a laboratory standards and three cesium clocks (Oscilloquarz) owned by finnish post and telegraf administration will work as an independent source of time in the near future. All those clocks and receivers and environmental parameters are monitored by a data acquisition and prosessing system developed by the institute. The time interval counter has one nanosecond resolution and IEC interface to computer. The start pulse comes from one rubidium clock and the stop pulse from a 16 channel digital multiplexer.
The laboratory controls the frequency of the synchronizing pulses of the Finnish TV network, because accredited calibration laboratories trace their frequency via TV transmissions. The time signal is also supplied to the Finnish broadcasting company. As an Euromet project together with STA, Russian loran RNS(E-A) is also monitored.

3. Analog and digital channels

Sixteen analog channels from 32 possible and twelve digital channels from 16 are in use to day. Five analog channels are reserved to different temperature readings (inside and outside), 8 channels monitor Loran-C receivers (signal levels, phase tracking errors and crystal control voltages), 2 channels watch for backup batteries and one channel is for atmospheric pressure. Four digital channels are reserved for Loran-C receivers, one for the GPS receiver, two for remaining rubidium clocks and one for TV sync. pulse. Four channels are reserved for varying purposes.

4. Data acqusition

The basic data collection period can be varied but in our application 10 minutes is suitable. Every two hour the twelve measuring results from each channel are statistically treated (Chauvenets criterium) to form the cleaned result which is stored to a floppy disk. Original results at 10 minute intervals are normally stored to hard disk for 24 hours only. Those results and regression residuals can be viewed graphically, which helps to observe abnormal phase fluctuations and ambient changes.
A second computer is used to process the data aqcuired by the first one. Several computer programs are developed to handle (picking, smoothing, correcting, removing peaks, combining and so on) the data, which is finally treated using Lotus 123 program.
One program is connected to an Euromet project, which studies uncertainty level in time comparisons between nearby countries using Russian RNS navigation system. The details of this method are discussed in another poster.

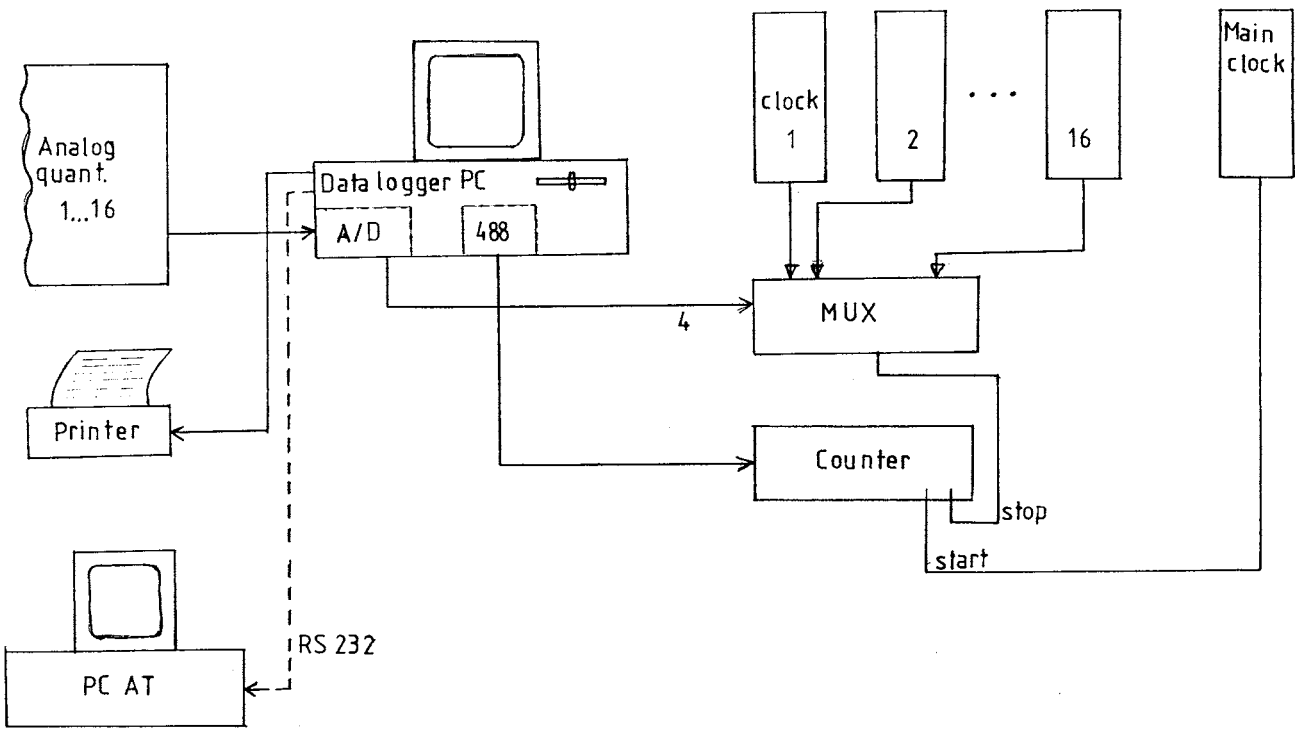
5. TV monitoring program

A monitoring and reporting program for the surveillance of the rubidium based, stabilized TV line frequency is relatively complicated. The problems in phase comparison are due to phase jumps of TV field sync. pulse. Associated with digital picture processing and program changes in TV studio 100 ns reversible phase changes occur 4...8 times a day. The phase seem to have "low" and "high" value. In addition phase jumps in microsecond or even millisecond level occur approximately every week. Those big jumps seem to associate with changes in propagation paths (TV links) or with power failure of the synchronization generator in TV centrum. If we do'nt take into account those reversible phase jumps, the possible error in average line frequency determination will be nearly 2*10exp-12 for one day comparison.

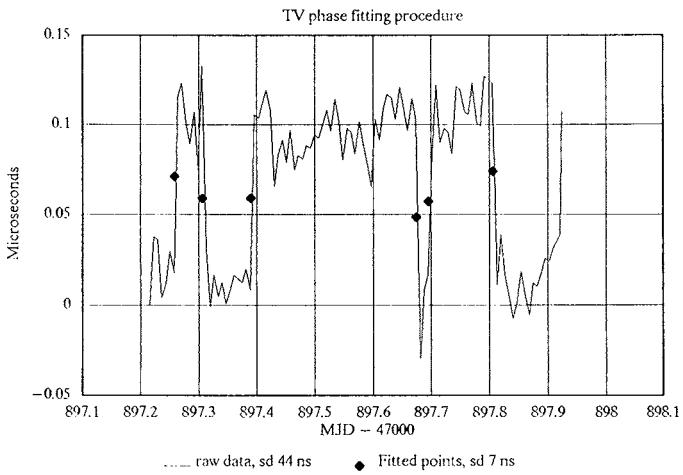
To see every phase change we have to sample the TV phase quite often, a regular 10 minutes sampling intervall is fast enough. It means approximately 4500 data points during one month for TV only instead of the one point per day that is really needed.
Our solution for data reduction was to save the average phase around a 100 ns phase change only. In other words the program seeks for phase changes and saves the average phase, a value exactly halfaway between the "low" and "high" phase. Thus it is not necessary to know whether the direction of a phase jump nor the real phase (high/low). A tiny drawback is that the those reversible phase jumps are needed to catch the phase data. As a result of data compression, one year TV data can be easily handled using standard software.

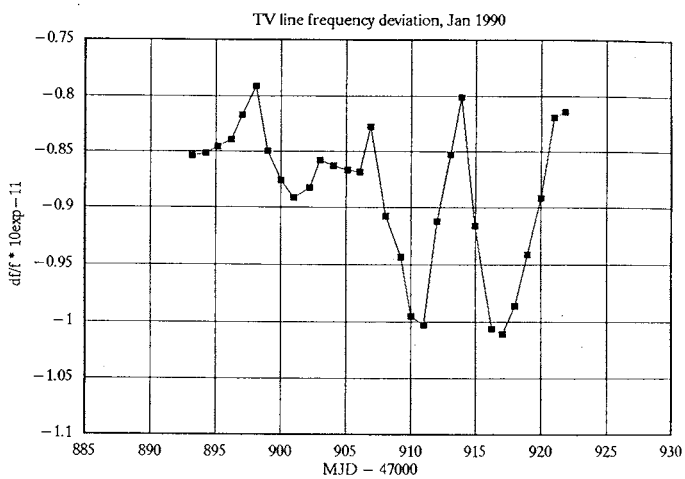
6. Conclusion

A computer controlled multichannel data acquisiton and prosessing system developed for time laboratories and suitable tools to handle the phase and frequency data automatically is developed. Environmental parameters like temperature and pressure, important voltages in receivers and relative phases between various clocks are monitored at 10 minute intervalls. Then the data is compressed using statistical and other suitable criteria. A TV line frequency monitoring program is specially designed to utilise reversible 100 ns phase jumps associated with digital TV picture processing.



A COMPUTERIZED TIME AND FREQUENCY STANDARDS MONITORING SYSTEM


TV phase fitting procedure
0.15
0.1
0.05
0
−0.05
Microseconds
897.1
897.2
897.3
897.4
897.5
897.6
897.7
897.8
897.9
898
898.1
MJD − 47000
raw data, sd 44 ns
Fitted points, sd 7 ns


TV line frequency deviation, Jan 1990
−0.75
−0.8
−0.85
−0.9
−0.95
−1
−1.05
−1.1
df/f * 10exp−11
885
890
895
900
905
910
915
920
925
930
MJD − 47000

Session PO-01 Number 7

REMOTE CONTROL OF A VLBI TIMING STATION

Roberto Ambrosini

Istituto di Radioastronomia, CNR
via Irnerio, 46 Bologna (I)

To operate the second Italian VLBI station in Noto, Sicily (see Fig. 1), it has been developed a new system to allow the remote control in real time of all the locally generated timing parameters.

A standard terminal connected via modems and a normal telephone line to the Personal Computer located at the site (see Fig. 2) can check the correct operation of the two local atomic frequency standards and of the Loran-C receiver.
In particular it is possible to restart from a remote location a complete syncronization sequence of this receiver, which gives the local UTC reference (see Fig. 3). Also the 32 operating parameters of the H-Maser, utilized as the primary frequency standard for the station, can be monitored remotely in real time (see Fig. 4).
The Maser Interface (MI) developed for this application allows also to record if, at any time, one of those parameters has exceeded some preset high and low thresold values (see Fig. 5).

Daily measurements of all these quantities toghether with delays (respect the Loran-C signal) of the time scales associated with the H-Maser and the Rb standards (this last used as an emergency back-up for the first) are performed in unattended mode; the results are hard copied on paper while only the most significant ones are stored on disk for further elaborations.

A similar system has been implemented also at the VLBI station of Medicina, Bologna, but it is not open to the remote control.

In the poster presented at the Forum we have shown data for the H-Maser located in Noto. These last have been possible to be updated up to the last week **(05/03/1990)**, utilizing the remote control and the programs GET.FILE and R.FILE written for this purpose. Just for space convenience, we present here instead data relatives to the H-Maser of the first Italian VLBI station in Medicina. The graphs plot for every day of year 1989 the recorded atomic signal amplitude (Fig. 6) and the delay of the local UT scale respect to the Loran-C chain (Fig. 7). Similar plots are available also for similar measurements against the IEN scale, as obtained via TV-sync comparisons. Finally a program called ELAB allows to display and compute multiple regressions on all these type of data.

REFERENCE: R. Ambrosini, M. Roma "Il sistema di controllo remoto della stazione T&F per uso VLBI", Rapporto Interno IRA 120/89.

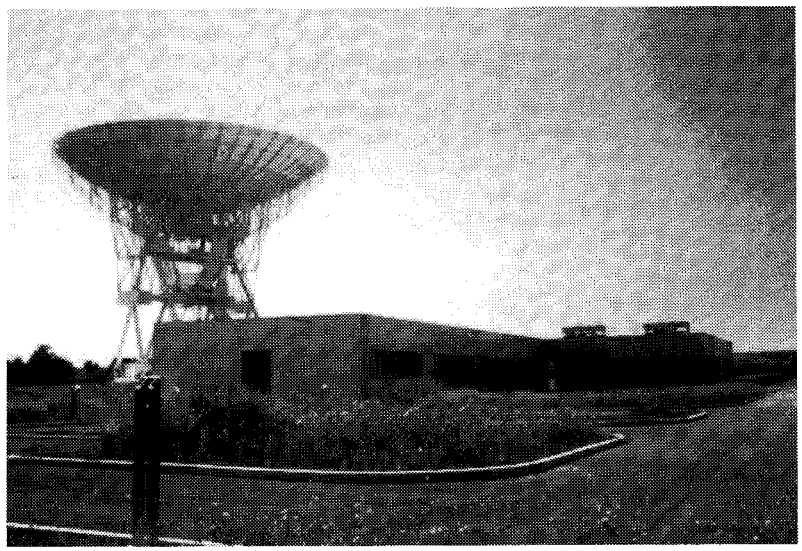


Fig 1. The VLBI station in Noto, Sicily (I).

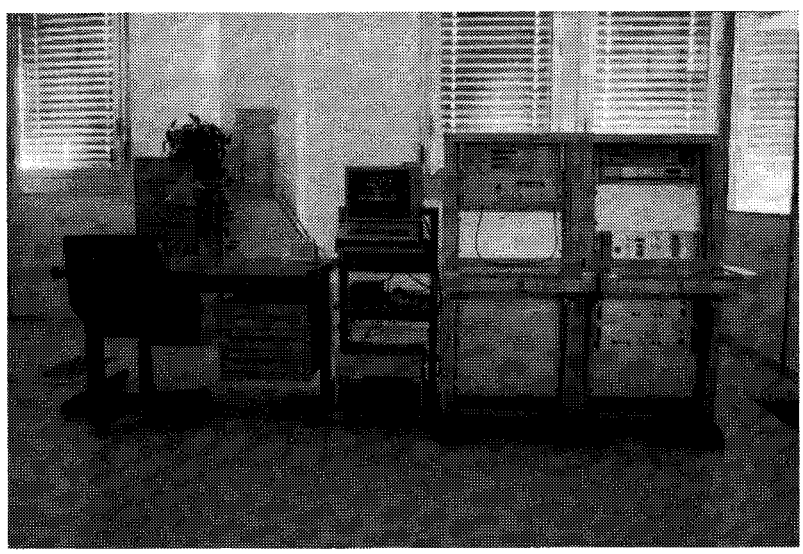


Fig. 2. The T&F console in Noto. On the left the Apple computer with Modem, disk and printer; a (spare) Rubidium clock and the decoding receiver of the IEN timing signals in the middle; finally on the right the programmable Loran-C receiver, with the H-Maer clock and monitoring unit.

```
]RUN R.SET.AUSTRON

R.SET.AUSTRON   program of 26/01/89
Actual values at Time   10:40:24

<1>  Status byte is          00000000
<2>  LORAN-C delay  (uS) = -000001.36  <-- First Frequency Standard
<3>  Signal/Noise   (dB) = > +9
<4>  Receiver gain  (dB) = 66
<5>  Frequency Offset    = -2.3E12
<6>  Cycle Number (2.xx) = 3
<7>  TOC ADJUST constant = -000955.43
<8>  SLEW difference     = -000954.80
<9>  GRI            (us) = 079900.00
<10> SEC.TD.        (us) = 000000.00
<11> Tracked STATION     = MASTER
<12> Time constant (0-4) = 0
<13> Last valid TOC      = 10:27:41
<14> 1 PPS-EXT.-UTC (uS) = -4.79  <-- Second Frequency Standard
<15> Start Acquisition
<16> Read or Set UT time
<17> 1 BEGIN TOC

What number (another key to Exit) ?
```

Program **R.AUSTRON** allows to start a complete syncronization sequence of the Loran-C receiver or just to read the 1PPS delays of the first and second frequency standards.

Fig. 3

```
]RUN R.MASER

R.MASER  program of 10/03/89

@
UUUUUUUUUUU
#99(
#99:02 103A 00 00 60 DD
#99:02 103E 00 FF FF DD
#99!
#99=27/
$
#99"
#99)
#00
#99(
#99>
#99/
#99!
#99=0/
$


#99!
#99=37/
$
#99>
#99/
#99!
#99=25/
$
#99"
#99)
#00
```

```
R.MASER  program of 10/03/89

Alarm = LOW

Ch/Value Ch/Value  Ch/Value  Ch/Value

0 106    10 64     20 41     30 87
1 16     11 62     21 41     31 62
2 106    12 49     22 81     32 100
3 17     13 59     23 25     33 100
4 37     14 47     24 -109   34 -102
5 95     15 67     25 -21    35 100
6 80     16 -5     26 -108   36 100
7 80     17 9      27 -21    37 -101
```

Program **R.MASER** reads the 32 operating parameters of the H-Maser in Noto.

Fig. 4

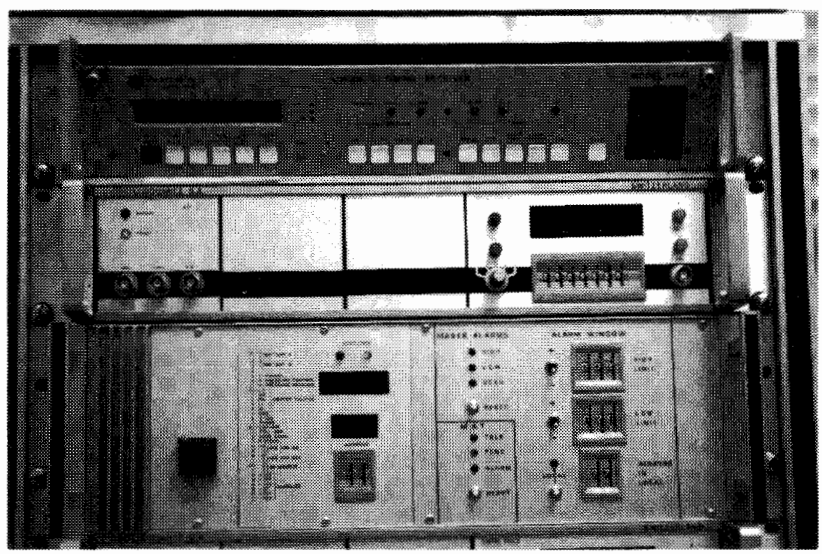


Fig. 5 A close-up view of the Loran-c receiver, the H-Maser clock and its upgraded monitoring unit. This last includes a MAT card and the new Maser Interface which samples the operating parameters and records if the high or low limits have been exceeded. Also the unlocked condition sets an alarm flag.

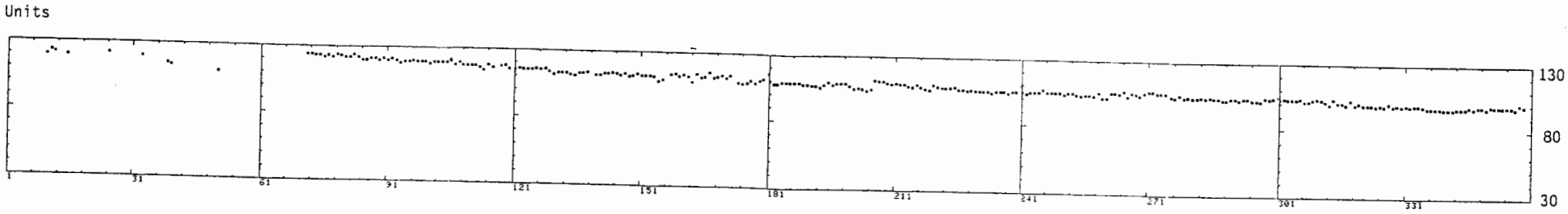


Fig. 6 - FILE : AMP1.89 - YEAR 1989 - 6 days/div.

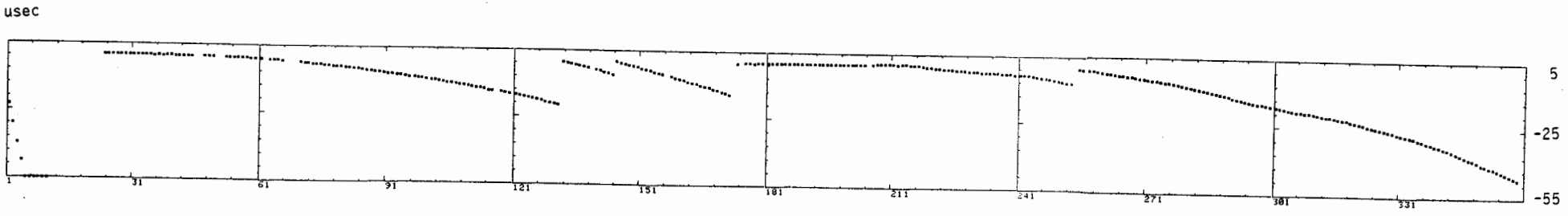


Fig. 7 - FILE : LOR.1.89 - YEAR 1989 - 6 days/div.

Session PO-01 Number 9

APPLICATION OF TIME AND FREQUENCY IN GEODESY

WOLFGANG SCHLÜTER
INSTITUT FÜR ANGEWANDTE GEODÄSIE
FUNDAMENTALSTATION WETTZELL
D-8493 KÖTZTING
F.R. GERMANY

Abstract. Geodetic space techniques employ atomic frequency standards and make use of the global timescale UTC. This paper points out the requirements for the geodetic techniques to T/F and showes the T/F system as realized at the Fundamentalstation Wettzell. Results of Cs-clock-comparisons and of the H-Masers EFOS 1 and EFOS 2 are given.

1. Observation Techniques and Requirements

The high accuracy of the geodetic observation techniques as Laser Ranging (SLR, LLR) and Very Long Baseline Interferometry (VLBI) enables to investigate plate tectonics and the dynamical behavior of the earth. The accuracy achieved up to now also has been dependant on the progress made in building clocks and frequency standards. The geodetic space observation techniques employ time delay- or phase measurements and make use of the availability of a global time scale such as UTC.

1.1 Laser Ranging Technique

The Laser Ranging Technique (Satellite or Lunar Laser Ranging/SLR or LLR) is based on the measurement of the turn-around-time of a light pulse transmitted to the target (artificial satellite or the moon) and returned by retroreflectors on the target. The distance results from multiplying the measured turn-around-time with the speed of light and applying corrections due to atmospheric refraction and instrumental offsets. The accuracy achieved today is better than 1cm. The distances (s) to satellites varies from 800 km to 40.000 km; the distance to the moon is approximately 400.000 km. The T/F-system has to provide the reference frequency and the epoch for the measurement. The influence of the T/F-System should be less than 5 mm. The requirements to the frequency stability are estimated from

$$\frac{df}{f} = \frac{ds}{s}$$

to be

for satellites ($\tau = 6 \ldots 260$ms) $6 \cdot 10^{-9} \ldots 1.3 \cdot 10^{-10}$,

for the moon ($\tau = 3$s) $1.3 \cdot 10^{-11}$.

The epoch timing with respect to a global timescale must be better than 500ns regarding the high velocity (approx. 7km/s) of the satellites. These requirement are met by atomic clocks (Rb-, Cs-clocks).

1.2 Very Long Baseline Interferometry (VLBI)

The VLBI-technique employes common observations with at least two radiotelescopes of radiosignals transmitted from extragalactic radiosources (quasars). The phasedifference of the received microwave signal at the various telescopes can be determined by correlation of the signals if a stable

local timescale is recorded together with the signal. The observation of a set of quasars covering the common horizon permits the determination of the baselines, of the earth rotation parameters and of the clock-parameters. The requirements to the local oscillator can be derived from the basic formula

$$c\tau = xX + yY + zZ + cdT + c(T - T_0)\frac{df}{f}$$

with

c	speed of light,
τ	propagation delay,
x,y,z	baseline vector between the telescopes,
X,Y,Z	position of the quasar (unit vector),
$T - T_0$	observation period,
f	frequency

as

$$\frac{df}{f} < \frac{0.5}{c(T - T_0)} = \frac{1.6 \cdot 10^{-11}}{T - T_0}.$$

Dependant on the signal strength the duration of the observation of a quasar last from several seconds up to several minutes. The duration of an observation session including sufficient radiosources last from 1 hour to several hours. As the clock parameters e.g. time- and frequency-offsets and drifts are included in the parameters to be determined, a coherent frequency must be provided from the local oscillator at least for a period of 3 hours. This demand results in the requirement to the frequency of

$$\frac{df}{f} < \ldots 1.5 \cdot 10^{-15}.$$

H-Masers have to be employed to meet these requirements.

The correlation process of the signals could be accelerated if the epoch offsets of the local clocks involved in the observations are known as precise as possible. The current requirement is 100ns with respect to the global time scale UTC(USNO).

1.3 Geodetic Applications of GPS

Precise geodetic applications of GPS make use of simultaneous phasemeasurements of at least two GPS-receivers to the carrierfrequencies transmitted from in minimum four satellites against ground frequencies generated by the receivers. The estimation of the frequency stability of the receiver oscillators could be performed in the same mannor as it has been done for the VLBI-technique, neglecting that the GPS-satellites are in 20.000km height orbits. The interval to take phasemeasurements are between 3s and some minutes. Employing double-difference-techniques the clock behavior of the satellite and receiver clock could mainly be eliminated. Nevertheless for orbit determination or for using only phasemeasurements precise ground frequencies are still required. For the modelling of the clock behavior coherent intervals up to some minutes still are requested. For permanent GPS-tracking stations, which provide data for orbit determination, the use of H-Masers is recommended.

The requirements to frequency from the geodetic space techniques is summarized in figure 1. The stability plots of commercial frequency standards is also shown. It is obvious that the requirements leads to the use of atomic frequency standards.

2. The T/F-System at Wettzell

The T/F-system (figure 2) at the Fundamentalstation Wettzell consists of various frequency standards to meet the requirement for the provision of

- reference frequencies,
- local timescale, referenced to UTC.

The timescale is generated by 5 Cesium atomic clocks, the connection to UTC is performed by LORAN C and GPS-measurements. The SLR-, LLR- and VLBI- and also the GPS- and Doppler-System are equipped with their own atomic clock to take into account the long- and shortterm requirements. The Laser Ranging Systems are equipped with Rubidiumstandards while the VLBI-facility employes H-Masers. The permanent GPS-tracking-receiver is connected to the H-Maser of the VLBI-system. All clocks and frequencystandards are integrated in the T/F-System. Comparisons of the various clocks are carried out by using a timeinterval counter.

3. Results of the Clock Comparisons.

Based on dayly clock comparisons via GPS common-view technique, the timeoffset of the timescale UTC(Wettzell) against UTC is derived. Figure 3 shows the offsets of the UTC(Wettzell) compared to UTC from 1987 to 1989. Time and frequency corrections are introduced, to keep UTC(Wettzell) within 2 microseconds to UTC(USNO).

The offsets of two cesiumclocks against UTC which are running free are given as an example in figure 4. A time- and frequencystep has been introduced for Cs(131) after a short failure of the clock.

The offsets of the H-Masers against UTC is given in figure 5 to 7. EFOS 1 shows a smaller frequency drift than EFOS 3. Frequencysteps and timesteps are only introduced if the offsets are out of the accepted limits. As EFOS 1 was free and undisturbed running for more than 1 year the drift behavior of the Maser can be seen obviously in figure 5.

4. Contribution of VLBI and SLR to T/F

Geodetic space techniques can contribute to time and frequency distribution. The VLBI-technique is not only a precise tool for baseline determination and for the estimation of the earth rotation parameter. The technique provides very precise time- and frequency offsets of the H-Masers involved. Figure 9 shows impressivly the timedelay residuals after the correlation. Placing more emphasis to the calibration of the telescopes and the related electronics, the data could be definetively used as comparisons of the corresponding clocks.

The SLR-technique could be employed for precise timesynchronisation as it is proposed for the LASSO-experiment. As LASSO is onboard of the geostationary satellite Meteosat not all SLR sites have the capability to range. First ranging results have been detected with the new Wettzell Laser Rangig System (WLRS) (figure 8). The results are encouraging to participate in the two way synchronisation experiment LASSO beside the French LLR-station in Grasse and the Austrian station in Graz.

References

Carter, W. E., Robertson, D.S., MacKay, J.R., *Geodetic Radio Interferometric Surveying: Applications an Results, Journal of Geophysical Research, Vol. 90, No. 86, 1985.*

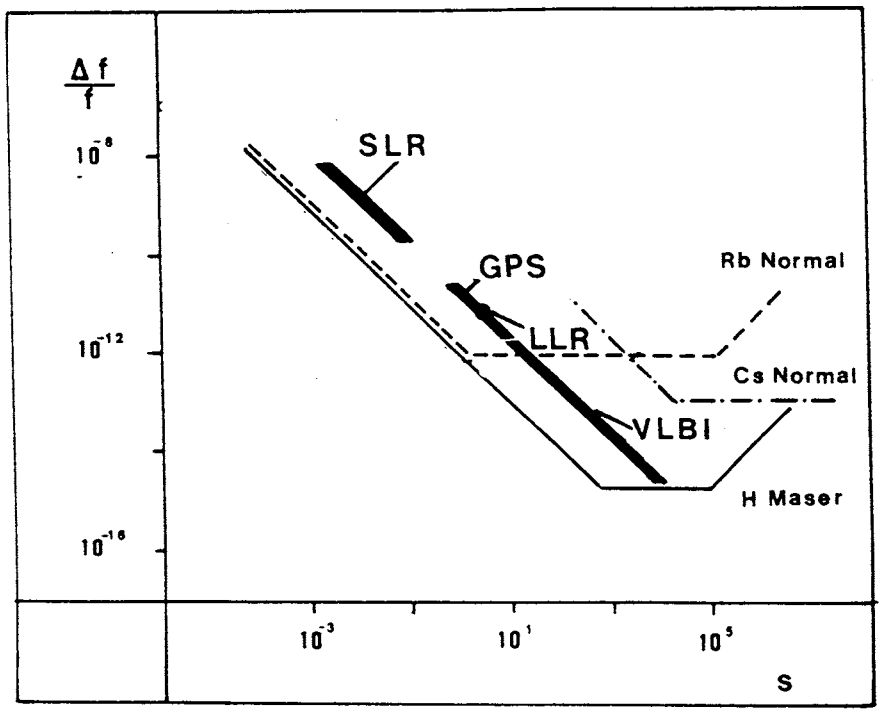


Figure 1: Frequency requirements in geodetic space techniques

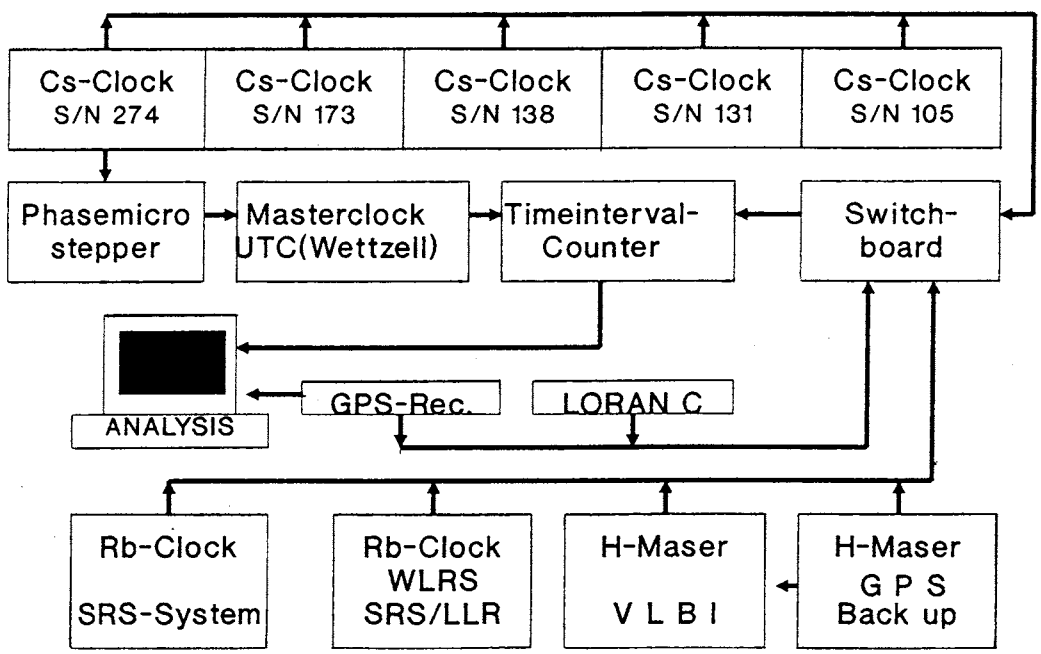


Figure 2: Diagram of T/F-System of the Fundamentalstation Wettzell

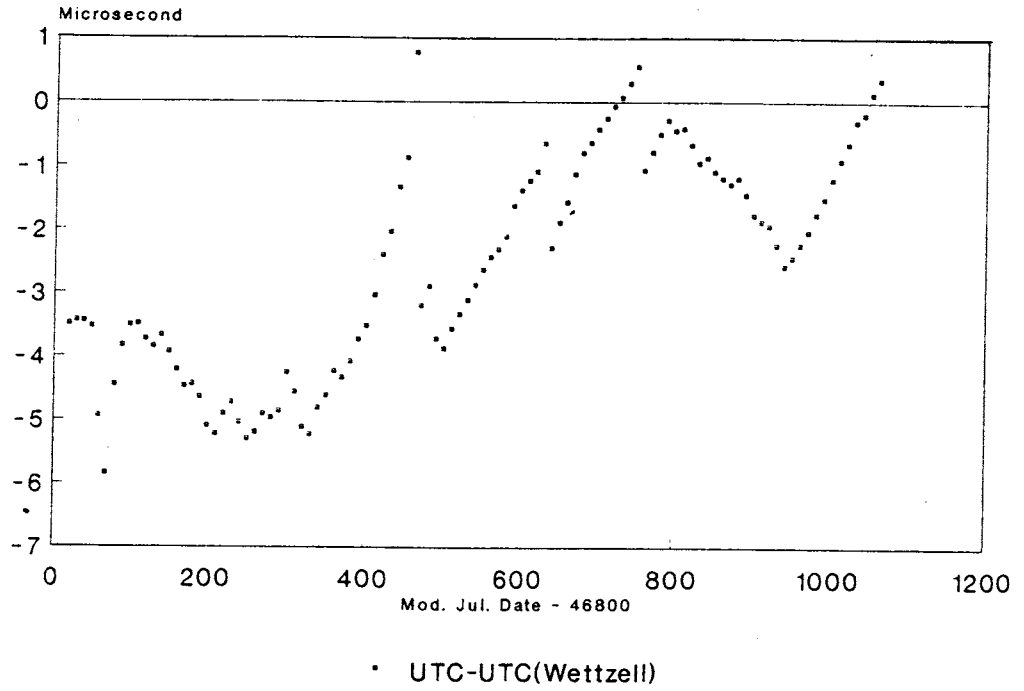


Figure 3: Clock offset UTC(Wettzell) against UTC

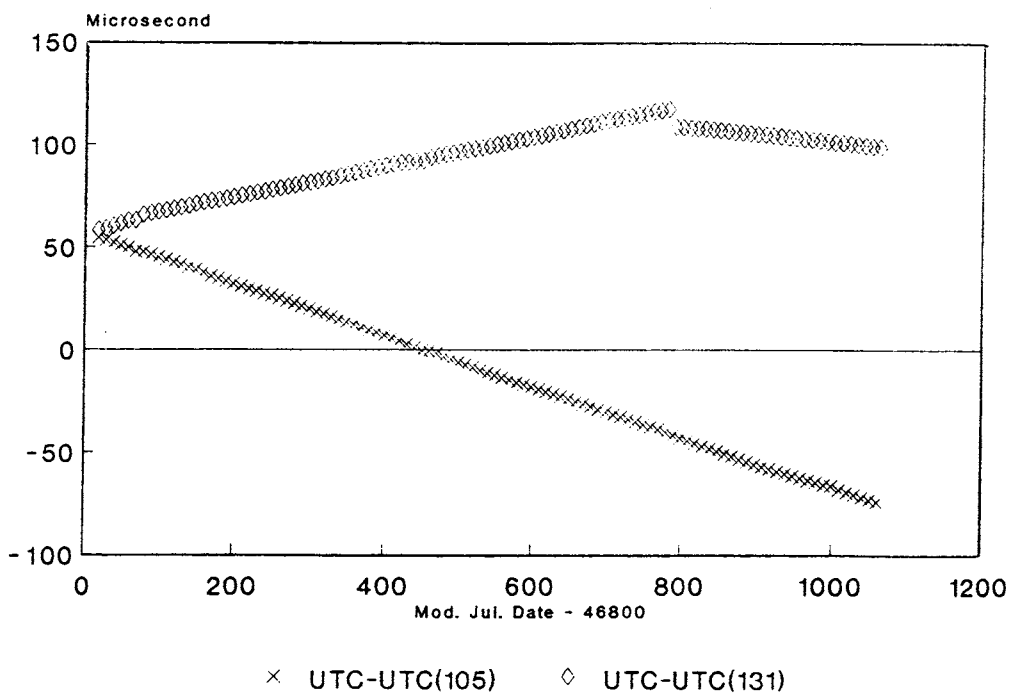


Figure 4: Clock offset of 2 Cesiumclocks against UTC

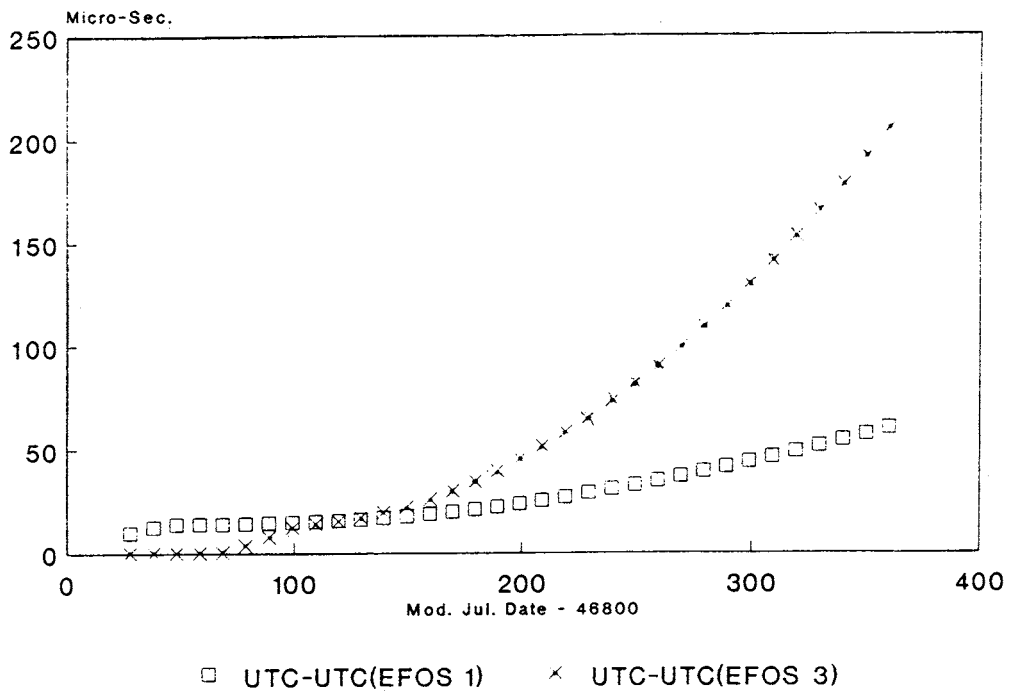


Figure 5: Clock offset of EFOS 1 and EFOS 3 against UTC in 1987

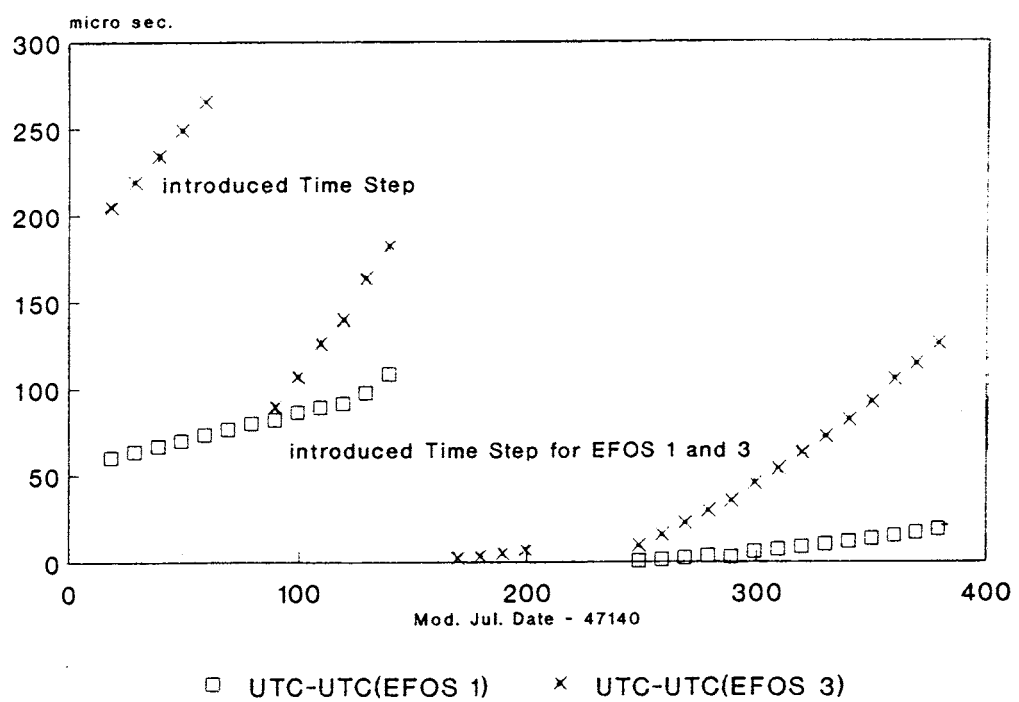


Figure 6: Clock offset of EFOS 1 and EFOS 3 against UTC in 1988

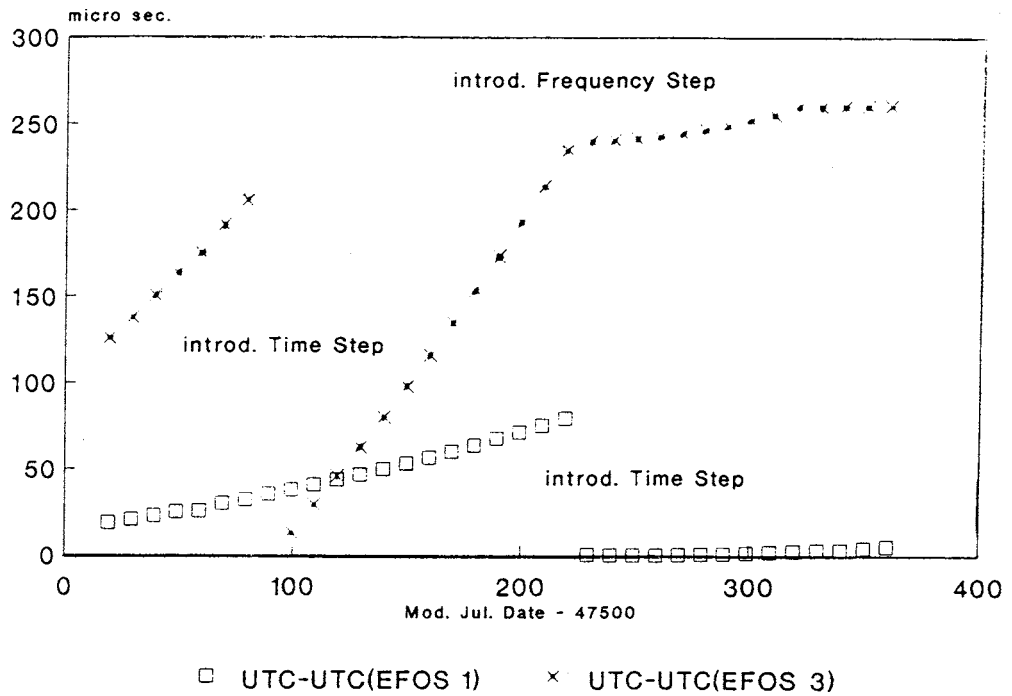


Figure 7: Clock offset of EFOS 1 and EFOS 3 against UTC in 1989

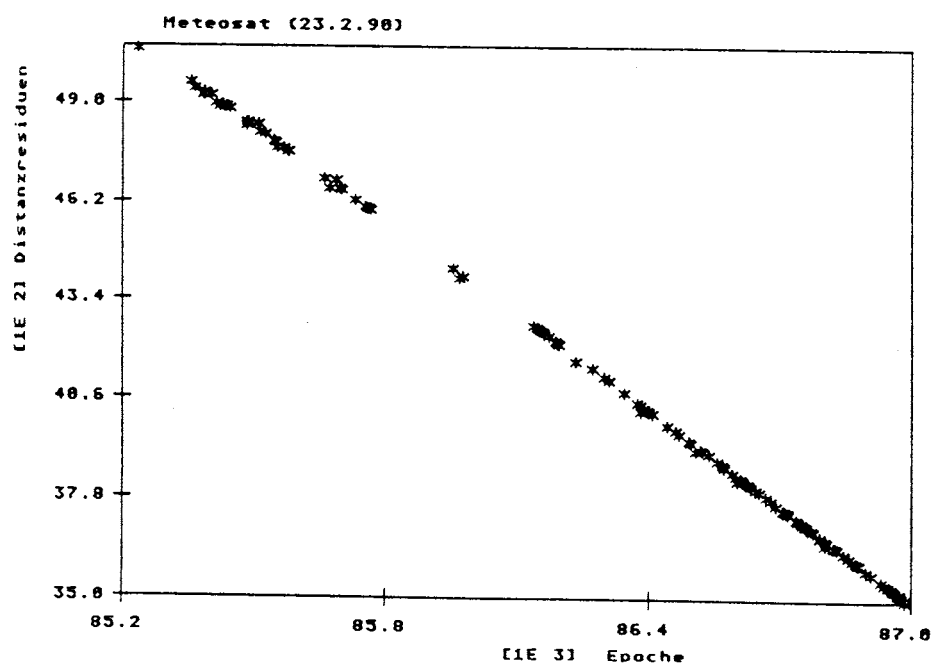


Figure 8: SLR-Ranging results to METEOSAT with WLRS

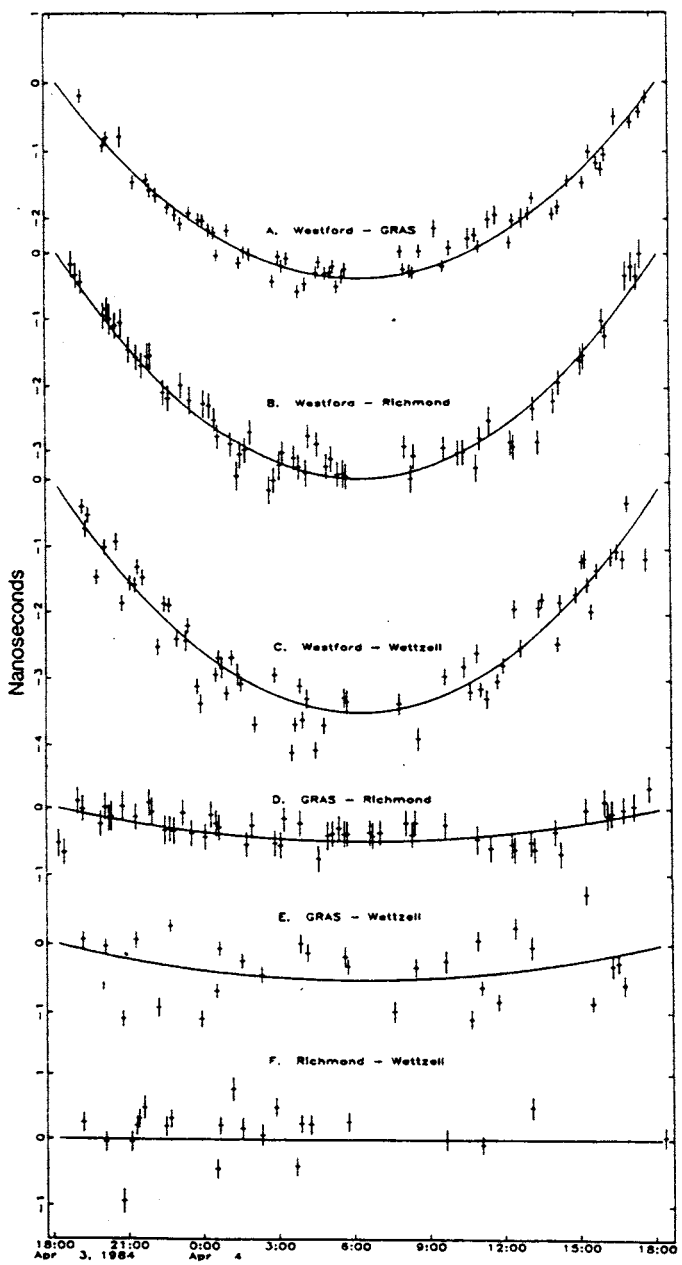


Figure 9: Timedelay residuals after correlation of a VLBI-Experiment

Session PO-01 Number 10

Specification of the BVA-8600 OCXO for Space Use in ESA's Data Relay Satellite

Laurent-Guy Bernier
Giovanni Busca
Observatoire Cantonal, Neuchâtel, Switzerland

1.0 Introduction

The Data Relay System (DRS) is based on 2 geostationary satellites that shall provide data-relay services between the client Low Earth Orbiting (LEO) satellites and the ground stations [1]. The Data Relay System Satellites (DRSS) will be located at 59° E and 44° W. In addition to the data-relay services, the DRS will provide a LEO localization service using one DRSS as a ranging relay. The OCXO (USO for Ultra Stable Oscillator) embedded into the Frequency Generation Unit (FGU) of each DRSS must therefore generate a 5 MHz reference signal with both an excellent frequency stability, for ranging purposes, and an excellent spectral purity, for the production of the various carriers at microwave frequencies generated on board [2].

The goal of the phase B1 of the DRS project, which now comes to an end, was to produce equipment specifications on the basis of the system specification. In that context the OCXO BVA-8600b, manufactured by Oscilloquartz in Switzerland, was examined as a potential candidate for the DRSS-USO. Besides, the prospect of space qualifying the BVA-8600 OCXO has already generated a lot of interest [3, 4, 5] because of its excellent phase noise and frequency stability characteristics. This paper presents selected environmental considerations about the specification of the new BVA-8600b, now in preproduction (BVA-8603b: option 3, synthesis grade), for space use in the DRSS.

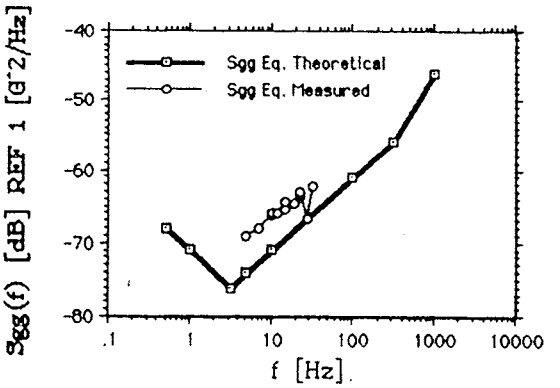


Figure 1
BVA-8600b Equivalent Vivration Spectrum

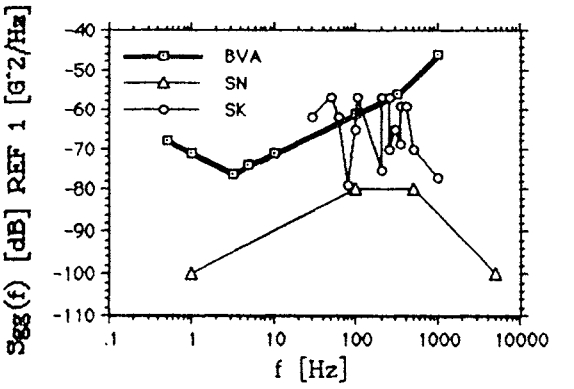


Figure 2
Olympus Vibration Environment vs BVA-8600b Equivalent Vibration Spectrum

2.0 USO specifications

Among other specifications, the satellite specification document calls for a frequency stability of 5×10^{-13} for $\tau = 1$ s and the absolute frequency offset must be kept smaller than 9×10^{-9} at any time over a temperature range of -10 C° to 50 C° with a remote calibration rate limited to once per month. The high spectral purity of the OCXO must be conserved despite the environmental fluctuations present in the satellite.

3.0 Acceleration and Vibration

The specified acceleration coefficient of the BVA-8603b is 1x10-9 [1/g]. Using this coefficient, it is possible to compute an equivalent vibration spectrum which we define as the vibration spectral density in $[g^2/Hz]$ which would produce an excess phase noise equal to the phase noise specification of the BVA-8603b. The dynamic vibration sensitivity of the original BVA-8600 was measured experimentally several years ago at Thompson, in France. Figure 1 shows the experimental equivalent vibration spectrum, as

computed from the Thompson data, versus the theoretical equivalent vibration spectrum, computed using the static acceleration coefficient of the BVA-8603b.

Figure 2 shows the vibration spectrum on the Olympus platform both under normal conditions (curve SN) and during station keeping maneuvers (curve SK). It is assumed that the vibration environment on the DRSS would be similar. The equivalent vibration spectrum of the BVA-8603b (curve BVA) is shown on the same figure. It appears that the BVA-8603b phase noise performance would not be degraded in normal operation. However the phase noise performance would be affected during station keeping maneuvers.

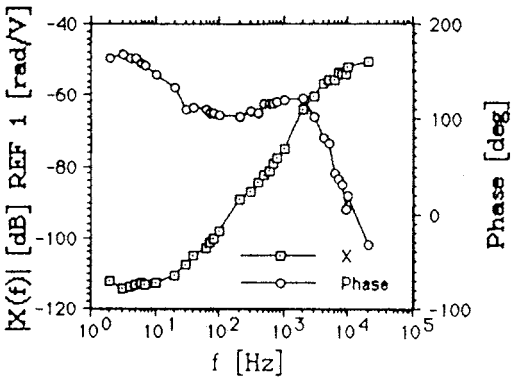


Figure 3
BVA-8600b DC Supply Voltage to Phase Transfer Function

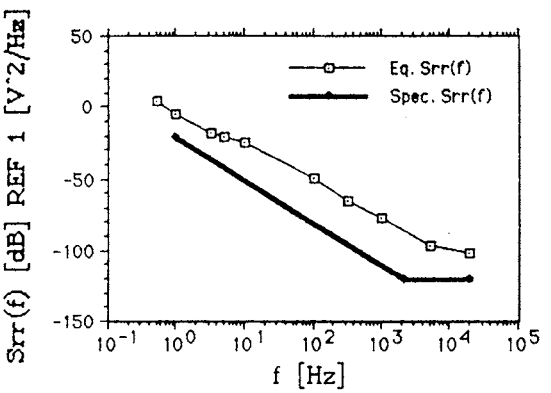


Figure 4
BVA-8600b Equivalent DC Supply Noise Spectrum vs DC Supply Noise Specification

4.0 Power Supply Ripple

The ripple to phase transfer function of a pre-production BVA-8600b was measured in order to define a power supply ripple specification. The measured transfer function is shown on figure 3. Using this transfer function, it is possible to compute an equivalent DC supply noise spectrum which we define as the voltage spectral density in $[V^2/Hz]$ which added to the DC supply voltage would produce an excess phase noise equal to the phase noise specification of the BVA-8603b. The equivalent power supply ripple is shown on figure 4. The high-pass effect of the series voltage regulator combined with the BVA-8603b phase noise specification yields an equivalent spectrum with a frequency dependence close to f^{-1}. This can be modelled by a flicker noise process. The power supply ripple specification can be defined by taking a sufficient safety margin below the equivalent ripple. The specification shown on figure 4 assumes that the power supply noise can be described by the combination of a flicker noise process plus a white noise process. This noise level is easy to achieve in a power supply and guaranties the full phase noise performance of the BVA-8600b.

5.0 Conclusion

There are many other environmental topics that are relevant to the space qualification of the BVA-8600b [4, 5]. The vibration and power supply ripple specifications were selected for this paper because they have not been described elsewhere. From the present study we conclude that the power supply requirements are very easy to achieve and that the acceleration sensitivity of the BVA resonator may induce a phase noise degradation during station keeping maneuvers.

6.0 References

[1] De Agostini A., "Data-Relay Satellite (DRS) Ground-Segment Design," ESA Journal, vol. 12, 1988, pp. 249-266.

[2] Bernier L.G., Busca G., "Frequency Stability and Spectral Purity Requirements for ESA's data Relay Satellite," *Proc. 3rd European Time and Frequency Forum*, Besançon, March 1989, pp. 215-219.

[3] Peier U.R., Jenni G., "Space Qualified 100 MHz BVA-VHF-Resonator, " *Proc. 2nd European Frequency and Time Forum*, Neuchâtel, March 1988, pp. 701-713.

[4] Schäfer W., "Results of the Qualification Programme of a BVA-Type Crystal Oscillator for Satellite Use, " *Proc. 2nd European Frequency and Time Forum*, Neuchâtel, March 1988, pp. 715-718.

[5] Suter J.J., Bates A.G., Cloeren J.M., "The Radiation Susceptibility of BVA-AT Resonators to Simulated Medium Earth Orbit Radiation, " *Proc. 2nd European Frequency and Time Forum*, Neuchâtel, March 1988, pp. 679-686.

Acknowledgements

The work reported in this paper was performed for ESA(ESTEC) at Neuchâtel Observatory under a contract with Marconi Space Systems, U.K.

Session PO-01 Number 11

CALCULATED FROM THE FREQUENCY DOMAIN: AN UPDATE

F. L. Walls, John Gary, Abbie O'Gallagher, Linda Sweet, and Roland Sweet

National Institute of Standards and Technology
Boulder, Colorado 80303 USA

Abstract

In this paper we investigate the dependence of the fractional frequency stability measures $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ on the parameter $2\pi f_h \tau$ for common power-law noise types. Previous investigators have calculated the value of $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ in the limit that $2\pi f_h \tau$ is either small or large compared to 1. We have implemented the calculations using numerical techniques which make it possible to perform the calculations for virtually any value of $2\pi f_h \tau$ and a wide variety of high-frequency filters. We find significant differences for $\sigma_y(\tau)$ depending on the filter, when $2\pi f_h \tau \ll 1$. The values for $\sigma_y(\tau)$ are nearly independent of filter slope for $2\pi f_h \tau \gg 1$. Of particular importance is R(n), the ratio of $\sigma_y(\tau)$ to $\mathrm{mod}\sigma_y(\tau)$, as a function of n, the number of measurements averaged together in the calculations of $\mathrm{mod}\sigma_y(\tau)$. We find differences from the values of R(n) versus n given in the literature for both flicker-phase and flicker-frequency noise types.

Introduction

The calculation of fractional frequency stability expressed in terms of the two sample or Allan Variance, $\sigma_y^2(\tau)$, from frequency domain measures such as the spectral density of frequency fluctuations, $S_y(f)$, is well known [1-3]:

$$\sigma_y^2 = 2 \int_0^{f_h} S_y(f) \frac{\sin^4(\pi f n \tau_0)}{(\pi f n \tau_0)^2} \, df.$$

$S_y(f)$, for most physical systems used in signal handling, is characterized by random noise whose power spectral density varies as a simple power law over wide ranges of Fourier frequencies, f. $\sigma_y(\tau)$ can be

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

calculated analytically under the condition that $2\pi f_h \tau$ is either very large or very small compared to 1. Here f_h is the equivalent noise bandwidth for the time domain characterization and τ is the measurement time [1-4]. When these conditions are not fulfilled there are corrections to $\sigma_y(\tau)$ that are not well known.

Numerical Techniques

The integral,

$$\sigma_y^2 = 2 \int_0^{f_h} S_y(f) \frac{\sin^4(\pi f n \tau_0)}{(\pi f n \tau_0)^2} \, df,$$

can be treated as an oscillatory integral except in the neighborhood of the origin. τ_0 is the minimum measurement time. Therefore, the integral is computed using one approximation in the neighborhood of the origin and a second method away from the origin, that is

$$\sigma_y^2 = I_1 + I_2,$$

where

$$I_1 = \int_0^{\alpha} 2F_1(f) df,$$

$$I_2 = \int_{\alpha}^{f_h} 2F_2(f) \sin^4(\pi f n \tau_0) df,$$

$$F_1(f) = S_y(f) \frac{\sin^4(\pi f n \tau_0)}{(\pi (f n \tau_0)^2},$$

and

$$F_2(f) = \frac{S_y(f)}{(\pi f n \tau_0)^2},$$

with α chosen so that the first integral is taken over $\eta/2$ cycles of the sin function, provided $\tau = n\tau_0$ is large enough. Otherwise it is taken over 2/3 of the interval f_h. That is, $\alpha = \text{minimum } (\frac{2f_h}{3}, \frac{\eta}{\pi})$. If the function F_1 grows more slowly than f^{-3} at f = 0, then the first integral exists. We assume that $S_y(f)$ grows no more rapidly than f^{-2}, and thus the integrand in this first integral has a removable singularity and can be computed using a general purpose quadrature routine.

The second integral is oscillatory. It can be transformed into canonical form by use of the trigonometric substitution,

$$\sin^4(\pi f\tau) = \frac{3}{8} - \frac{1}{2}\cos(2\pi f\tau) + \frac{1}{8}\cos(4\pi f\tau),$$

and can therefore be written as the sum of three integrals,

$$I_2 = I_{21} + I_{22} + I_{23}.$$

Since α is chosen large enough to avoid the singularity at the origin, the first of these three integrals can be computed with the standard quadrature routine. The remaining two integrals are oscillatory and are therefore computed using a routine, designed for integrals of the form $\int f(x)\sin(\omega x)dx$ or $\int f(x)\cos(\omega x)dx$ where ω can be very large. These integrals may have values of $n\tau_0$ which easily exceed 10^4 (here $n\tau_0$ corresponds to ω). Additional details can be found in [4].

The numerical code for $S_y(f)$ has the following form.

$$S_y(f) = \frac{C_1 f^{-2} + C_2 f^{-1} + C_3 + C_4 f + C_5 f^2}{K(f)M(f)} + C_6,$$

where K(f) is used to describe the offset on $S_y(f)$ due to locking to an external reference characterized by $S_y(f) = C_6$, and M(f) is used to provide high-frequency filtering. Alternately we can use an infinitely sharp cut-off, f_h to attenuate the high frequency noise. The values of the Cs are stored in an array. The filter function K(f) must be chosen from the following four functions.

$$K(f) = 1,$$

$$K(f) = \left(1 + \frac{1}{K_1 f}\right)^2,$$

$$K(f) = \left(\left(1 + \frac{1}{K_1 f}\right)\left(1 + \frac{1}{K_2 f}\right)\right)^2,$$

$$K(f) = \left(\left(1 + \frac{1}{K_1 f}\right)\left(1 + \frac{1}{K_2 f}\right)\left(1 + \frac{1}{K_3 f}\right)\right)^2.$$

$$M(f) = [(1 + M_1 f(1 + M_2 f)\ (1 + M_3 f)]^2,$$

where any or all of the M_i can be 0.

Results

Figure 1 shows the well known dependence of $\sigma_y(\tau)$ on τ for the common noise types when $2\pi f_h\tau \gg 1$ and an infinitely sharp filter at f_h is used to attenuate the high frequency noise.[1-3] The only difference between the results shown here, and that obtained for a single low pass filter is that the values for flicker phase are about 10% smaller. These calculations, performed using our numerical techniques, agree to within 1% with previous work. They serve to verify that the codes are operating correctly.

Figure 2 shows the results of our numerical calculation of $\sigma_y(\tau)$ for $\alpha = 2$ (white phase modulation noise) as a function of measurement time τ and three values of f_h. The calculations were made using both an infinitely sharp filter of width f_h and a single low pass filter of the form $M(f) = (1 + f/f_h)^2$. From the variations of $\sigma_y(\tau)$ from one f_h curve to the next we can uniquely determine the dependence of $\sigma_y(\tau)$ on f_h and on τ in the limits where $2\pi f_h\tau$ is either very small or large compared to 1. Once the dependence on f_h and τ has been determined, the numerical coefficient can be calculated. In the limit that $2\pi f_h\tau$ is either very small or large compared to 1 they generally agree to within 1% with those obtained from the analytical calculations [1-3]. The dependence of $\sigma_y(\tau)$ on τ and $2\pi f_h\tau$ near $2\pi f_h\tau = 1$ for the two filter shapes has not, to our knowledge, been investigated before. Figures 3 to 6 show similar results for the other common noise types. For a single noise type, the slope of $\sigma_y(\tau)$ versus τ, averaged over a decade in τ, can vary widely depending on the value of $2\pi f_h\tau$ and the type of filter used. The asymptotic expressions of $\sigma_y(\tau)$ for the various noise types with the infinitely sharp and single low-pass filters are given in Table 1. [1-3]

We have adapted the numerical techniques used to calculate $\sigma_y(\tau)$ to calculate $\mathrm{mod}\sigma_y(\tau)$ in the limit of an infinitely sharp filter of width f_h.

The mod sigma integral

$$\mathrm{mod}\sigma_y^2(n) = \frac{2}{n^4\pi^2\tau_0^2}\left[n\int_0^{f_h}\frac{S_y(f)}{f^2}\sin^4(\pi fn\tau_0)df + 2\int_0^{f_h}\sum_{k=1}^{n-1}(n-k)\frac{S_y(f)}{f^2}\cos(2\pi fk\tau_0)\sin^4(\pi fn\tau_0)df\right],$$

is more difficult to approximate because of the sum over k. Large values of n require computation of many integrals which takes a long time. The sum over k is eliminated by use of the following identity [5,6]

$$\sum_{k=1}^{n-1} (n-k)\cos\theta k = -\frac{n}{2} + \frac{1}{2}\frac{\sin^2\frac{\theta n}{2}}{\sin^2\frac{\theta}{2}}.$$

This identity is derived by looking at the real part of a sum of complex exponentials. Since cos is the real part of the complex exponential, this transforms the sum into a geometric series. Such a series can be summed explicitly yielding the final formula. Application of this identify to the original integral leads to the following form for the integral:

$$\mathrm{mod}\sigma_y^2(n) = \frac{2}{n^4\pi^2\tau_0^2}\int_0^{f_h} \frac{S_y(f)\sin^6(\pi\tau_0 nf)}{f^2\sin^2(\pi\tau_0 f)}df.$$

This integral can be evaluated in the same way as the $\sigma_y^2(n)$ integral, except that this integrand has a singularity for integral values of $\tau_0 f$, instead of the single singularity at f = 0. As in the case of the $\sigma_y^2(n)$ integral, it is necessary to use one routine near the singularities and another routine for those subintervals in which the integrand is highly oscillatory. It is therefore necessary to break up the interval of integration into $f_h\tau_0$ subintervals and sum the values of the integral over each of these smaller intervals. If f_h is large this involves a large number of subintervals. Often, later subintervals contribute negligibly to the value of the integral. For this reason an estimate of the error produced by ignoring further contributions is made every few subintervals. If this error is small relative to the computed value of the integral up to that point, then no further calculation is done and the value returned by the program as $\mathrm{mod}\sigma_y^2(n)$ is the sum of the integrations done over the subintervals up to that point. Additional details can be found in [4].

Rather than show the explicit values of $\mathrm{mod}\sigma_y(\tau)$, we have chosen to represent the results as

$$R(n) = \frac{[\mathrm{mod}\sigma_y(\tau)]^2}{[\sigma_y(\tau)]^2}.$$

This form clearly shows the differences between $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ as a function of noise type. It is interesting that R(n) shows a dependence on $2\pi f_h\tau$ only for flicker phase noise (α = 1) using an infinitely sharp filter. For flicker phase and flicker frequency noise types the results given in Figure 7 and Table 2 differ somewhat from these given by previous investigators [7,8].

Discussion

We believe that this numerical approach to the calculation of both $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ will prove to be very useful. It offers more flexibility in the types of noise which can be considered and can easily accommodate various filter shapes. Such filter shapes can be used to limit the noise bandwidth or define the many types of servo systems found in modern frequency standards. Using this numerical approach, we have investigated the dependence of $\sigma_y(\tau)$and $\mathrm{mod}\sigma_y(\tau)$ on measurement time as a function of noise type, the value of $2\pi f_h\tau$, and filter shape. We have shown that the slope of $\sigma_y(\tau)$ or $\mathrm{mod}\sigma_y(\tau)$, which is normally used to identify the noise type, depends strongly on the value of $2\pi f_h\tau$ and the high-frequency filter shape.

References

1. J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, G.M. Winkler, Characterization of Frequency Stability, Proc. IEEE Trans. on I&M 20, 105-120 (1971).
2. D.W. Allan, H. Hellwig, P. Kartaschoff, J. Vanier, J. Vig, G.M.R. Winkler, and N.F. Yannoni, Standard Terminology for Fundamental Frequency and Time Metrology, Proc. of the 42nd Ann. SFC, 419-425 (1988).
3. P. LeSage and C. Audoin, Characterization of Measurement of Time and Frequency Stability, Radio Science 14, 521-539 (1979).
4. F.L. Walls, J. Gary, A. O'Gallagher, R. Sweet and L. Sweet, Time Domain Frequency Stability Calculated from the Frequency Domain Description: Use of the SIGINT Software Package to Calculate Time Domain Frequency Stability from the Frequency Domain, NISTIR 89-3916.
5. V.F. Kroupa and L. Sojdr, Modified Allan Variance: Transfer Functions and Practical Applications, CPEM 1984 Digest.
6. P. Tremblay in Etude de l'Effet du Filliage du Signal sur la Caracterisation de la Stabilite de Frequence, These Universite Laval, Quebec 1985.
7. D.W. Allan and J.A. Barnes, A Modified "Allan Variance" with Increased Oscillator Characterization Ability, Proc. of the 35th Ann. SFC, 470-475 (1981).
8. P. Lesage and T. Ayi, Characterization of Frequency Stability: Analysis of the Modifed Allan Variance and Properties of Its Estimate, IEEE Trans. on I&M, 1984, IM-33, 332-337.

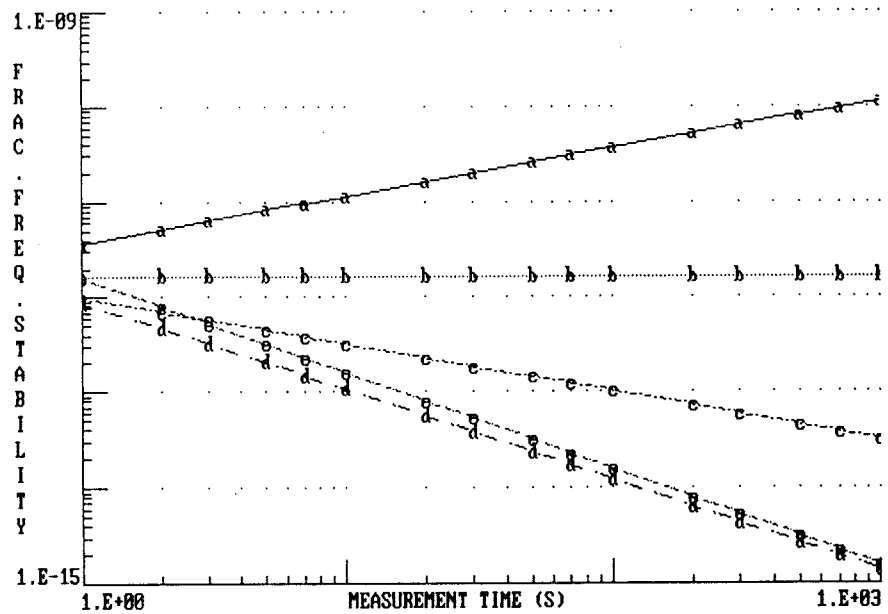


Figure 1. $\sigma_y(\tau)$ versus τ for the five common power-law noise types in the limit that $2\pi f_h \tau$ is large compared to 1 and an infinitely sharp filter is used. Curve a is for random-walk frequency modulation, $S_y(f) = h_{-2}\ f^{-2}$. Curve b is for flicker frequency modulation, $S_y(f) = h_{-1}\ f^{-1}$. Curve c is for white frequency modulation, $S_y(f) = h_0$. Curve d is for flicker phase modulation, $S_y(f) = h_1\ f$. Curve e is for white phase modulation, $S_y(f) = h_2\ f^2$.

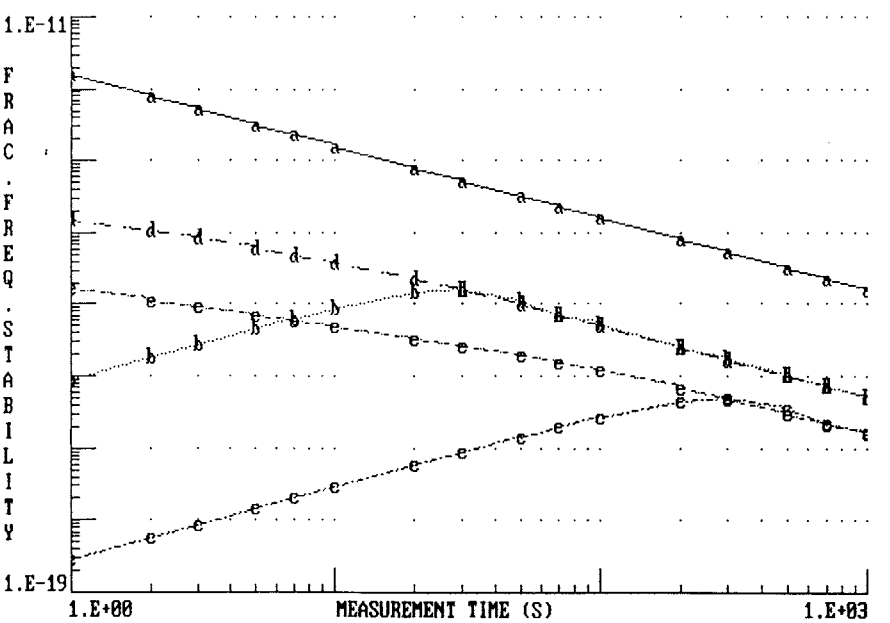


Figure 2. $\sigma_y(\tau)$ for white phase modulation ($\alpha = 2$) as a function of measurement time, τ, and measurement bandwidth, f_h. Curves a, b, and c have an infinitely sharp filter with width, f_h = 16 Hz, f_h = 0.016 Hz, f_h = 0.0016 Hz respectively. Curves d and e have a single pole filter width, f_h = 0.016 Hz and f_h = 0.0016 respectively.

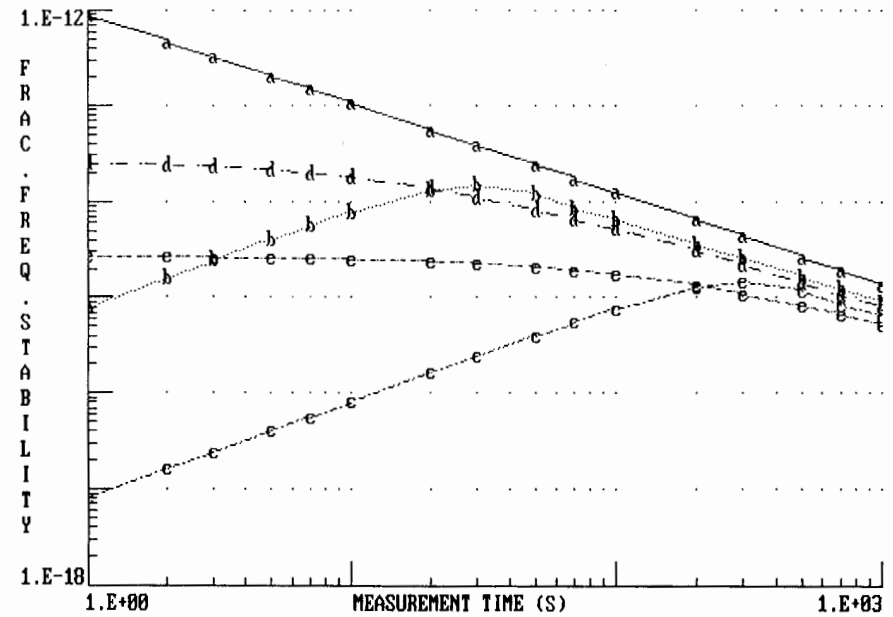


Figure 3. $\sigma_y(\tau)$ for flicker phase frequency modulation ($\alpha = 1$) as a function of measurement time, τ, and measurement bandwidth, f_h. Curves a, b, and c have an infinitely sharp filter with width, f_h = 16 Hz, f_h = 0.016 Hz, f_h = 0.0016 Hz respectively. Curves d and e have a single pole filter width, f_h = 0.016 Hz and f_h = 0.0016 respectively.

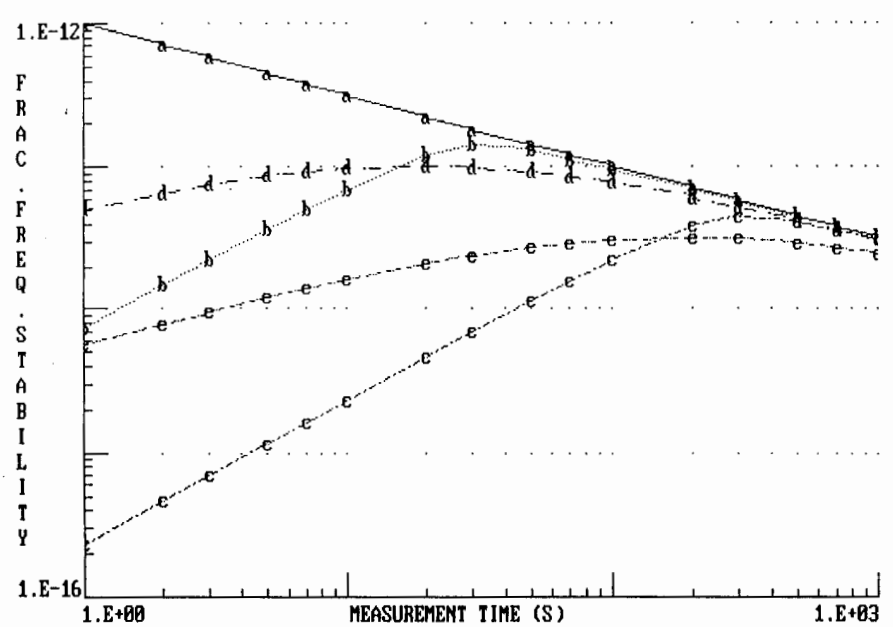


Figure 4. $\sigma_y(\tau)$ for white frequency modulation ($\alpha = 0$) as a function of measurement time, τ, and measurement bandwidth, f_h. Curves a, b, and c have an infinitely sharp filter with width, f_h = 16 Hz, f_h = 0.016 Hz, f_h = 0.0016 Hz respectively. Curves d and e have a single pole filter width, f_h = 0.016 Hz and f_h = 0.0016 respectively.

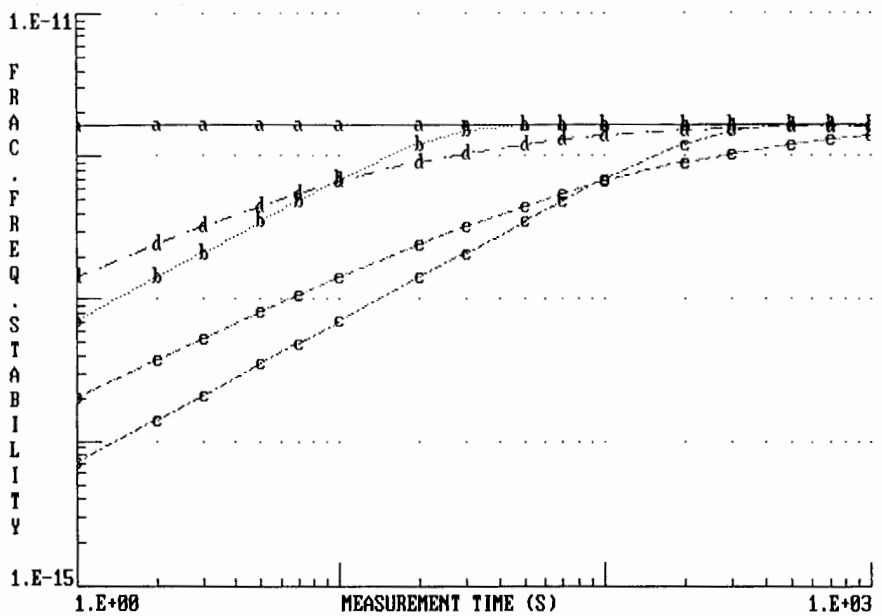


Figure 5. $\sigma_y(\tau)$ for flicker frequency modulation (α = -1) as a function of measurement time, τ, and measurement bandwidth, f_h. Curves a, b, and c have an infinitely sharp filter with width, f_h = 16 Hz, f_h = 0.016 Hz, f_h = 0.0016 Hz respectively. Curves d and e have a single pole filter width, f_h = 0.016 Hz and f_h = 0.0016 respectively.

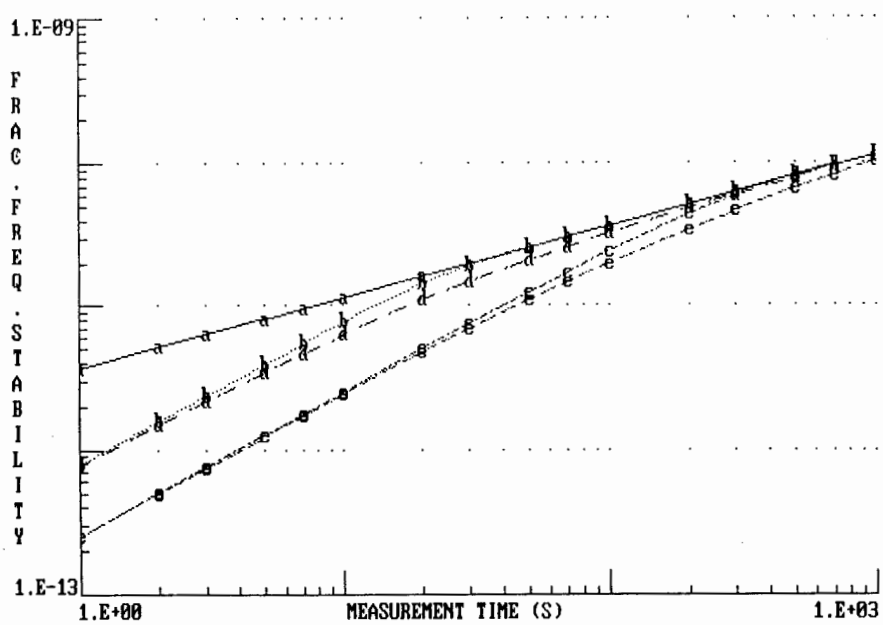


Figure 6. $\sigma_y(\tau)$ for random walk frequency modulation (α = -2) as a function of measurement time and measurement bandwidth, f_h for π. Curves a, b, and c have an infinitely sharp filter with width, f_h = 16 Hz, f_h = 0.016 Hz, f_h = 0.0016 Hz respectively. Curves d and e have a single pole filter width, f_h = 0.016 Hz and f_h = 0.0016 respectively.

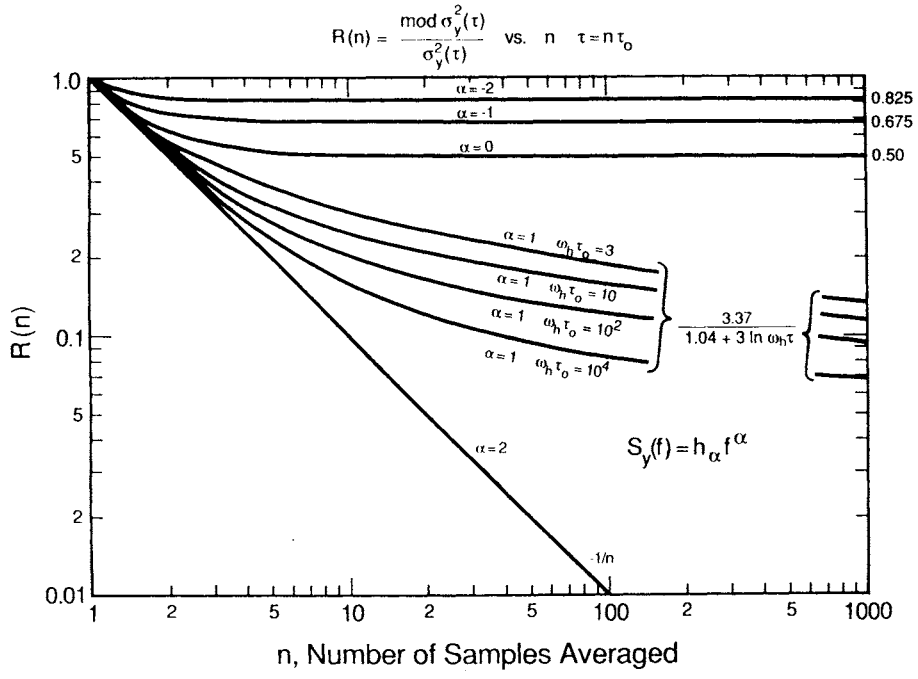


Figure 7. Ratio of $\mathrm{mod}\sigma_y(\tau)^2$ to $\sigma_y(\tau)^2$ as a function of n, the number of points averaged to obtain $\mathrm{mod}\sigma_y(\tau)$. The measurement time $\tau = n\tau_0$, where τ_0 is the minimum data interval.

TABLE I. Asymptotic forms of $\sigma_y^2(\tau)$ for various power-law noise types and two filter types. Note: $\omega_h/2\pi = f_h$ is the measurement system bandwidth--often called the high-frequency cutoff. ln ≡ loge.

Name of Noise	α	$S_y(f)$	$\sigma_y^2(\tau)$			
			$\omega_h\tau>>1$ Infinite Sharp Filter	$\omega_h\tau>>1$ Single Pole Filter	$\omega_h\tau<<1$ Infinite Sharp Filter	$\omega_h\tau<<1$ Single Pole Filter
White Phase	2	$h_2 f^2$	$\frac{3f_h h_2}{(2\pi)^2\tau^2}$	$\frac{3f_h h_2}{(2\pi)^2\tau^2}$	$2/5\pi^2 f_h{}^5\tau^2 h_2$	$\frac{f_h{}^2 h_2}{2\tau}$
Flicker Phase	1	$h_1 f$	$\frac{(1.038 + 3\ln(\omega_h\tau))h_1}{(2\pi)^2\tau^2}$	$\frac{(3\ln(\omega_h\tau))h_1}{(2\pi)^2\tau^2}$	$\frac{1}{2}\pi^2 f_h{}^4\tau^2 h_1$	$2f_h{}^2(\ln(2))h_1$
White Frequency	0	h_0	$\frac{h_0}{2\tau}$	$\frac{h_0}{2\tau}$	$2/3\pi^2 f_h{}^3\tau^2 h_0$	$2/3\pi^2 f_h{}^2\tau h_0$
Flicker Frequency	-1	$h_{-1} f^{-1}$	$2(\ln(2))h_{-1}$	$2(\ln(2))h_{-1}$	$\pi^2 f_h{}^2\tau^2 h_{-1}$	$8\pi^2 f_h{}^2\tau^2 h_{-1}$
Random-Walk Frequency	-2	$h_{-2} f^{-2}$	$\frac{2\pi^2\tau h_{-2}}{3}$	$\frac{2\pi^2\tau h_{-2}}{3}$	$2\pi^2 f_h\tau^2 h_{-2}$	$2\pi^2 f_h\tau^2 h_{-2}$

TABLE II. Ratio of $\mathrm{mod}\sigma_y^2(\tau)$ to $\sigma_y^2(\tau)$ vs n, for common power-law noise types $S_y(f) = h_\alpha f^\alpha$. n is the number of time or phase samples averaged to obtain $\mathrm{mod}\sigma_y^2(\tau = n\tau_0)$ where τ_0 is the minimum sample time, and ω_h is 2π times the measurement bandwidth f_h.

n	$\alpha = -2$	$\alpha = -1$	$\alpha = 0$	$\alpha = +1$				$\alpha = +2$
				$\omega_h\tau_0 = 3$	$\omega_h\tau_0 = 10$	$\omega_h\tau_0 = 100$	$\omega_h\tau_0 = 10^4$	
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.859	0.738	0.616	0.568	0.543	0.525	0.504	0.500
3	0.840	0.701	0.551	0.481	0.418	0.384	0.355	0.330
4	0.831	0.681	0.530	0.405	0.359	0.317	0.284	0.250
5	0.830	0.684	0.517	0.386	0.324	0.279	0.241	0.200
6	0.828	0.681	0.514	0.349	0.301	0.251	0.214	0.167
7	0.827	0.679	0.507	0.343	0.283	0.235	0.195	0.143
8	0.827	0.678	0.506	0.319	0.271	0.219	0.180	0.125
10	0.826	0.677	0.504	0.299	0.253	0.203	0.160	0.100
14	0.826	0.675	0.502	0.274	0.230	0.179	0.137	0.0714
20	0.825	0.675	0.501	0.253	0.210	0.163	0.119	0.0500
30	0.825	0.675	0.500	0.233	0.194	0.148	0.106	0.0333
50	0.825	0.675	0.500	0.210	0.176	0.134	0.0938	0.0200
100	0.825	0.675	0.500	0.186	0.159	0.121	0.0837	0.0100
Limit	0.825	0.675	0.500		$\longleftarrow \left(\frac{3.37}{1.04 + 3\ \ln\omega_h\tau}\right) \longrightarrow$			1/n

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-01 Number 12

Analysis of Fractal and Pareto-Levy Sets : Theory and Application

Stephen D. Casey
The American University
The Department of Mathematics and Statistics
4400 Massachusetts Ave.,N.W.
Washington, D.C. 20016

1 Introduction : Synthesis and Analysis of Fractal and Pareto-Levy Sets

Most natural objects and many man-made processes exhibit great detail and scaling invariance. Given some underlying geometry, these objects and processes can be modeled as *fractal sets*. If no such geometry exists, the theory of Pareto-Levy distributions can be used to analyze these objects. (In this case, we refer to the objects or processes as *Pareto-Levy sets*.) The study of these self-similar and statistically self-similar sets is unified under the theory of fractals.

The theory of fractals goes back to K. Weierstrass, who developed a continuous but nowhere differentiable curve (1874). This area became quite active in the 1970's, thanks in part to the work of B. Mandelbrot [M], who unified the theory as a single branch of mathematics. The theory is an extension (but by no means a replacement) of our current models of the world which are based on Euclidean geometry and smooth generalizations of that geometry, e.g. differentiable manifolds. It includes highly detailed objects produced by recursive applications of maps, while Euclidean geometry includes objects made by the conglomeration of Euclidean shapes — circles, lines, polygons, etc. — each of which can be generated by simple maps. Both visual inspection and statistical analysis lead us to believe that much of the natural world (e.g., clouds, mountains, coastlines, trees) and many man-made objects (e.g., spatial distributions of communication networks, temporal flow of messages in these networks, economic distribution of resources) can be modeled by this theory.

The digital computer has played a large role in the growth of this area. It is ideally suited to the recursive procedures needed to generate images which approximate fractal sets. These images, then, in turn fuel the interest in fractals. The images can be surreal and quite beautiful (as produced by iterates of f(z) = tan(z) [C]), can approximate images of natural objects (as produced by a mixed set of affine mappings producing an image of an orchard) or can approximate the structure of some man-made object (as in the case of diffusion limited aggregation approximating a phone network [ACI]). Because these synthesized images visually resemble the objects that they are modeling so well, we are lead to believe that fractal geometry will extend our ability to model natural and man-made phenomena beyond our current models.

There is, however, a major gap in this approach to modeling. At this time, there is no mathematically rigorous method for synthesizing an arbitrary image to an arbitrary degree of accuracy. Significant progress on this problem had been made by Barnsley and others (see [B] and the references therein), who have unified the field of deterministic fractal sets by introducing *Iterated Function Systems* (IFS). (An IFS is a finite set of contraction mappings on a given complete metric space.) They have developed an existence theorem, which says that given an arbitrary image, there exists an IFS which has an attactor which is arbitrarily close (as measured by the Hausdorff metric on sets) to the given image. However, the specific structure of this IFS is not known - even the number of mappings contained in it is an open question.

We look into developing this model from another angle, that of analysis. Given an arbitrary set which exhibits a high level of detail and a statisical scaling invariance, we wish to extact measures of that set that reflect its structure. Moreover, we wish to determine what these measures tell us about the set, e.g., if the structure too sparse, dense, or irregular, if it breaks down when it should not, or even if it appears unexpectedly.

As experimentalists know, real world data usually does not conform nicely to any model. Fortunately, sets that have scaling invariance can be measured by processes that are sufficiently robust to extract meaningful information.

2 Topological, Fractal, and Hausdorff-Besicovitch Dimensions

Given a compact set, we wish to calculate the *dimension* of that set. We want this dimension to reflect the highly detailed structure of the set, if it has one. Consider the following six examples. The first two are the unit interval $I = [0, 1]$ and the unit square $I^2 = [0, 1] \times [0, 1]$. The next three are more complicated, and require some construction.

In 1883 G. Cantor constructed a perfect infinitely disconnected set as follows. Given $[0,1]$, remove the open middle third segment $(1/3, 2/3)$. Then, remove the open middle thirds of the remaining two segments, i.e., remove $(1/9, 2/9)$ from $[0, 1/3]$ and $(7/9, 8/9)$ from $[2/3, 1]$. Repeat. At the n^{th} stage of this process, there will be 2^n segments, each of length $(1/3)^n$. Continue ad infinitum.

The total length of the removed set is $(1/3)\sum_{i=0}^{\infty}(2/3)^n$, which equals one. However, the remaining set C is a perfect, and thus an uncountable, set which is totally disconnected (see [HY]).

H. von Koch's (1904) snowflake K curve is an example of a simple closed curve of infinite length fitting inside a finite area. The limit curve also has no tangent lines.

The Koch construction is to replace the segment $[0,1]$ with four segments of length $1/3$ with vertices $(1,0), (\frac{1}{3},0), (\frac{1}{2},\frac{\sqrt{3}}{6}), (\frac{2}{3},0), (1,0)$. Then, each of the four segments is replaced by this pattern, and so on. At the n^{th} stage the curve consists of 4^n segments of length $(1/3)^n$. As n goes to infinity, the total length also goes to infinity.

Starting with a filled-in equilateral triangle, perform a Cantor-like removal process by removing the middle equilateral triangle whose vertices are the midpoints of the three edges. Repeat this process on the three remaining filled-in triangles, and then the nine remaining ones, and so on. The remaining set S_g has a removed triangle in any neighborhood of any point in the set. This construction is due to W. Sierpinski (1915).

The final example is the set $H = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\} \cup \{0\}$.

We discuss three methods of measuring these sets. The first is the topological dimension, which can be thought of as an integer n which describes which Euclidean n-space $\mathbf{R}^n$ that a given set resembles. Locally the topological dimension is an accurate determination of topological information about the set - it does not measure the geometry of the set.

Definition: *The Topological Dimension* of a set X ($D_T(X)$) (F. Browder (1911), H. Poincare (1912)) : Let X be in $\mathbf{R}^n$, and A be any open cover of X. Then B is a refinement of A of order m if some point covered by A lies in m elements of B, and no point covered by A lies in more than m elements of B. Then,

$$D_T(X) = \inf_{covers}\{m+1 : A \text{ has refinment of order } m+1\}.$$

The topological dimensions of C and H equal zero. As I, K, and S_g are curves, they have topological dimension 1. For I^2 , $D_T = 2$. Proofs of these facts can be quite involved (see [HW] for further reference).

The second dimension is, for our purposes, the most practical of the three, as it can not only measure more of the structure of a set, but can also be determined experimentally. This is the Kolmogorov-Mandelbrot, or fractal dimension D_{KM} of the set. We shall see, however, that although it is easier to calculate, it may give results which inaccurately reflect the geometry of a set. It was introduced by L. Pontrjagin (1932), and studied by Kolmogorov (1950's) and Mandelbrot (1970's) (see [M] for more information).

Consider any set in n-dimensional Euclidean space, $\mathbf{R}^n$. Let r be any positive number, and let $\mathcal{N}(r)$ be the *minimal* number of closed line segments, balls, or spheres of radius r centered on an element of the set needed to cover the set. Then we want to calculate a number D such that as $r \to 0$,

$$\mathcal{N}(r) \cdot r^D \sim 1. \quad \text{(Scaling Relationship)}$$

Consider first a single point. For any $r > 0$, we need only one line segment to cover. Thus, $\mathcal{N}(r) \cdot r^D = 1 \cdot r^D = 1$ for $D_{KM} = 0$.

Next consider $[0,1]$. Let $\lceil r \rceil$ be the smallest integer greater than or equal to r. Given any $r > 0$, we need $\lceil 1/r \rceil$ line segments of length r to cover $[0,1]$. Thus $\mathcal{N}(r) \cdot r^D = \lceil 1/r \rceil \cdot r^D \sim 1$, for $D_{KM} = 1$.

The set $[0,1] \times [0,1]$ can be covered by $\lceil 1/r \rceil^2$ balls of radius r, so in this case D_{KM} is 2.

In these first three examples, the value D_{KM} agrees with our intuitive concept of dimension which is topological dimension. The next example breaks from this concept.

Consider the Cantor Middle Thirds set C. At the nth stage of its construction, we need 2^n line segments of radius $(1/3)^n$ to cover. As this is true for all n,

$$\mathcal{N}(r) \cdot r^D = 2^n(1/3)^{nD}.$$

Therefore,

$$\begin{array}{rrcl} & \mathcal{N}(r) \cdot r^D & = & 1 \\ \Leftrightarrow & \log(2^n(1/3)^{nD}) & = & \log 1 \\ \Leftrightarrow & n(\log 2 - D\log 3) & = & 0. \end{array}$$

(where log denotes the natural logarithm). Thus, $D_{KM} = \log 2/\log 3$.

These examples motivate the following:

Definition: *The fractal dimension* $D_{KM}(X)$ of a set X is:

$$D_{KM} = \lim_{r \to 0} \frac{\log(\mathcal{N}(r))}{\log(1/r)}.$$

We note that D_{KM} is the number necessary to preserve the scaling relationship $\mathcal{N}(r) \cdot r^D = 1$ as $r \to 0$. The fractal dimension of a self-similar seed fractal set can be derived from its generating pattern. We also note that if $r = a^n, 0 < a < 1$, then

$$D_{KM} = \lim_{n \to \infty} \frac{\log(\mathcal{N}(a^n))}{\log(1/a^n)}.$$

For the Koch curve described above we have, at the nth stage, 4^n segments of length $(1/3)^n$. Thus $D_{KM} = \frac{\log 4}{\log 3}$. The Sierpinski gasket at the n^{th} stage consists of 3^n equilateral triangles with sides of length $(1/2)^n$. Therefore $D_{KM} = \frac{\log 3}{\log 2}$.

Finally, we introduce the Hausdorff-Besicovitch dimension. This dimension is an essential tool of geometric measure theory.

Definition—*The Hausdorff-Besicovitch Dimension* $D_{HB}(X)$ *of a set* X: Consider the set of all countable covers of X by sets of diameter less than $r > 0$, and let r_n be the diameter of the n-th set in any given cover. Then, for $\beta > 0$, the *β-th Hausdorff-Besicovitch outer measure of* X is

$$\mu^*_\beta(X) = \lim_{r \to 0}\left[\inf_{\{\text{covers}\}}\left[\sum_{n=1}^{\infty} r_n^\beta\right]\right].$$

The Hausdorff-Besicovitch dimension is then

$$D_{HB}(X) = \inf\{\beta : \mu^*_\beta(X) = 0\}$$

(Hausdorff (1919), Besicovitch (1934)).

Equivalently, we could define

$$D_{HB}(X) = \sup\{\beta : \mu^*_\beta(X) = \infty\}.$$

For the examples above, we have that $D_{HB}(I) = 1$, $D_{HB}(I^2) = 2$, $D_{HB}(C) = \frac{\log 2}{\log 3}$, $D_{HB}(K) = \frac{\log 4}{\log 3}$, and $D_{HB}(S_g) = \frac{\log 3}{\log 2}$.

Verification of these values requires more involved estimates than in the case of D_{KM} (see [F]).

The following theorem follows almost directly from the definitions.

Theorem: Let X be any compact subset of a complete metric space. Then

$$D_T(X) \leq D_{HB}(X) \leq D_{KM}(X).$$

The set H is an example of a set for which $D_{HB}(X) < D_{KM}(X)$. In fact, $D_{HB}(X) = 0$, while $D_{KM}(H) = 1/2$. Verification of these numbers is left as an exercise for the reader. Note that in this case, as H has only one "scaling point" (namely 0), $D_{HB}(X)$ is the correct number.

3 Experimental Determination of Fractal Dimension

Sets that exhibit a scaling relationship can have this structure measured in various ways. The first and most obvious method is to extract this measure from a graphical representation of the set. This is most easily demonstrated by considering subsets of Euclidean m-space $\mathbf{R}^m$.

Let I, I^2 be as in 2. Let X be the set we wish to measure. If has topological dimension 0 or 1, we can get the "fractal length" of X by measuring X by rulers of length $r \cdot I$. Let $\mathcal{L}(r)$ be the minimal number of units $r \cdot I$ placed end to end needed to measure X . Then, X has "length" $\mathcal{L}(r) \cdot r$, which asymptotically gives

$$\mathcal{L} \sim \frac{1}{r^{D_{KM}-1}}, \text{ or } D_{KM} \sim \frac{\log \mathcal{L}(r)}{\log(1/r)} + 1.$$

Alternatively, we could calculate a "fractal box dimension." A closed circle of radius p can be contained in a closed square of side length 2ρ, which in turn can be contained in a closed circle of radius $\sqrt{2}\rho$. Therefore, if we let $\mathcal{N}(r)$ denote the minimal number of closed squares of side length r needed to cover the set, we get (by the "pinching theorem" from calculus)

$$D_{KM} \sim \frac{\log \mathcal{N}(r)}{\log(1/r)}.$$

Finally, we can get a "fractal grid dimension" by using a grid of mesh size $\frac{1}{2^n}$. Let $\mathcal{G}(n)$ denote the number of elements of the grid (including boundary) that contain an element of X. Then, for $\mathcal{N}(r)$ as above, we have in $\mathbf{R}^m$, m=1,2,...

$$2^{-m}\mathcal{G}(n-1) \leq \mathcal{N}(1/2^n) \leq \mathcal{G}(k(n))$$

where $k(n)$ is the smallest integer satisfying $k \geq n - 1 + \log_2 m$. This inequality then in turn gives

$$D_{KM} \sim \frac{\log \mathcal{G}(n)}{\log(1/2^n)}$$

(see [B]).

A more general and more robust technique is to use statistical measurement. For the more complicated Hausdorff-Besicovitch measure, a related probability distribution function is the Pareto-Levy distribution, introduced by P. Levy and A. Khinchine (1936) (see [W]). These distributions are stable, and most can be written down only in terms of their Fourier transforms. The Pareto-Levy is linked to the Hausdorff-Besicovitch dimension in that the distribution scales in the same manner as the outer measure, and, given a controlled random walk or more general process producing a family of curves or surfaces, those surfaces will have HB dimension which is a function of a parameter of that distribution.

Just as the HB dimension has its simpler counterpart, so too does the Pareto-Levy reduces to the simpler Pareto distribution, first studied by V. Pareto (1897).

This has probability distribution function

$$f(x) = \begin{cases} \frac{\alpha x_0^\alpha}{x^{\alpha+1}} & x \geq x_0 \\ 0 & x < x_0. \end{cases}$$

Then, the cumulative distribution function is

$$F(x) = 1 - (\frac{x}{x_0})^{-\alpha} = 1 - \exp(-\alpha(\log(x) - \log(x_0))).$$

Thus by mapping the domain onto log((domain)), we transform the Pareto into the exponential. Now, the exponential is "memoryless," i.e.,

$$P_e(x > a + b \mid x > b) = P_e(x > a).$$

Thus, for the Pareto,

$$P_p(x > \tilde{a} \cdot \tilde{b} \mid x > \tilde{b}) = P_p(x > \tilde{a})$$

where $\tilde{a}$, $\tilde{b}$ are of the form e^a, e^b and $\tilde{a}, \tilde{b} \geq x_0$,i.e., the Pareto is "scale invariant." Also note, the r-th moment of the Pareto about the origin is

$$\langle |x|^r \rangle = \frac{\alpha}{\alpha - r}$$

which converges for $r < \alpha$, and diverges for $\alpha \geq r$. Finally note that $f(x)$ is linear on a log-log scale. But

$$D = \frac{\log \mathcal{N}(x)}{\log(1/x)},$$
$$\log \mathcal{N}(x) = -D \log(x), \text{ and}$$
$$\mathcal{N}(x) = (x)^{-D}.$$

Thus we have the relationship

$$\textbf{Pareto}(\alpha + 1) \longleftrightarrow \textbf{KM Dimension}(D).$$

Both have a linear relation in log-log, namely

$$\log \mathcal{N}(x) = -D \log(x)$$
$$\log f(x) = -(\alpha + 1) \log(x).$$

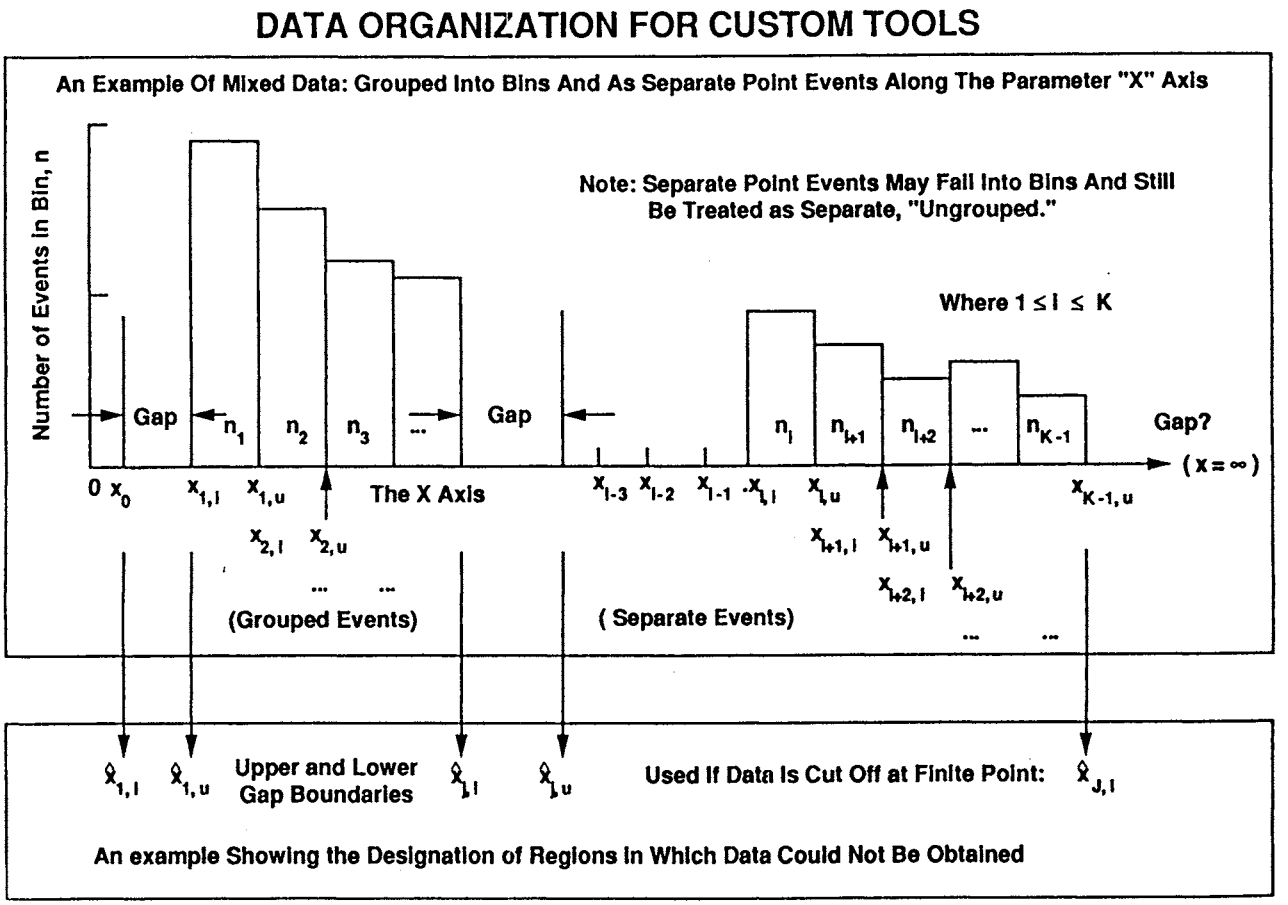
The connection between the Pareto distribution and the fractal dimension has proven to be the most effective method of extracting fractal dimension (see [AC1, AC2]). In [AC1], numerous data sets were examined. First, the log of the data was taken, and then this data was fit to an expotential by maximumum likelihood estimation. The quality of this fit was then tested by χ^2 and Kolmogorov-Smirnov tests. A scaling structure was evident in many different sets, including errors in communication lines, delays in the delivery of messages, the structure of communication networks, and even European city sizes (see [AC1]). (This last example gives more evidence to a theorem in human geography

introduced by G. Zipf.) Work of Mandelbrot and others (see [M]) tells us that these methods also will extract useful measures of many varieties of noise.

We close this paper with the following result (see also [AC1, AC2]). Assume that the data we have gathered is a subset of a larger data set which has a scaling structure, and that data is truncated, grouped into bins and/or consists of separate data points. Then we have

Theorem: Let X be a truncated, binned and/or isolated subset of a set with a scaling structure of parameter D. Then this parameter D may be accurately estimated by using the maximum likelihood estimation on the Pareto distribution.

proof: The proof is sketched in the following figure:



LIKELIHOOD FOR EXPONENTIAL AND HYPERBOLIC

If the Distribution Is Hyperbolic, It Is Converted to an Equivalent Exponential

Hyperbolic

Exponential

$$P(u_0 < U < u) = 1 - \left(\frac{u}{u_0}\right)^{-\alpha}$$

$$= 1 - e^{-\alpha(\log u - \log u_0)}$$

$$\equiv 1 - e^{-\alpha(x' - x'_0)}$$

The Cumulative Hyperbolic Distribution (and its Representation as an Exponential)

$$p(u) = \alpha \left(\frac{u}{u_0}\right)^{-\alpha-1} du$$

$$= \alpha\, e^{-\alpha(\log u - \log u_0)} \frac{du}{u}$$

$$= \alpha\, e^{-\alpha(x'' - x''_0)}\, dx''$$

Corresponding Density Function (Agrees).

$$P(x_0 < x' < x) = 1 - e^{-\alpha(x - x_0)}$$

The Cumulative Exponential Distribution

$$p(x') = \alpha e^{-\alpha(x' - x_0)}\, dx'$$

The Exponential Probability Distribution Density Function

Grouped | Event Probability Depends on Whether Data Is Binned or Discrete: | Separate Point Events

$$L_l = \frac{e^{-\alpha(x_{l,l} - x_0)} - e^{-\alpha(x_{l,u} - x_0)}}{\sum_{j \in B} \left\{ e^{-\alpha(\hat{x}_{j,l} - x_0)} - e^{-\alpha(\hat{x}_{j,u} - x_0)} \right\}}$$

Probability Of An Event Landing In Bin "l"

$$\tilde{L}_l = \frac{\alpha\, e^{-\alpha(x_l - x_0)}}{\sum_{j \in B} \left\{ e^{-\alpha(\hat{x}_{j,l} - x_0)} - e^{-\alpha(\hat{x}_{j,u} - x_0)} \right\}}$$

Contribution from a Discrete Event at x_l .

LIKELIHOOD FOR EXPONENTIAL DISTRIBUTION

The Likelihood of Observing Exactly the Number of Counts Per Bin Seen Is:

$$L = \prod_{l \in B} L_l^{n_l} \cdot \prod_{j \in D}^{N_D} \tilde{L}_j$$

Note that L Is a Function of α, and Can Be Expressed as a Natural Lograrithm:

$$\text{Log}\, L(\alpha) = \sum_{l \in B} n_l \log L_l(\alpha) + \sum_{j \in D}^{N_D} \log \tilde{L}_j(\alpha)$$

Define New Variables for Computation of Likelihood:

Shift the origin: $x - x_0 \rightarrow x$

Let $(x_{l,l} + x_{l,u})/2 \equiv x_l \geq 0$

Let $(x_{l,l} - x_{l,u})/2 \equiv \Delta x_l \leq 0$

Let the normalizing denominator be simply $\sum_{j \in B}(\alpha)$

Then the numerator for the L_l can be written:

$$e^{-\alpha x_l} \cdot \{ e^{-\alpha \Delta x_l} - e^{+\alpha \Delta x_l} \}$$

Now the Expression for the Logarithm of the Likelihood Becomes:

$$\text{Log}\, L(\alpha) = \sum_{l \in B} n_l(-\alpha x_l) + \sum_{l \in B} n_l \log \{ e^{-\alpha \Delta x_l} - e^{+\alpha \Delta x_l} \} - \sum_{l \in B} n_l \log \sum_{j \in D} \left(e^{-\alpha \hat{x}_{j,l}} - e^{-\alpha \hat{x}_{j,u}} \right) + N_D \log \alpha$$

If There Is a Last Bin That Extends to Infinity, One Must Add Another Term:

$- \alpha\, n_K x_K$ Where x_K Is the Lower Bound of the Bin Extending to Infinity.

MAXIMUM LIKELIHOOD SOLUTION

The Maximum Likelihood as α Varies Occurs where the First Derivative Vanishes (If the Second Is Negative, and the Third Derivative Is Bounded); the First Derivative of Log L (α) Is:

$$\frac{d}{d\alpha}\log L(\alpha) = -\sum_{I\in B,D} n_I x_I + (-n_K x_K\ ?) - \sum_{I\in B} n_I \Delta x_I \frac{e^{-\alpha\Delta x_I} + e^{+\alpha\Delta x_I}}{e^{-\alpha\Delta x_I} - e^{+\alpha\Delta x_I}} + N\frac{\sum_{k\in B}\{\hat{x}_{k,l}e^{-\alpha\hat{x}_{k,l}} - \hat{x}_{k,u}e^{-\alpha\hat{x}_{ku}}\}}{\sum_{j\in B}(\alpha)} + \frac{N_D}{\alpha} = 0$$

(1) (Include for Both Separate and Grouped Events.)

(2) (Include If Last Bin Extends to Infinity.)

(3) (Finite Bin Width Term)

(4) (Normalization for Missing Data)

(5) Separate Events

(Maximum Likelihood Condition.)

By Taylor Expansion (3) Becomes:

$$(3) = \frac{-1}{\alpha}\sum_{I\in B} n_I \frac{1 + \frac{(\alpha\Delta x_I)^2}{2!} + \frac{(\alpha\Delta x_I)^4}{4!} + \ldots}{1 + \frac{(\alpha\Delta x_I)^2}{3!} + \frac{(\alpha\Delta x_I)^4}{5!} + \ldots}$$

$$= \frac{-1}{\alpha}\sum_{I\in B} n_I\{1 + \frac{(\alpha\Delta x_I)^2}{3} + \ldots\}$$

$$= \frac{-N}{\alpha} + \frac{-1}{3\alpha}\sum_{I\in B} n_I(\alpha\Delta x_I)^2 + \text{4th Order On Up}$$

So (1) (3) & (5) Yield the Familiar Estimator for α:

$$\alpha = \frac{N_D + N - n_k}{\sum_{I\in B,D} n_I x_I + n_k x_k}$$

(Exact for Small Bins e.g., $|\alpha\Delta x_I| << 1.0$, and No Missing Data.)

Solve by Iteration Unless (3) and (4) Vanish.

Initial Approximation of α Is Given by:

Iteration Algorithm Is As Follows:

Estimate the Standard Deviation as Indicated on the Next Chart.

Step to Right and Left by Standard Deviation Increments Until (1) Changes Sign.

Linearly Interpolate to Find Zero Estimate.

Evaluate (1) at Interpolated Value, and Check Sign Change for Right or Left Interval.

Linearly Interpolate Again, and Iterate Until Convergence Is Reached.

UNCERTAINTY ON THE FITTED EXPONENT

The Uncertainty on the Fitted Exponent Can Be Considered to Be the Standard Deviation of α which Is Known to Be Related to the Second Derivative of the Log Likelihood for Distributions with Finite Variance. For the Fit to an Exponential Distribution, this Second Derivative Is Simple to Evaluate If the Bins Are Vanishingly Small and the Full Range of Data Is Available. From the Approximation to (3) the Leading Term Is the Same as (5), and Both Can Be Differentiated to Yield:

$$\frac{d}{d\alpha}(5) = \frac{d}{d\alpha}\left(\frac{N}{\alpha}\right) = -\frac{N}{\alpha^2} = -\frac{1}{\sigma_{\bar{\alpha}}^2}$$ So the Standard Deviation Is $\sigma_{\bar{\alpha}} = \frac{\bar{\alpha}}{\sqrt{N}}$ the Familiar Result.

When the Bins Are Large or the Data Range Is Restricted, the Second Derivative of the Full Expression Must Be Used. The Second Derivative Is the Derivative of the First Derivative, Expression. By Inspection, only Two More Terms Contribute, as Shown Below:

$$\frac{d}{d\alpha}(3) = -\sum_{I\in B}\frac{4n_I\Delta x_I^2}{(e^{-\alpha\Delta x_I} - e^{+\alpha\Delta x_I})^2}$$ If Δx_I Not Zero, Variance Is Increased Due to Bin Width.

$$\frac{d}{d\alpha}(4) = \frac{d^2}{d\alpha^2}\{\sum_{I\in B} n_I \log\frac{1}{\sum_{j\in\hat{B}}(\alpha)}\} = -N\frac{\sum_{j\in\hat{B}}(\alpha)\cdot\frac{d^2}{d\alpha^2}\sum_{k\in\hat{B}}(\alpha) - \{\frac{d}{d\alpha}\sum_{k\in\hat{B}}(\alpha)\}^2}{\{\sum_{j\in\hat{B}}(\alpha)\}^2}$$ Variance Growth Due to Limited Data Range.

$$= -N\frac{\sum_{j\in\hat{B}}(e^{-\alpha\hat{x}_{jl}} - e^{-\alpha\hat{x}_{ju}})\cdot\sum_{k\in\hat{B}}(\hat{x}_{k,l}^2 e^{-\alpha\hat{x}_{kl}} - \hat{x}_{k,u}^2 e^{-\alpha\hat{x}_{ku}}) - \{\sum_{k\in\hat{B}}(\hat{x}_{k,l}e^{-\alpha\hat{x}_{kl}} - \hat{x}_{k,u}e^{-\alpha\hat{x}_{ku}})\}^2}{\{\sum_{j\in\hat{B}}(e^{-\alpha\hat{x}_{jl}} - e^{-\alpha\hat{x}_{ju}})\}^2}$$

CHI SQUARE ESTIMATE OF QUALITY OF FIT

Chi Square Test of Quality of Fit Requires Binned Data, Especially in the Highly Populated Regions of the Independent Variable. The Bins in the Sparsely Populated Tails of the Distribution Can Be Empty or of Low Count and Still Obtain Reasonable Chi Square Values.

The Chi Square Is Formed by First Estimating the Expected Number of Counts $\bar{n}_i$ in Each Bin:

$$\bar{n}_i = N\ L_i(\bar{\alpha})$$

And Chi Square Is then:

$$X^2 = \sum_{i \in B} \frac{(n_i - \bar{n}_i)^2}{\bar{n}_i}$$

Note that Discrete Data Must Be Binned for the Chi-Square Test.

Note that Some n_i Can Vanish, yet $\bar{n}_i$ Will Not, So the Expression Will Not Be Pathological.

Note that Highlighting Bins with Large Values of:

$$\chi_i = \frac{n_i - \bar{n}_i}{\sqrt{\bar{n}_i}}$$

Reveals Extreme Deviates.

CHARACTERISTIC LARGEST VALUES FOR THE PARAMETERS, X OR U

Whenever there Are N Events in a Data Sample, the Characteristic Largest Value of the Parameter Is that for which the Cumulative Probability of Being Further Out in the Tail of the Distribution Is 1/N. Moreover, the Probability Distribution for the Value of the Parameter for the Largest Event Cuts Off Sharply Beyond the Characteristic Largest Value, So the Characteristic Largest Value Is a Good Estimate of what the Largest Value Will Be.

For the Exponential:

$$e^{-\alpha(x_N - x_0)} = \frac{1}{N} \quad , \text{so } x_N = \frac{\log N}{\alpha} + x_0$$

For the Hyperbolic:

$$\left(\frac{u_N}{u_0}\right)^{-\alpha} = \frac{1}{N} \quad , \text{so } u_N = N^{\frac{1}{\alpha}} u_0$$

Acknowledgements : The author would like to thank his colleagues Larry Crone and Jeff Hakim of AU, and Robert Anthony of IDA. Work reported on in this paper was done at IDA under Anthony's direction.

The interested reader should note that the books referenced in the bibliography all contain extensive bibliographies themselves. See especially [M].

References

[AC1] Anthony, R., and Casey, S., "Protype fractal reprsentation and analysis of C^3 systems (Pro-Fracs)," Institute of Defense Analysis Technical Report, 1990.

[AC2] Anthony, R., and Casey, S., "A Guide to the Pro-Fracs methods and tools," Institute of Defense Analysis Technical Report, 1990.

[B] Barnsely, M., *Fractals Everywhere*, Academic Press, Inc., Boston (1988).

[C] Casey, S., "Formulating fractals," *Computer Language*, Vol. 4, No. 4, April 1987, pp. 28-40, cover.

[F] Falconer, K., *The Geometry of Fractal Sets*, Cambridge University Press, Cambridge (1985).

[HY] Hocking, J. and Young, G., *Topology*, Dover Publications, New York (1988).

[HW] Hurewicz, W., and Wallman, H., *Dimension Theory*, Princeton University Press, Princeton, N.J. (1941).

[M] Mandelbrot, B., *The Fractal Geometry of Nature*, W. H. Freeman and Co., New York (1983).

[W] West, B., *An Essay on the Importance of Being Non-Linear*, Springer-Verlag, Inc., Berlin (1985).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO 01 Number 13

Investigations on time scale algorithms with simulated clock data

Helmut Brandenberger* and Léon Prost

Swiss Federal Office of Metrology, Wabern, Switzerland

Several algorithms are known to compute a time scale based on the time comparisons of a set of clocks that may be located at several remote sites. The results of the comparisons are corrupted by noise depending on the type of the clock and its method of comparison, especially for remote clocks. The clock parameters (deterministic and statistical) may only be estimated.
Based on the parameter estimation and the actual experimental results of the clock comparisons, a mean synthetic clock can be calculated. Two of the well known algorithms are the "ALGOS" algorithms used at BIPM, and the so called "Paperclock" algorithm developped and used at NIST. Both algorithms differ rather significantly from each other, especially for the estimation and prediction periods. To test their behaviour, a set of seven clocks having a relative frequency spectral density of:

$$S_y = h_{-1} f^{-1} + h_0 f^0$$

is simulated with different h_0 and h_{-1} parameters. The results of the simulation show, that the two algorithms react differently to slightly different initial conditions.

* H. Brandenberger has retired since 1987

Session PO-01 Number 14

IDENTIFICATION AND MEASUREMENT OF DIFFERENT TYPES OF NOISES PRESENT IN A FREQUENCY SOURCE.

F. Vernotte*, J. Groslambert** and J. J. Gagnepain**

*Observatoire de Besançon
41 bis, Avenue de l'Observatoire - 25044 BESANCON Cedex - FRANCE

**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 BESANCON - FRANCE

Oscillators are affected by phase and frequency instabilities due to several noise sources. It is important to identify these different noise sources in order to know their causes and to be able to correct them. Generally, a variance is used (Allan variance, Modified Allan variance, ...) to estimate the different types of noises present in a frequency source. However, the sensitivity of a variance, which is different for each type of noise, limits the precision of this method.

A new method has been developed, using, for the same signal, a system of several variances. These variances have been chosen in order to get a set of N (N = 3, 4 or 5) variances which present different types of selectivity (high-pass variances, band-pass variances, ...). Knowing the response of each variance of this set for each type of noise, it is possible, by solving a N equations system, to measure the contribution of N types of noises.

Different types of signals have been studied by this method and by a one-variance method. The results of these two methods will be compared and an estimation of precision, dynamics, and computing time will be given.

1) PRINCIPLE

11. Single-variance method

Generally, a variance exhibits different slopes for different types of noises. Thus, it is possible to estimate the contribution of all these noises with the asymptotes to the variance curve. However, some variances present the same slope for different types of noises (e. g. : Allan variance for white and flicker phase noises). Furthermore, the sensitivity of a variance is not the same for the different types of noises : there are high-pass variances, low-pass variances ... Therefore, it is useful to know what type of noise is the most representative in a signal in order to know what sort of variance will produce the greatest accuracy for this noise.

12. Multi-variances method

An improvement of this first method is to use several different variances on the same signal. Thus, it is possible to choose variances which present different kinds of selectivity for the different types of noises. Moreover, if we use N variances to measure the contribution of N types of noises, the resolution of the associated NxN system will generate, for each value of τ, the measure of the contribution of the N types of noises. The accuracy of these measures can be increased by averaging along τ values.

As an exemple, let us consider the 3x3 system using Allan variance $\sigma_y^2(\tau)$, High-pass variance $\sigma_{HP}^2(\tau)$ and Band-pass variance $\sigma_{BP}^2(\tau)$ to retrieve the contributions h_0 of white frequency noise, h_{-1} of flicker frequency noise and h_{-2} of random walk frequency noise. The system (1)(2)(3) below, shows the response of each variance for each type of noise. Of course, more the responses of these variances for one type of noise are different more the measure of the contribution of this type of noise is accurate. Similarly, in the case of low-frequency noise (random walk and flicker frequency noise) the greater values of τ produce the best precision and vice versa.

$$\begin{cases} \sigma_y^2(\tau) = \dfrac{h_0}{2\tau} + 2h_{-1}\mathrm{Log}2 + \dfrac{2\pi^2 h_{-2}\tau}{3} & (1) \\ \sigma_{HP}^2(\tau) = \dfrac{h_0}{\sqrt{2}\,\tau} + \dfrac{\pi h_{-1}}{2} + \dfrac{\pi^2 h_{-2}\tau}{\sqrt{2}} & (2) \\ \sigma_{BP}^2(\tau) = \dfrac{\sqrt{2}\,h_0}{\pi\tau} + \dfrac{5\,h_{-1}}{\pi\sqrt{2}} + \dfrac{8\sqrt{2}\,h_{-2}\tau}{\pi} & (3) \end{cases}$$

The determinant of this system is :

$$\Delta = 2\sqrt{2} - \frac{5\pi}{4} + 2\pi\mathrm{Log}2 - \frac{16\mathrm{Log}2}{\pi} + \frac{5\pi}{3} - \frac{\sqrt{2}\,\pi^2}{3} \qquad (4)$$

The resolution of this system allows us to calculate the coefficients h_0, h_{-1} and h_{-2} :

$$\begin{cases} h_0 = \dfrac{4\sqrt{2} - \frac{5\pi}{2}}{\Delta}\tau\sigma_y^2(\tau) + \dfrac{\frac{5\sqrt{2}\pi}{3} - \frac{16\sqrt{2}\mathrm{Log}2}{\pi}}{\Delta}\tau\sigma_{HP}^2(\tau) + \dfrac{\sqrt{2}\pi^2\mathrm{Log}2 - \frac{\pi^2}{3}}{\Delta}\tau\sigma_{BP}^2(\tau) & (5) \\ h_{-1} = \dfrac{\pi - \frac{8}{\pi}}{\Delta}\sigma_y^2(\tau) + \dfrac{\frac{4\sqrt{2}}{\pi} - \frac{2\sqrt{2}\pi}{3}}{\Delta}\sigma_{HP}^2(\tau) + \dfrac{\frac{\sqrt{2}\pi^2}{3} - \frac{\pi^2}{2\sqrt{2}}}{\Delta}\sigma_{BP}^2(\tau) & (6) \\ h_{-2} = \dfrac{\frac{5}{2\pi} - \frac{1}{\sqrt{2}}}{\Delta}\dfrac{\sigma_y^2(\tau)}{\tau} + \dfrac{\frac{2\sqrt{2}\mathrm{Log}2}{\pi} - \frac{2}{2\sqrt{2}\pi}}{\Delta}\dfrac{\sigma_{HP}^2(\tau)}{\tau} + \dfrac{\frac{\pi}{4} - \sqrt{2}\mathrm{Log}2}{\Delta}\dfrac{\sigma_{BP}^2(\tau)}{\tau} & (7) \end{cases}$$

2) COMPARISON BETWEEN THESE TWO METHODS

The table below compares the results obtained with the single-variance method and with the multi-variances method (N = 3 as above). These values have been obtained with a large number of numerical simulations (all the simulated sequences contain 8192 samples). The advantage to work with simulation is, of course, the exact knowledge of the values of noise coefficients.

Thus, the item precision shows the difference between this real value and the measured value. This precision was obtained for 3 types of noises with the same magnitude in the frequency domain corresponding to the samples. It is clear that, if a type of noise is smaller than the others, the precision on the measure of this noise will be worse.

Dynamics shows the range in which one type of noise can vary while the precision on all the types of noises remains better than 20 %. In fact, it seems that this quantity depends on the varying type of noise. With this set of variances we found that for the flicker frequency noise, dynamics is about 40 dB, instead of 30 dB for the other types of noises.

Computing time is given for a PC/AT (80286, 16 MHz). In the multi-variances method with N different variances, the computing time is about N times the computing time for the single-variance method, because the time to solve the system is very short in comparison to the time to compute variances.

	Single-variance	Multi-variances
Precision	20 %	< 5 %
Dynamics	10 dB	30 dB
Computing time	1 minute	3 minutes

Posters

PO-02 Short presentation of each poster by the chairman

Chairman: *S. Starker*

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 1

RUSSIAN RADIONAVIGATION SYSTEM AS A TIME COMPARISON METHOD

Tapio Mansten (1), Kalevi Kalliomäki (2)
(1) Technical Research Centre of Finland, Metrology Research Institute,
Otakaari 7 B, SF-02150 Espoo.
(2) University of Oulu, Department of Electrical Engineering, SF-90570 Oulu.

1. Introduction

The Metrology Research Institute (MTI) contains, amongst others, the National Measurement Centre of Time and Frequency in Finland. Its time and frequency is based on a GPS satellite receiver and two rubidium oscillators. MTI time and frequency are traceable to BIPM via atomic clock transports with SU and STA and via time reports of BIBM and USNO. We also monitor the Norvegian Sea Loran-C chain (Sylt) as a back-up and compare the results with those of Sweden. The clocks in our time laboratory are in computer surveillance. The computer collects data of the time differences and laboratory environment. Results are taken at 10 min intervals and cleaned statistically to 2 hour averages, so we get 12 measured points per day.

Soviet Union has a radionavigation system, RNS, that closely resembles the Loran-C system. The master station situates at Brjansk, about 300 km from Moscow and about 1000 km from our laboratory. VNIIFTRI (a Soviet metrology institute) near Moscow monitors the phase of the master station every day at 9.45 UTC and publishes the results in time information letters.

Signals from RNS have been used in Finland for timing purposes since 1972 due to strong signals and suitable repetition interval (80 000 us). In April 1988 one of our Loran-C receivers was locked to the RNS master station and connected to our automatic measurement system. Since the we have measured the phase and the amplitude of the Brjansk transmission.

The National Measurement Centre for time and frequency in Sweden (STA) monitors the same transmitter with us in an agreed Euromet project. In the future Austria (TUG) might be a third monitoring party, too. Therefore the comparison of "Eastern" and "Western" time and study of propagation conditions over flat land has a common interest.

The approximate distances of transmitters and receivers concerned are represented in table I.

TABLE I: Distances and approximate propagation delays between transmitters and receivers concerned.

Stations	Distance, km	Propagation delay, microseconds
Brjansk - STA	1201	4010
Brjansk - MTI	969	3236
Brjansk - VNIIFTRI	362	1209

2. The receiver

Our RNS receiver as well as our Loran-C receivers have been built in our laboratory. The basic construction is suitable both for Loran-C and RNS reception. Because of relatively weak signals we have built shielded loop antennas with 46 dB preamplifiers on our roof. We do not use AGC in our receivers because of fixed receiver location. The receiver measures RNS phase relative to our time base by using vector volt meter principle. The phase integration time constant is a few minutes.

3. Measurements

We have collected continuous data of the phase of Brjansk RNS transmissions, first from April 1988 to the end of May 1988. In June 1988 the transmissions from Brjansk became extremely irregular because of the RNS network reorganization and proper measurements could not be resumed until late October 1988. After that the data collection has been continuous with only a few short pauses.

In this paper we have included the measurements from periods 9th of April - 29th of May 1988 (period 1, STA, VNIIFTRI, MTI), 1st of March - 31st of May 1989 (period 2, VNIIFTRI, MTI) and 25th of November - 31st of December 1989 (period 3, STA, MTI). During the first two periods our time was maintained by Loran-C receivers only. During the third period we have traced our time to USNO MC via our GPS satellite receiver.

4. Results

In figures 1, 2 and 3 are represented the relative results of the participated laboratories traced back to

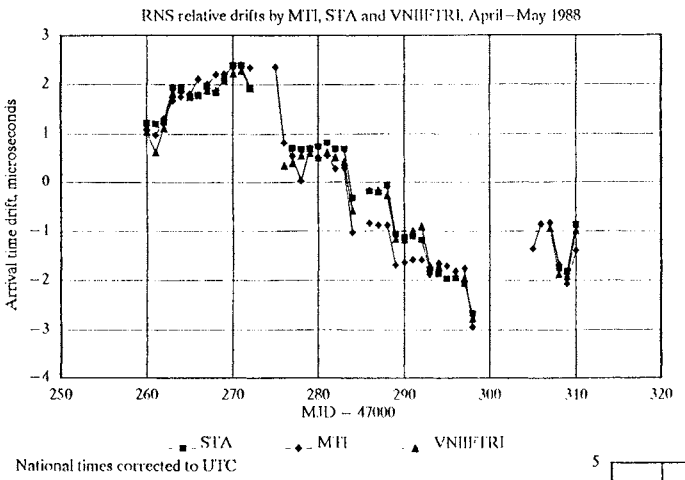


Figure 1: First period of RNS-measurement
9. Apr. - 29. May 1988

Average values:

MTI:	3371.8 us
STA:	4181.1 us
VNIIFTRI:	1291.4 us

Difference SD:s:

MTI-STA:	$360/\sqrt{2}$ ns
MTI-VNIIFTRI:	$400/\sqrt{2}$ ns
STA-VNIIFTRI:	$150/\sqrt{2}$ ns

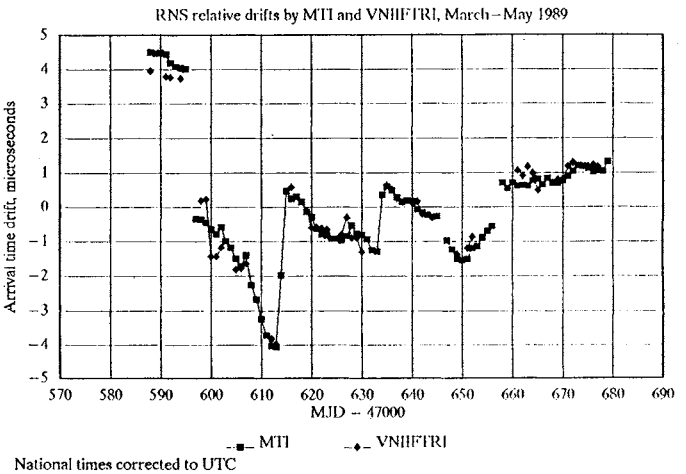


Figure 2: Second period of RNS-measurement
1. March - 31. May 1989

Average values:

MTI:	3373.2 us
VNIIFTRI:	1205.5 us

Difference SD:

MTI-VNIIFTRI:	$335/\sqrt{2}$ ns

UTC by using the appropriate time reports from BIPM, USNO and VNIIFTRI. Because the results contain some unknown constants due to uncertainty in propagation and receiver delays we have subtracted averages from them. Some obvious +/- 10 us receiver jumps have also been corrected.

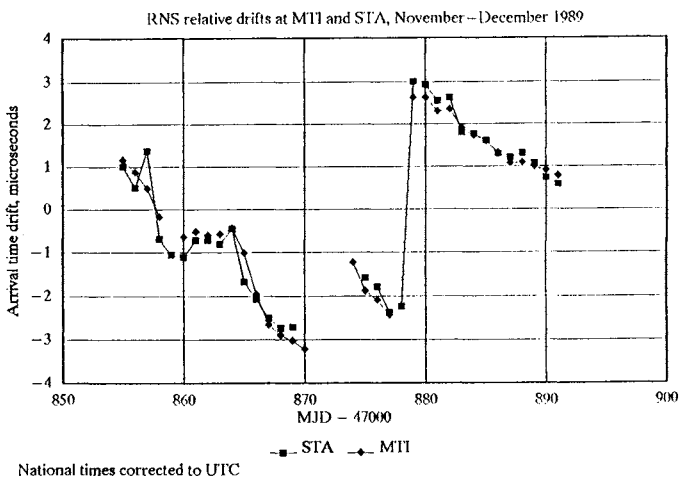
Figure 3: Third period of RNS-measurement
25. Nov. - 31. Dec. 1989

Average values:

MTI:	3371.2 us
STA:	4011.4 us

Difference SD:

MTI-STA:	$270/\sqrt{2}$ ns



5. Conclusions

The arrival time of RNS pulses have been measured and corrected for our time base error versus UTC. The parallel measurements at VNIIFTRI and STA, if available, have been traced back to UTC by using appropriate time reports. The results show a systematic daily amplitude variation of +/-20% and phase variation of +/-200 ns. The drift rate of the phase due to Brjansk master clock is typically about 200 ns/day. The Brjansk main clock seems to be kept within +/- 5 us limits compared to UTC by using jump corrections when needed. The average arrival times of VNIIFTRI and STA have changed during the measurement period due to new receivers (VNIIFTRI), new measurement site (STA) and some possible changes in computational constants.

RNS phase measurement gives us a fair continuous comparison possibility between the national times of Finland, Sweden and Soviet Union.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 2

The Advanced Mitrex Modem with Small Satellite Earth Station

Ph. Hartl, K-H. Thiel
University Stuttgart
Institute of Navigation
Keplerstrasse 11
D-7000 Stuttgart

H. Reder, F. Schmidt
Dornier GmbH
P.O.B. 1420

D-7990 Friedrichshafen

The synchronization via telecommunication satellites and or other geosynchronous satellites can be performed by means of the Mitrex-Modem with an accuracy of the order of a nanosecond. This has been proved by various tests with the Mitrex-Modem (1 - 4). In the original Mitrex-Modem mainly analog circuits have been used.The advanced Modem applies a digital design. It allows much more flexibility for the type of operation, is PC- compatible and contains already the time interval counter as interegral part of the Modem. The basic operational concept is very similar to the original Modem; fig. 1. The advantage of representing the interface between the ground station´s IF- input/output and the clock-standard is remained. It can, therefore, be used in connection with the various existing ground stations. However, it should be mentioned that the outmost accuracy can be achieved, if the time comparison is performed between identical ground stations.

In addiiton, it is preferable, to have very compact ground stations, which would minimize the internal signal delays and the uncertainties involved. If such compact stations are used and if they are mobile, then one could also determine and calibrate the individual stations by comparison at the same location. So, for calibration pusposes, one would ship a "reference station" from one ground station of the network to another and operate for a short time both systems first as station 1 and 2 (colocation test) and afterwards in parallel with some remote station.

As an example for such a ground station, the HPST - ground station of Dornier is described in table 1 in its performance characteristics. This "Highly PORTABLE GROUND STATION " can be adjusted to the various applications as regrads the transmitter power amplifier and the additional transmission of data.

For easy handling and for providing this accuracy in conjunction with the ground station, a small and compact Satellite Earth Station is foreseen to establish the link via the satellite. Using the modem with such a small station will provide an overall higher performance due to short and well known connections, thus allowing for removing the station path delays.

With this improvement not only the time synchronization will be improved, but also the range measurement can be performed in a more flexible way at different positions. This will provide the possibility of controlling a satellites position and could easily be realized with three Mitrex-Stations as shown in fig. 2.

Literature:

1) Ph. Hart et al. : Spread spectrum time transfer experiment via Intelsat, Proceedings of the Fifteenth Annual Precise Time and Time Interval Applications and Planning Meeting, December 6-8, 1983

2) Ph. Hartl et al.: Tracking experiments with OTS using the MITREX modem, Acta Astronautica Vol. 12 pp. 629-639

3) David A. Howe : High accurate time transfer via geostationary satellites. Preliminary results, IEEE Transactions on Instrumentation and Measurement, Vol. 37, September 1988

4) D. Kirchner et al.: A two-way time transfer experiment via ECS-1 using the MITREX modem, IEEE Transactions on Instrumentation and Measurement, Vol. 37, September 1988

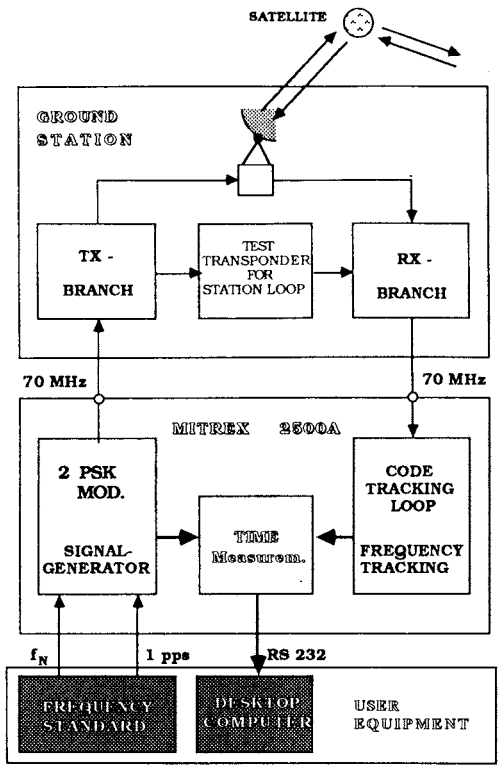


Fig. 1: GROUND STATION WITH MITREX MODEM

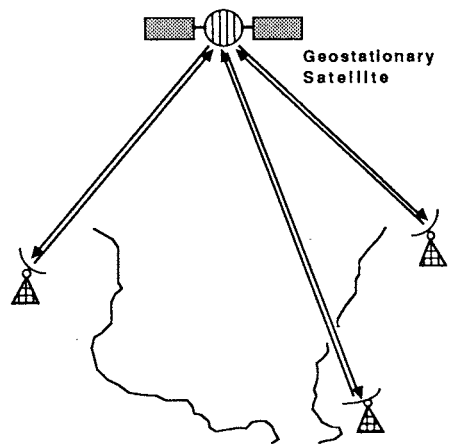


Figure 2: Satellite position monitoring with three Mitrex stations

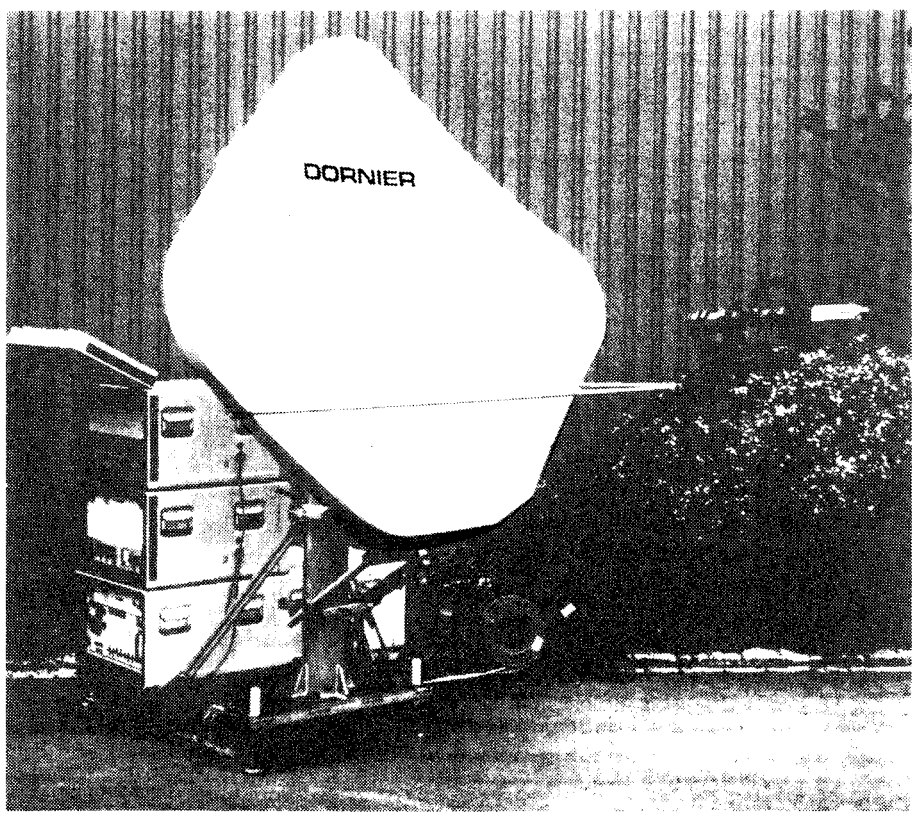


Figure 3: High Portable Satellite Terminal (HPST)

SYSTEM

The satellite earth station type MIE is designed to comply with all relevant communication satellite systems.

G/T (Figure of Merit)	22 dB/K (10.95 GHz)
EIRP (Standard)	45 dBW
EIRP (Option HP)	48 dBW

ANTENNA

Type	Dual-offset
Equiv. Diameter	1.8 m
Sidelobe Performance	Complies with CCIR 580-1 across the entire transmit and receive band
Cross polar isolation	35 dB within the main beam down to the -1 dB copolar contour
Feed	Two port Rx/Tx 10.95 GHz - 11.7 GHz and 14.0 - 14.5 GHz
Transmit Gain	46.5 dB: at 14.25 GHz
Receive Gain	44.9 dB: at 11.70 GHz
Noise Temp.	36 K at 20° elevation
Antenna Mount	El/Az pedestal

SUBSYSTEMS RX/TX

Transmit

Output Frequency Range	14.0 - 14.5 GHz, synthesized, 1 MHz step size (others upon request)
IF	70 MHz
IF Bandwidth	36 MHz (standard), others upon request
Transmit Power	1 W (standard), 2 W (option HP)

Receive

Input Frequency Range	10.95 GHz - 11.7 GHz, synthesized, 1 MHz step size (others upon request)
Antenna Electronics	Low Noise Block Converter (LNC) for data applications
Noise Temp. LNC	140 K max.
1st IF (ext.)	950 MHz - 1700 MHz
2nd IF (ext.)	70 MHz
Bandwidth (2nd IF)	36 MHz (standard)

MECHANICAL DESIGN

Antenna Pedestal	Several types of hot dip galvanized supporting systems are available for • Roof mounting • Flat roof mounting • Ground installation with ballast
Housing of Electronic Equipment	The broadband up and down converter (e.g. 1/14 GHz and 11/1 GHz) are mounted near the feed system of the antenna. A weatherproof housing with internal 19" support frame is installed behind the antenna pedestal.

Dornier reserves the right to make changes to the products described in this document to improve performance, reliability or manufacturability at any time without notice.

Table 1: Specification of Satellite Earth Station for MITREX Applications

Up-link (14.2 GHz)				
EIRP (0.5 W transmit power)	dBW	42.00		42.00
Clear sky atmospheric attenuation	dB		0.25	
Free space loss	dB	207.90		207.90
Satellite beam centre G/T	dB/K	9.00		6.50
Clear sky up-link C/N_0	dBHz	71.45		68.95
Satellite				
IPFD for saturation	dBW/m^2	-79.10		-77.60
Spreading factor	dB		163.05	
IPFD resulting from carrier	dBW/m^2	-121.30		-121.30
Carrier input back-off	dB	42.20		43.70
Carrier output back-off	dB	37.20		38.70
Down-Link (11.1 GHz)				
Beam centre satellite EIRP at saturation	dBW	50.00		46.50
Carrier output back-off	dB	37.20		38.70
Carrier EIRP towards earth station	dBW	12.80		7.80
Clear sky atmospheric attenuation	dB		0.20	
Free space loss	dB	205.90		205.90
Earth station G/T	dB/K		22.00	
Clear sky down-link C/N_0	dBHz	57.30		52.30
Clear sky overall C/N_0	dBHz	57.20		52.20

Table 2: Link Budget for two way time transfer communication using Intelsat V West spot / East spot transponder

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 3

PRARE ON GPS AND/OR GLONASS

PH. HARTL, W.SCHÄFER
University Stuttgart
Institute of Navigation
D-7000 Stuttgart 1
F.R.Germany

CH. REIGBER, W. LECHNER
Deutsches Geodätisches Forschungsinstitut
Abteilung 1
D-8000 München 1
F.R.GERMANY

ABSTRACT

The Precise Range and Range Rate Tracking Equipment PRARE, which will be flown on the ESA remote sensing satellite ERS-1 end of 1990 for the first time, is a selfcontained system, which could also be used for other spacecrafts. If installed in a spacecraft of the Global Positioning System GPS and/ or on board a GLONASS satellite it could serve for a detailed error analysis and for collection of data for a better modelling of the orbit parameters, the propagation phenomena, and the on board clock of these two navigation systems. The basic proposal is described. Simulation results are presented.

INTRODUCTION

PRARE is a two- way tracking system , which can offer extremely precise ranging data. In addition, it allows data transmission between the spacecraft and the PRARE groundstations. About 30 groundstations are distributed around the globe. The PRARE system is fully autonomous. It will be flown for the first time on board the ESA- satellite ERS-1 and of 1990.

Both, the on board equipment as well as the ground stations, are very compact. The ground stations are mobile, the ERS-1 PRARE on board equipment consumes very little power (27 W during transmission, 9 W in standby) and is small in volume(40 cm x 24 cm x 18.8 cm) and mass (17 kg). The corresponding values for GPS or GLONASS would be 20 kg and 40 W operational /10 W in stand-by.

It should be easily possible from a technical point of view to install the space segment on board of the GPS- and/or GLONASS spacecraft. The instrument would hardly be modified as compared to the ERS-1 flight unit. The main change would be the on board antenna, which would be very similar to the antenna used at the PRARE ground stations. The link budget would then be almost identical for both cases.

With the PRARE system the navigational satellite systems GPS and/or GLONASS could be investigated in details as regards their error sources. One could perform a detailed error analysis, which would allow improved modelling of these navigation systems. It is more a political question than a technical one, if such a proposal will be accepted by the owners of the two systems.

This paper will shortly describe the possible measurements. An estimation is given about the accuracy improvements. These results are based on simulation results.

THE INTEGRATED SYSTEM

Fig. 1 shows the block diagram of the integrated system. The PRARE spaceborne equipment would be installed on board the GPS or GLONASS satellite. It would be recommended, but not unconditionally needed that the signals of the GPS/GLONASS on board clocks can be used as master oscillators for the PRARE space segment. Apart from this the PRARE interface could be very simple: PRARE would require the unregulated, raw electric power from the spacecraft. All facilities for telemetry, telecommand, data storage, programmed tracking procedures etc. are contained in the PRARE equipment itself. Also the uploading and dumping of the data could be performed via one or two master stations of the PRARE network. One of the master stations is installed for the ERS-1 operation in Oberpfaffenhofen, Germany. The command station and another dumping station are located at Stuttgart, Germany and Tromsoe, Norway, resp. It would be possible to apply anyone of the ground stations of the PRARE network to become a type of master station, if this seems to be appropriate for the proposed mission. The PRARE ERS-1 tracking network with 20 fixed stations (others could be added or selected at different places, or could be used as mobile stations) is shown in fig.2. The ground stations are small and compact.

It is proposed that at the location of each ground station at least one L1 plus L2- frequency GPS/ receiver (or GLONASS receiver) is installed. This means that the PRARE and the navigational satellite signals could be received simultaneously on ground. The main differences between the GPS/ GLONASS- systems on the one side and the PRARE system on the other side are the following:

- PRARE applies a two- way system in the 7/8 GHz band and a one way down-link in the S- band. This S- band down link allows to determine the total electron content TEC very precisely by comparison of the S- band delay with the X-band delay. This measurement is much more precise than in the GPS/ GLONASS- case, where 1,2 GHz and 1,5 GHz signal delays are applied, which are close together, as far as the frequency distance is concerned. Besides, by comparing the delays in the 4 frequencies mentioned it should be possible to determine the ionospheric effects to such a degree that the residue uncertainty will be below 1 mm. So, one could determine very precisely the ionospheric effects as func-

tion of time, geography, solar activity etc. and use it later for a better ionospheric prediction and correction.

- The PRARE ground station consists of two receivers- one for the S- band and one for the X- band. The X- band ranging data as well as some "additional data " modulate the X- band transmitter, which sends the signals up to the spacecraft. On board, the modulated signals are detected, the time delay between transmission and reception of the X- band ranging code is counted and used as a measure for the distance. The Doppler shift of the transmitted and the received X- band signal is used to measure the relative velocity and the phase information. Calibration procedures are performed on board as well as on ground to compensate for the internal time delays in the instruments. At the X-band uplink, which is primarily used for the two-way ranging signal and the PRARE data measured on ground (signal delay difference S-band vs. X-band, meteorological data, and calibration data) one could transmit also the GPS/GLONASS information. All this information and the PRARE two- way signal delay can be stored in the PRARE memory aboard.

- So, on board, all "raw tracking data", the time delay between S- and X- band down link, the corresponding time delays of the GPS/ GLONASS signals, and local meteorological information, such as pressure, temperature, moisture is stored in the PRARE memory for each PRARE ground station.

- During overflight the global information from the various ground stations is sent to the master station.

THE COLLECTED INFORMATION

The information collected at the master station can be immediately analyzed. It can be used for the following aspects:

- Determination of the GPS / GLONASS S/C - clocks´drifts: This is possible, because the two- way ranging system PRARE allows to determine the on-board time more precisely due to the fact that only the uncertainties of the two- way propagation time and the uncertainty of the slant range cause uncertainties. The short term stability of the PRARE on board clock is in the order of 10 exp -13, the uncertainty in slant range in the order of some centimeters. The total uncertainty may be on the order of nanoseconds and are related to the atomic clocks on ground.

- Orbit uncertainty: The ionospheric effect is compensated. The uncertainties about the gravity field can be eliminated, if 3 to 4 ground stations measure the orbit simultaneously and determine by this way the position by geometry and not by dynamic modelling of the orbit and the forces effective on the spacecraft.

- Tropospheric effects: The ranging data of the PRARE system contain information about the surface temperature, moisture and pressure at the ground station. The dry component of the troposphere can, therefore, be corrected by means of the well known models for the standard atmosphere. The wet component can also be modelled. There are various models in use, such as the Hopfield model and others. We have developed a new model, which allows to improve the wet component by use of the met data of the ground station in connection with the meteorological global map of the day of measurement. The accuracy is then in the order of about 2 to 4 cm or better for the zenith. It is important that in the new model we can also determine the reliability of the correction.

- Multipath effect: The PRARE ground station applies a directional antenna. It is, therefore, less sensitive to multipath than the omnidirectional antenna of the navigational satellite receivers. A comparison of the measurement data can be used for the analysis of the multipath effects and the physical background.

- It would also be possible to determine from the carrier phase measurements unambiguously the range data of GPS / GLONASS, because of the extremely precise information of the PRARE ranging techniques.

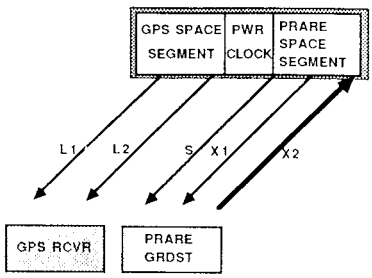
SIMULATION RESULTS

Under conservative assumptions of 20 PRARE ground stations the orbit determination we achieved the following simulation results for absolute positioning, relative positioning, orbit determination, orbit prediction, and pole coordinate determination:

REFERENCES

Hopfield, H.S. (1979) : " Improvements in the Tropospheric Refraction Correction for Range Measurements". Phil. Trans. R. Soc., London, A. 294, 341-352

Rahnemoon, M. : "Ein neues Korrekturmodell fürr Mikrowellen - Entfernungsmessungen zu Satelliten". DGK, München, 1988



X1 = 8.45-8.50 GHZ; X2 = 7.19-7.235 GHZ; S = 2.20-2.29 GHZ

10 MHZ 10 MHZ 1 MHZ

PRARE GROUNDSTATION: 60 CM DISH
TRIPOD
BOX FOR
REGENERATIVE TRANSPONDER AND
DATA XMISSION

FIGURE 1: CONFIGURATION OF PRARE/GPS

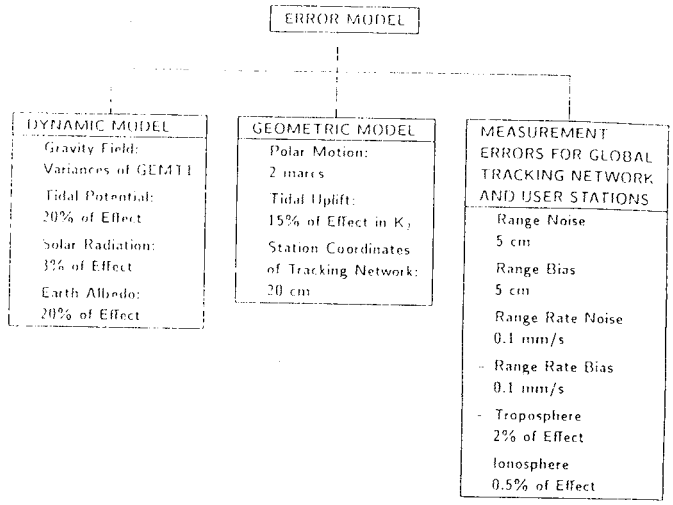
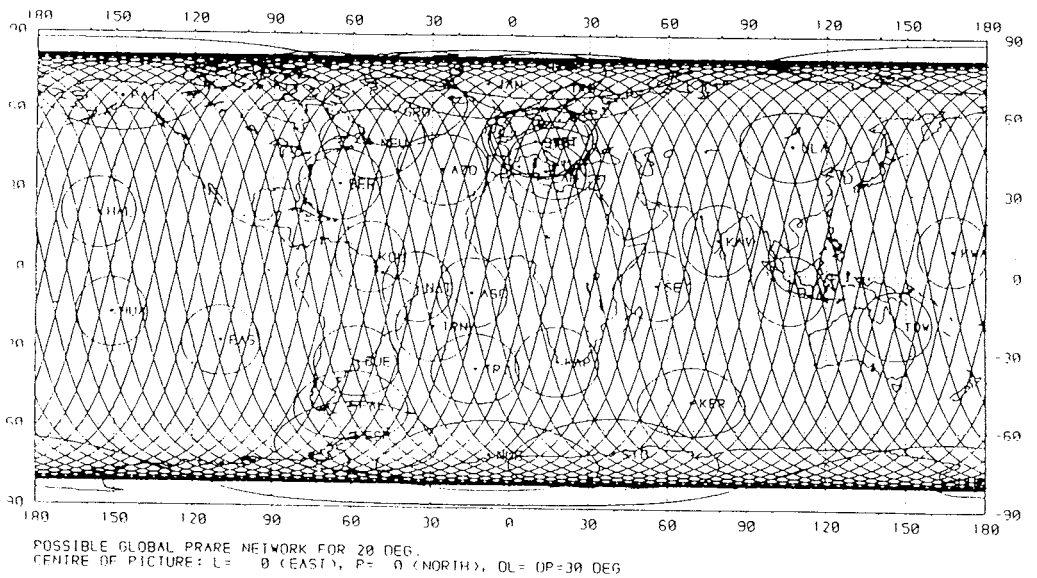


Table 1 : Basic assumption for the model



POSSIBLE GLOBAL PRARE NETWORK FOR 20 DEG.
CENTRE OF PICTURE: L= 0 (EAST), P= 0 (NORTH), OL= OP=30 DEG

Fig. 2

		1 day arc	7 days arc
▪ **Orbit determination:**	Radial:	< 0.1 m	< 0.1 m
	Position:	< 0.5 m	< 0.5 m
▪ **Orbit prediction:**			
1 day	Radial:	< 0.1 m	< 0.1 m
	Position:	< 1.5 m	< 1.0 m
7 days	Radial:	< 0.1 m	< 0.1 m
	Position:	< 7.5 m	< 4.0 m
▪ **Pole coordinates:**		< 3 marcs	< 2 marcs
▪ **Dominating Error:**		coordinates of tracking stations	gravity field

	1 day arc	7 day arc
▪ **Absolute Positioning:**		
- Wettzell (W)	8.4 cm	7.6 cm
- Oberpfaffenhofen (O)	8.4 cm	7.6 cm
- Kootwijk (K)	8.1 cm	7.3 cm
- Greenbelt (G)	11.1 cm	9.2 cm
- **Dominating Error**	coordinates of tracking stations	coordinates of tracking stations
▪ **Relative Positioning**		
- W - O (ca. 165 km):	1.5 cm ($9.1 \cdot 10^{-8}$)	0.7 cm ($4.2 \cdot 10^{-8}$)
- W - K (ca. 600 km):	1.3 cm ($2.2 \cdot 10^{-8}$)	0.9 cm ($1{,}5 \cdot 10^{-8}$)
- W - G (ca. 6500 km):	13.7 cm ($2.1 \cdot 10^{-8}$)	9.0 cm ($1.4 \cdot 10^{-8}$)
- **Dominating Error:**	range bias, troposphere	range bias troposphere

Table 2 : Simulation results for GPS/PRARE

Session PO 02 Number 4

SLALOM - Intersatellite ranging and Doppler measurement

Ph. Hartl and Ch. Reigber
Institut für Navigation
Universität Stuttgart
Keplerstrasse 11
D - 7000 Stuttgart 1

For the precise gravity field measurement one can use small transponder satellites as gravity field probes. The required sensitivity must be in the order of 10 μm/sec. The concept, which is presently under study, is presented.

Session PO-02 Number 5

HIGH ACCURACY TIME MANAGEMENT STUDY (HAT STUDY)

Author: D. MAEUSLI, European Space Agency, Onboard Information Management Section, Postbus 299, 2200 AG Noordwijk (NL), fax 01719-17400
Electrical Engineer, Graduate of EPF Lausanne
Co-authors: Dr. U. DENSKAT and T. SCHWAMMLEIN, Dornier, Friedrichshafen (Germany)
M. LOMBARDI and J.M. OBERTO, Matra, Toulouse (France)

Abstract: On board spacecraft, time information is required to time tag events (such as: solar eruption) and to start actions at precise instants (such as: engine ignition). During the HAT study (performed by ESA-Dornier-Matra), a system study has been made to evaluate the operation and equipment requirements to maintain the different types of on board time reference. The required use of existing medias, ground stations and worldwide time formats imposed strong constraints. Some simple protocols to distribute an high accurate time information to users connected to one of the most common on board bus (OBDH, MIL-1553B) were defined and simulated.

1. THE DIFFERENT TYPES OF ON BOARD TIME REFERENCES AND THE CCSDS FORMATS

The **Elapsed Time reference (ET)** is the simpler on board time reference, generated by an unsegmented binary counter in free-running mode: the epoch is arbitrary. No synchronisation is performed between ground and space segments (i.e no need of TC uplink). The basic time unit is the second. ET has a good short term stability as no correction is applied on the counter. The ET monotony is guaranteed implicitly. The drawback is that a correlation is required on ground before the use of these data (off-line process).
The **Universal Time reference (UT)** has a fixed epoch (1.1.1958) and leap second corrections must be performed when required. The definition of the required synchronisation process shall be such as to guarantee a long term stability and the UT monotony. UT is segmented in day/msec/usec/(nanosec).
The Consultative Committee for Space Data Systems has defined two formats for ET and UT. The HAT study uses them and consequently shall be compatible with the ground equipments of most of the world space agencies. However the HAT study as shown that these CCSDS formats shall be eventually modified to reduce the on board time hardware: the same basic time unit shall be selected for the 2 formats (e.g. 1 msec.). In the HAT study, **ET is always implemented in the spacecraft (S/C)**. Following the mission/user requirements (on-line process, data broadcasting), UT can be added: in this case, ET is used for UT synchronisation.

2. THE TIME DISTRIBUTION CHAIN

At the **Ground Station (GS)** node is located UT(GS), the UT reference of the whole chain; instead of ET, a correlation table is available to convert ET in UT and vice versa.
The **Central Time Management System (CTMS)** node is located onboard the S/C in the Bus Controller Terminal: the CTMS is the time reference for all the other nodes located onboard. The ETC(CTMS) is the ET reference of the whole chain. Optional is UT(CTMS) which is a copy of UT(GS).
The **Local Time Management System (LTMS)** nodes regenerate locally a copy of the CTMS time information: ET(LTMS) is mandatory while UT(LTMS) availability depends on UT(CTMS) availability and on User requirements.
The **Telemetry (TM)** link is required for correlation and monitoring (in accordance to ESA Packet Telemetry Standard).
The **On Board Data Bus** link is used to keep LTMS nodes in synchronism with CTMS node.
The **Telecommand (TC)** link is used for synchronisation of UT(C/LTMS) with UT(GS) (in accordance to ESA Packet Telecommand Standard).

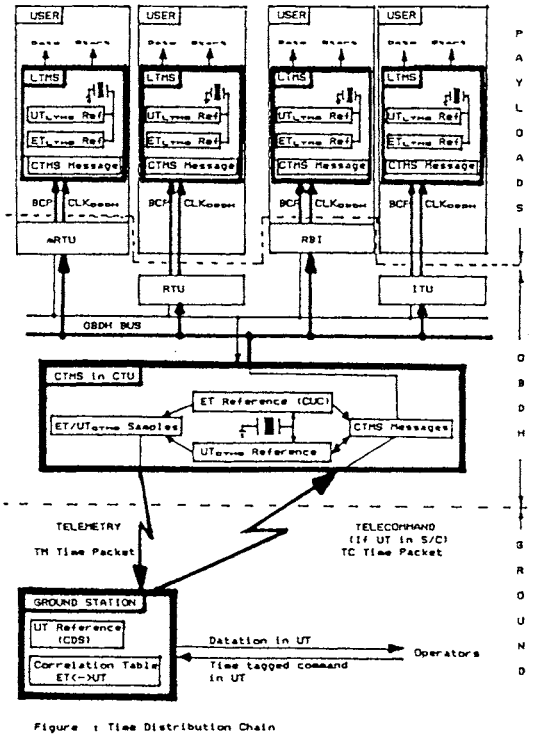


Figure 1 Time Distribution Chain

3. THE ET(CTMS)-UT(GS) CORRELATION

To make the ET(CTMS)-UT(GS) correlation, the GS has permanent access to a very accurate UT(GS) (UT chain reference). To generate the correlation table the TM information is used by the GS. On board the S/C, ET(CTMS) (and UT(CTMS) if available) is sampled at regular occurrence of a particular edge in the TM flow. The GS samples UT(GS) at reception of the same edges (but with transmission delays). The on board ET(CTMS) samples are transmitted to GS in dedicated TM Time Packets. With the knowledge of ET(CTMS) and UT(GS) samples and the S/C position, the GS can generate its correlation table. By interpolations or extrapolations, ET can be converted in UT and vice versa.
As a single oscillator is used to drive ET(CTMS) and UT(CTMS), the UT(CTMS) samples transmitted by TM are used mainly for monitoring. Other error and status

flags can be included in the TM Time Packets.

4. THE GS-C/LTMS SYNCHRONISATION

If UT(CTMS) is used on board the S/C, synchronisation via the TC link has to be performed to keep UT(GS)-UT(CTMS) difference under an admissible limit. An extrapolation in the future from the correlation table is performed by the GS. The extrapolation result, an advanced ET(CTMS) and its corresponding advanced UT(GS), is placed in a dedicated TC Time Packet (see 7 and 8).
At reception of a TC Time Packet, the CTMS stores the advanced ET(CTMS) and UT(GS) values in a register; then the CTMS broadcasts this information to the LTMS's via the Onboard Bus. The LTMS's shall store this information in a register too. Assuming that all the C/LTMS nodes are running synchronously (see 5), an equality between the advanced ET(CTMS) and the current ET(C/LTMS) shall be detected at the same instant and shall be used to load the stored UT(GS) value in the UT(C/LTMS) counters simultaneously.
Such a synchronisation shall not be performed more than once per day.
The advantage of this method is that the accuracy is only decreased by the down-link and extrapolation errors. In any method these errors will occur. However no TC transmission delay errors are introduced here!

5. THE CTMS-LTMS SYNCHRONISATION

To guarantee the synchronism of CTMS and LTMS's, the CTMS generates automatically each ET second a message similar to the GS one. The message includes now an advanced ET(CTMS) and UT(CTMS); it is broadcasted to the LTMS's which store it in a register. To inform the LTMS's when these time codes become valid, the CTMS broadcasts a dedicated synchronisation sequence. At detection of this sequence, the LTMS's load in their ET(LTMS) and UT(LTMS) counters the stored values ET(CTMS) and UT(CTMS).
The LTMS check if the ET(CTMS) fine time is available or not in the message received from the CTMS to differenciate between a GS-C/LTMS (with) and a CTMS-LTMS (without) synchronisation.

6. THE ON BOARD BUSSES AND THE HAT PROTOCOLS

The properties of the different on board busses differ drastically. Two protocols have been defined to comply with the two strongest candidates: the OBDH bus (used in ERS1, HYPPARCOS, SOHO,..) and the MIL-1553B bus (foreseen in ARIANE5, HERMES, COLUMBUS).
The OBDH (Interrogation) Bus has a permanent traffic of 32bit words (500kbit/s). Four bits of each 32 bit word, the Broadcast bits (BCP), are used by all the bus terminals simultaneously. A BCP is equivalent to a 15.5 kbit/s serial link. In the HAT distribution, one BCP is used. This BCP is modulated to transfer the synchronisation messages (ET, UT codes plus the synchronisation sequence). The permanent OBDH clock is used between the resynchronisations. The HAT distribution does not interfere with other bus traffic!
The MIL-1553B bus (500kbit/s) has no permanent traffic and no BCP. The HAT distribution requires to reserve a same HAT subaddress (1 of 32) in each terminal. The synchronisation messages are sent by conventional 1553 message (up to 32 words) broadcasted to every terminal at the specified HAT subaddress. The synchronisation sequence is replaced by the emmission of a broadcasted 1553 synchronisation mode command generating a pulse in each terminal. As the bus clock is not permanent, Digital Phase Lock Loops are used.

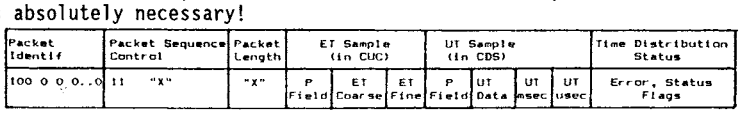
7. THE UT MONOTONY

While ET is implicitly monotonous, the UT synchronisation process must guarantee the UT monotony. For this reason, the resolution of the UT codes handled between GS, CTMS and LTMS is extended. For example, if the mission payloads require a UT resolution of 1 msec., the usec bits are transmitted between the different nodes and are used inside the nodes only. The drift corrections shall be performed only at the level of these additional bits (in our ex.: the usec), i.e the bits non available/visible to the time users.

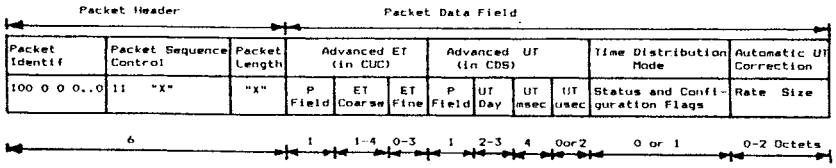
8. THE PROTOCOL SIMULATION

A simulation has proven the validity of the HAT protocols above (with an error management too). The simulation has shown also that a sophisticated oscillator is required to support UT(CTMS). The impacts of the oscillator ageing on the TC rate have been highlighted too (with TC Time Packets as defined in 4). For the SOHO mission (minimum interval between 2 TC Synchro=7 days; accuracy=2*10E-2), an oscillator with a short and long term stability of 5E-10 and 1E-9 has an acceptable weekly drift during 16.28 days only! As a TC increase is often not admissible, the solution is to include some information to correct the permanent drift in the TC Time Packets. However, due to the oscillator and TC requirements, the UT(CTMS) option shall be used only if it is absolutely necessary!

TM TIME PACKET FORMAT:



TC TIME PACKET FORMAT:



9. THE FUTURE

The implementation of these protocols has already started. A single VLSI component supporting the LTMS node (OBDH and MIL-1553B compatible) is designed actually (first prototypes available at mid-90). No modifications are required on the GS. For the CTMS, further investigations have to be performed before starting the implementation; in particular a detailed survey of the oscillators available on the market shall be made. For ESA, the EFTF 90 shall be an opportunity to start this activity.

Session PO-02 Number 6

A METHOD FOR DETERMINING THE POSITION OF THE ANTENNA OF ALLEN OSBORNE TYPE GPS RECEIVERS.

P. Uhrich, A-F. Paradon, R. Tourde.
Laboratoire Primaire du Temps et des Fréquences, Observatoire de Paris.
61, Avenue de l'Observatoire, 75014 PARIS, FRANCE.

INTRODUCTION.

For laboratories comparing atomic clocks with GPS NAVSTAR satellite navigation system, the problem of the coordinates of the GPS receiver antenna is well known. For distances between laboratories under 1000 km, systematic errors that were too large for the usual sources of noise have been noticed on the results of time comparisons [1] [2], and so could only come from the wrong coordinates of the antenna. The Time Section of the Bureau International des Poids et Mesures (BIPM) has carried out a computer program which uses the results of all the time comparisons during one six months schedule to get the error on the relative coordinates between two stations. They obtained precisions of a few tens of centimeters.

In this paper, we present a method, based on the navigation solution of the receiver, for determining the position of the antenna of Allen Osborne type GPS receivers. After having carefully selected four tracks of the BIPM common views time comparisons schedule, we have collected the navigation solutions day by day on the GPS receivers in two stations at the same time. Because the four tracks cannot be followed simultaneously with this kind of receivers, we have had to collect also the clocks drifts, and to introduce them in the daily calculation of the new coordinates. As GPS data are randomly distributed, the mean over the period of experiment gives the corrected coordinates of the antenna.

An example is given with the Paris Observatory and the Besançon Observatory GPS receivers. The use of the corrected relative coordinates on the link between these two receivers allow us to improve the precision of the time comparisons of about 40 %. The precision obtained on the relative positionning of the two antennas is about 1m. That seems to be the best value available by GPS receivers dedicated to time comparisons.

1. THE NAVIGATION SOLUTION.

The selection of the four tracks on the common view time comparison schedule is based on two criterions. The first one is the Geometrical Dilution Of Precision, or GDOP, which is a dimensionless number that caracterises the quality of the navigation solution with one choice of satellites. The GDOP is strongly related to the volume of the fictive tetrahedron formed by the four satellites and the ground station. We used a very fine satellite selection method [3] for choosing the tracks which minimize this GDOP. The theoretical minimum of this number is 1.58. It corresponds to a geometry that cannot be obtained. Only six satellites were available at the time of experiment, and we obtained GDOP which could reach a few hundreds.

The second criterion was the time interval between the first and the last track. Because the four tracks have to be done sequentially on the Allen Osborne type receivers, the influence of the clocks drifts (reference clock and satellites clocks) is not negligible. So we tried to minimize this time interval, including the minimization of the GDOP.

2. INTRODUCTION OF CLOCKS DRIFTS.

The basic system to be solved by the receiver to calculate the navigation solution is:

$$\sqrt{(X - X_i)^2 + (Y - Y_i)^2 + (Z - Z_i)^2} + cT = R_i \quad \text{for } i = 1 \text{ to } 4 \qquad (1)$$

where (X_i, Y_i, Z_i) are the coordinates of the satellites, (X, Y, Z, T) are the coordinates of the antenna and the clock difference between the receiver and the satellite, and R_i are the measurements of the so-called pseudo-range (geometrical range plus the clock drift) between the receiver and each satellite.

Because the time sources are not running at the same frequency, the receiver will measure some pseudo-range biased by the clocks drifts during the sequentially four tracks. We have to write from (1) the following system:

$$R_{Mi} + cT = R_i + cB_i \quad \text{for } i = 1 \text{ to } 4 \qquad (2)$$

where R_{Mi} is the geometrical range, with the coordinates (X_M, Y_M, Z_M) measured by the receiver, R_i the true geometrical range (without clock bias), and B_i contains all the clocks biases between a time t_o taken as an origin, and the time of the track number i. But we want to obtain corrected coordinates (X_C, Y_C, Z_C), given by:

$$R_{Ci} + cT = R_i \quad \text{for } i = 1 \text{ to } 4$$

from where it comes:

$$R_{Ci} = R_{Mi} - cB_i \quad \text{for } i = 1 \text{ to } 4 \qquad (3)$$

We then express the differences between the measured coordinates of the receiver's antenna and the corrected one's:

$$X_C - X_M = \Delta X, \; Y_C - Y_M = \Delta Y, \; Z_C - Z_M = \Delta Z \qquad (4)$$

If we neglect the second order terms (ΔX^2 ...), we finally get from (3) and (4) a linear system given by:

$$\begin{bmatrix} & A_{3,3} & \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = c \begin{bmatrix} B_{2\,1} \\ B_{3\,1} \\ B_{4\,1} \end{bmatrix}$$

where the $A_{i,j}$ only depend on the elevations and the azimuths of the satellites involved, and $B_{i,j}$ means $B_i - B_j$ (the problem of the choice of t_o disappears this way).

3. METHOD FOR DETERMINING NEW ANTENNA COORDINATES.

We have collected every day the coordinates measured with the GPS receiver (X_M, Y_M, Z_M), and calculated from the measured clocks drifts the corrections $(\Delta X, \Delta Y, \Delta Z)$ to be applied to these coordinates. It is then possible to determine the corrected coordinates (X_C, Y_C, Z_C).

At the end of the experimental session, we can get two kinds of results. The absolute corrected coordinates of one receiver's antenna are given by the mean over the set of measurements for each station. The relative positionning of one antenna towards another one is given by the daily difference between coordinates of the two stations, and the mean over the set of data.

4. USE OF CORRECTED COORDINATES FOR IMPROVEMENT OF TIME COMPARISONS.

The improvement on the time comparisons due to the corrected coordinates can be estimated in terms of a scalar product of two vectors [1] [2]. Let us consider two stations A and B, with corrected coordinates only applied to the station B, and let us call B' the calculated new position of the antenna (see figure 1).

During the track with one satellite i, if we call $\vec{u_{bi}}$ the unit vector of the direction station B/satellite i, the correction ∂t due to new coordinates is given by (α a constant):

$$\partial t = \alpha \, \overrightarrow{BB'} \, . \, \vec{u_{bi}} \qquad (5)$$

which can be written [1] [2]: $c\,\partial t = \beta_{Xi}\,\Delta X + \beta_{Yi}\,\Delta Y + \beta_{Zi}\,\Delta Z$

where the β are the director cosines of the direction station B/satellite i, which depend only on the elevation and azimuth of this satellite, seen from the station B. The so calculated ∂t has to be added to the time comparison result for this track. If the satellite is enough stable on its orbit, this ∂t is the same for one particular track during the whole schedule.

A good estimation of the improvement of the precision on the time comparisons is given by the residuals method [1], which can be summarized by: to report all the time comparisons to a standard date on ten days intervals, and to compare them to their mean.

5. APPLICATION TO THE LINK PARIS OBSERVATORY/BESANÇON OBSERVATORY.

DATA:
GDOP Paris (calculated) = 3.08
GDOP Besançon (calculated) = 3.09
Time interval between 1st and 4th track ≈ 4 hours.
Navigation Solutions from 1st may 1989 to 26th june 1989. We only had time to collect two months of daily navigation solutions between Paris Observatory and Besançon Observatory, which after withdraw of the accidental bad results gives us 52 data.
Clocks drifts: UTC(Paris) ≈ 3 ns/day, negligible on four hours.
UTC(Besançon) ≈ 150 ns/day.
SV/GPS Master Clock ≈ 20 ns/day.

RESULTS:
GDOP Paris (measured on GPS receiver) = 3.04
GDOP Besançon (measured on GPS receiver) = 3.00
All the data below are given in meters.

	PARIS (OP):	BESANÇON (OB):	DIFFERENTIAL OP/OB:
Initial coordinates in the receiver:	X = 4202780.73	X = 4314129.51	DX = 111348.78
	Y = 171367.71	Y = 452638.56	DY = 281270.85
	Z = 4778658.82	Z = 4660707.87	DZ = - 117950.95
Corrected coordinates calculated:	X_C = 4202781.56	X_C = 4314134.12	DX_C = 111352.56
	Y_C = 171371.08	Y_C = 452639.94	DY_C = 281268.86
	Z_C = 4778662.53	Z_C = 4660711.96	DZ_C = - 117950.57
Differences and standard deviations:	ΔX_C = 0.83 , σ = 4.8	ΔX_C = 4.61 , σ = 5.3	ΔDX_C = 3.78 , σ = 1.3
	ΔY_C = 3.37 , σ = 3.1	ΔY_C = 1.38 , σ = 3.1	ΔDY_C = - 1.99 , σ = 0.9
	ΔZ_C = 3.71 , σ = 4.2	ΔZ_C = 4.09 , σ = 4.5	ΔDZ_C = 0.38 , σ = 1.6

When one looks at these data, it is obvious that the differences at Paris and at Besançon (let us call them 'absolute differences') between the old coordinates and the new coordinates are far all less than the standard deviations. This can be explained by the variations due to the GPS satellites. The figure 2 shows these variations for one coordinate (here X, but you can get the same kind of figure for all coordinates), from the measured data at Paris and Besançon. It can be seen that, at two different points, distant of 420 km, the measured coordinates follow the same daily variations, which can only be due to the common part of the system of measurement, in other words the daily alteration of the GPS satellites.

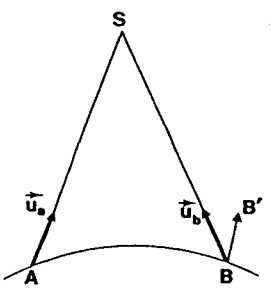


Figure 1. Geometry of the link between two stations.

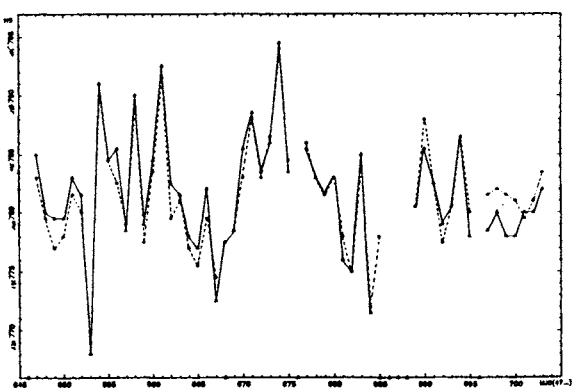


Figure 2. Daily variations of the X coordinates at Paris and Besançon.

In fact, the variations caused by the satellites are randomly distributed, and these differences are not so bad. This can be proved by the changes in the curves of the residuals (see next part) and by the differential differences, which are exactly equals to

the differences between the absolute differences at Paris and Besançon. They are given with standard deviations of about 1 m, because the daily alterations due to the GPS have nearly all been cancelled. The BIPM has obtained on the same link with more than 3000 data the following differential results [4] (all values in meters):

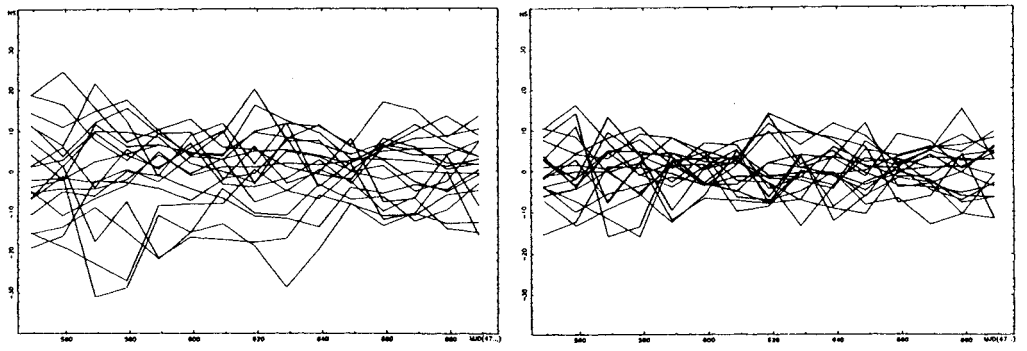
$$\Delta DX_C = 3.23\ ,\ \sigma = 0.22 \qquad \Delta DY_C = -2.43\ ,\ \sigma = 0.08 \qquad \Delta DZ_C = 0.01\ ,\ \sigma = 0.19$$

which differ by not more than 55 cm from our results, and with much better standard deviations.

Anyway, the greatest interest is to use the new relative coordinates to calculate the time correction that has to be added to the result of each track of the time comparisons schedule. This has been done for a long time for each schedule at the BIPM to improve the calculation of the TAI ('Temps Atomique International': international atomic time scale).

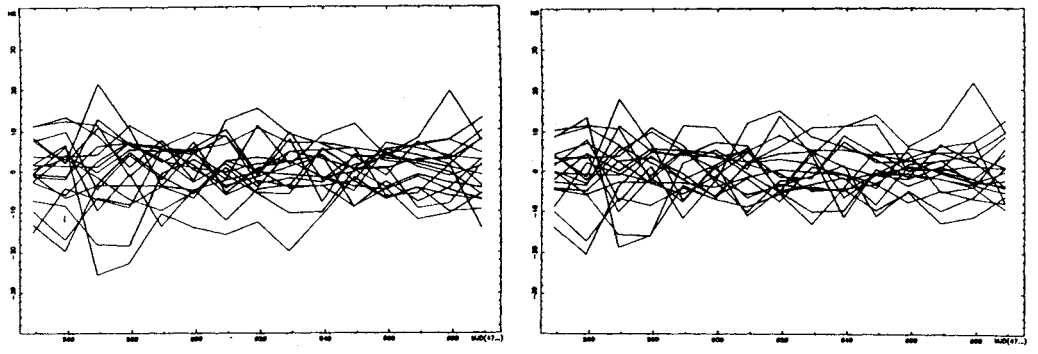
6. IMPROVEMENT ON THE LINK BETWEEN PARIS OBSERVATORY AND BESANÇON OBSERVATORY.

The improvement of the precision on the link Paris Observatory / Besançon Observatory can be seen on the curves of the residuals (figures 3 to 5). The influence of the wrong coordinates of the receiver's antenna is not the same for every tracks: it depends on the direction of the error vector related to the directions of the antenna/satellite vectors (see (5)). The correction of time due to new coordinates makes a global movement of the curve corresponding to each track, towards their general mean.



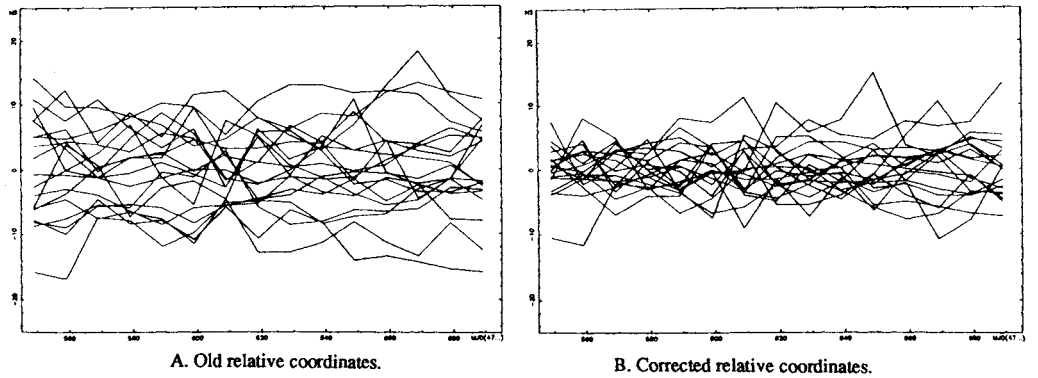
A. Old coordinates. B. Corrected coordinates.

Figure 3. Improvement at Paris using absolute differences.



A. Old coordinates. B. Corrected coordinates.

Figure 4. Improvement at Besançon using absolute differences.



A. Old relative coordinates. B. Corrected relative coordinates.

Figure 5. Improvement on the link Paris/Besançon using differential differences.

The gain on the precision of the time comparisons can be estimated by the changing of the standard deviations of the residuals, calculated day by day. The mean over the period of measurement of these standard deviations goes from 9.4 ns to 6.2 ns after correction at Paris (figure 3). The precision has then been improved of 34 %. At Besançon, the mean of the standard deviations goes from 7.1 ns to 6.6 ns. The precision of the time comparisons has only been improved of 7 %. It is a very low improvement, which can be explained by the geometry of the error vector. As it can be seen on the figure 6, the error vectors at Paris and Besançon are nearly orthogonal.

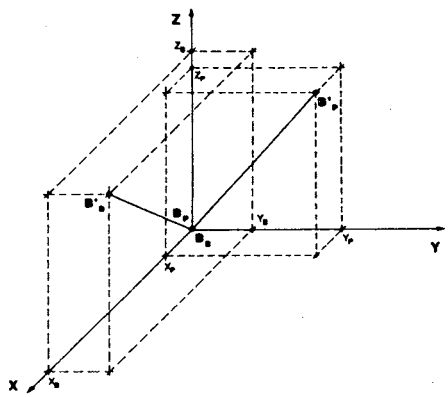


Figure 6. Error vectors of the antennas coordinates at Paris (index P) and Besançon (index B).

It also appears that the values of the time corrections were about the same for all the tracks at Besançon. This proves that the direction of the error vector is near ideally distributed towards the station/satellite direction of all the tracks. But this is not really a limitation of the method, because in this case the bad coordinates of the antenna have less influence on the time comparisons (see figure 4.A).

The highest improvement issued by the method concerns the correction on the link between the two receivers, given by the differential differences (figure 5). The mean over the period of measurement of the standard deviations goes from 6.7 ns to 4.0 ns. The precision of the time comparisons has then been improved of 40 %. This improvement is a very good result for such a simple method. It proves that the coordinates were really introducing a high bias in the time comparisons between these two stations.

The next step of the experiment was to put the new coordinates into the receivers, and to measure the improvement that could be obtained. Unfortunately, the GPS receiver at Besançon fails at this moment (october 1989), and this part of the experiment has still to be done.

CONCLUSIONS.

The most important result of this study is the conclusion that the antenna coordinates of the receivers are wrong! This could be the first step to initialize a new campaign of positionning in France, especially for the time and frequency laboratories taking part in the TA(F) ('Temps Atomique Français': french atomic time scale).

Nevertheless, we present a very simple method, based on the navigation solution of the receiver, to calculate corrections due to bad antenna coordinates for the improvement of time comparisons by GPS, for the people using Allen Osborne type GPS receivers. We include the clocks drifts into the measurement of the coordinates by the receiver. The relative positionning on the Paris/Besançon link (around 420 km) gives a precision and an estimated accuracy (by comparing the results of two completely different methods) of around 1 m. This means 3 ns on the time comparisons, which is for such a simple method a very good result. The improvement of the precision of the time comparisons, obtained after two months of measurements on the particular link Paris/Besançon is about 40 %.

We believe that this level of precision is the best numerical value attainable when using the navigation solution with this kind of GPS receiver dedicated to time comparisons. The use of a geodetic specialized GPS receiver is necessary to obtain a precision and an accuracy of 1 ns (or better) with the positionning of the antenna.

AKNOLEDGEMENTS.

The authors want to give a special thank to Pascal Blondé from LPTF for his help in computing the data, and a very special thank to Michel Vincent from Besançon Observatory for having kindly provided the data from his GPS receiver.

This work was supported by the Bureau National de Métrologie (french national bureau of standards).

BIBLIOGRAPHY.

[1] B. Guinot, W. Lewandowski.
Improvement of the GPS time comparisons by simultaneous relative positionning of the receiver antennas.
Bulletin Géodésique, 63, 1989, pp. 371-386.

[2] B. Guinot, W. Lewandowski.
Nanosecond time comparisons in Europe using the GPS.
Proceedings of the 2nd EFTF, Neuchâtel, 16-18 march 1988.

[3] M. Kihara, T. Okada.
A satellite selection method and accuracy for the Global Positionning System.
Navigation: Journal of the Institute of Navigation, Vol. 31, n° 1, 1984.

[4] Private communication with W. Lewandowski.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 7

THE PHASE STABILITY OF THE SYNCRONIZATION PULSE IN A 600 KM TV LINK SYSTEM

Jari Mannermaa
University of Oulu, Finland

Abstract

The phase stability of the Finnish TV network between Oulu and Helsinki (about 600 km) was studied by taking samples with an automatic measuring system every 10 minutes or 1 second. For the measured phase values differentials, residuals by the second order regression and power density spectra by the maximum entropy method (MEM) were calculated. The average phase residual is about 14.3 ns between Oulu and Helsinki and it can be considered as an approximation on the phase stability of the transmission path. The noise level of the system is about 2.7 ns which is the average frequency stability of the other rubidium standard. In the respect of noise both the spectra of the differentials and residuals seem to follow 1/f and $1/f^2$ law.

Introduction

The TV networks and signals are used generally all over the world to compare accurate clocks to each other. In Finland among the others the traceability of frequency is supplied through the TV network. The syncronization pulses of the Finnish TV network are stabilized using a high quality rubidium oscillator at Helsinki TV center. The National Standards Laboratory controls the frequency of the oscillator. The phases of syncronizing pulses are measured with a Loran- C controlled rubidium clock as a reference in the Laboratory of Electronic Measurements in University of Oulu, 600 km to the north of Helsinki. The phase and frequency instability caused by a long distance radio link system was the subject of the study.

There are no accurate measurements on the stability of the Finnish TV network and the knowledge of it is quite poor at this moment. However, the arctic climate conditions, especially in winters, may create unique effects on the characteristics of the transmission path.

To estimate the performances of the TV frequency standards and other devices it has become actual to increase the knowledge by collecting data of the Finnish TV-network and by analyzing it. To collect the data an automatic measurement system was developed and the measured values are analyzed with Lotus 123-software and with a MEM program for the spectral computation.

Method

The measurements were carried out by an automatic measurement system. The time difference between the frame syncronization pulse of the TV-signal and our main clock is measured in every 10 minutes or 1 second using a digital counter with 1 ns resolution. The main clock is a rubidium frequency standard (Rhode & Schwarz XSRB) but it is compared to another rubidium clock (Rb2, Efratom) to increase the accuracy and the reliability of the system. For the Rb2 the same measurements were carried out as for TV-signal data. All results during 10 days (or from 10 to 30 minutes if the sampling interval of 1 second is used) are saved on a computer disk. At first the obvious phase jumps are corrected. Then the linear trends of the daily data are removed using the second order regression method and the residual standard deviations are calculated. The standard deviations of the residuals and differentials are the phase stability and the frequency stability, respectively. Moreover, the distributions and the power density spectra of differentials and residuals are computed in order to determine the type of noise. The time periods which are buried in the TV signal can be found out from the power density spectra, too.

Results

The results are calculated based on the data of 84 days with 88 values of each one during the time period 01.09.1989-28.02.1990. The measuring time per day was from 5.00 to 19.30 UTC. The total number of samples is 7392. The average values of the standard deviations of the residuals and the differentials both on the TV and the Rb2 data are presented in the table 1. The distributions of the residuals of the TV network and Rb2 are plotted in Figure 1.

Table 1.

	A	B	C	D
September	12.7	4.6	14.1	2.2
October	14.3	5.2	13.1	2.4
November	11.6	6.1	13.2	2.7
December	11.5	7.2	13.0	3.1
January	16.9	5.7	19.9	3.3
February	18.8	4.0	18.9	2.6

where
A is the phase stability of the TV network
B is the phase stability of the Rb2
C is the frequency stability of the TV network
D is the frequency stability of the Rb2

It can be seen from the table 1 that the phase stability of the Finnish TV network and the Rb2 are approximately 14.3 ns and 5.5 ns, respectively and the frequency stabilities are 15.4 ns and 2.7 ns, respectively.

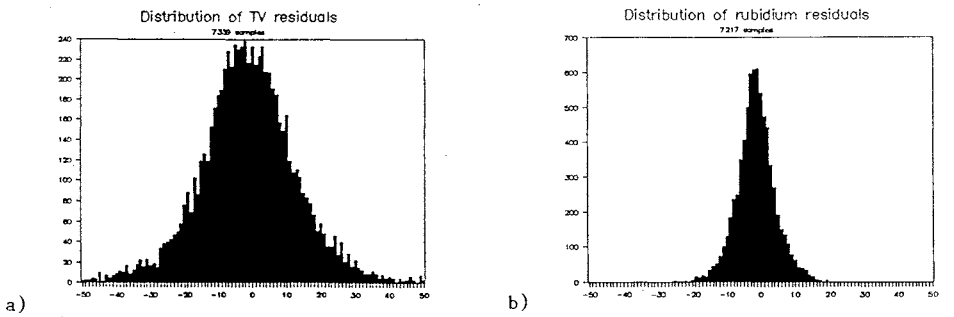


Figure 1. The distributions of the residuals and differentials of the TV network and the Rb2: a) the TV residuals and b) the rubidium residuals. The used time increment is 1 ns from -50 ns to +50 ns.

In Figure 2. the spectrum of the phase stability of the sampling time of 1 second are presented. The total number of datapoints is 2000. In the same figure a curve of the form $a/f+b/f^2$ is fitted to the data to describe the possible noise behaviour of the data.

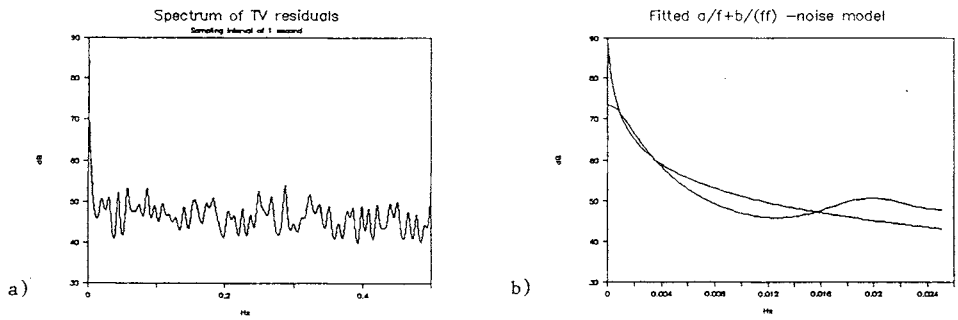


Figure 2. a) The spectrum of the TV residuals of the data of the sampling time 1 second and b) a curve of the form $a/f+b/f^2$ fitted to the the spectrum in a). The unit of the frequency axes is Hz.

In Figure 3. the MEM spectra for the phase and frequency stabilities of the TV network are shown. The spectra are sums of the specra of 84 days. The sampling interval was 10 minutes.

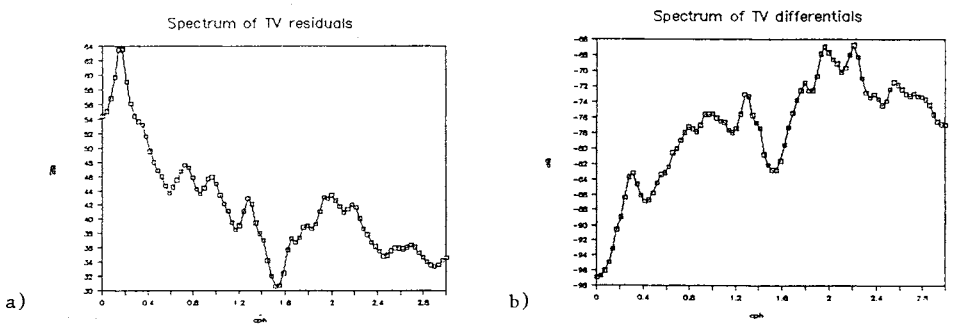


Figure 3. The MEM spectra of the residuals and differentials of TV network: a) the TV residuals and b) the TV differentials. The sampling interval was 10 minutes. The unit of the frequency axes is cph.

Conclusions

The values of the phase and frequency stability of the Rb2 show that with our automatic measurement system having a digital counter with 1 ns resolution (equal to the accuracy of $1.2\ 10^{-14}$ per day) is possible to reach the accuracy of $3.1\ 10^{-14}$ per day. This means the noise level of 2.7 ns, which is the frequency stability of the Rb2.

The phase stability of Finnish TV network is in average 14.3 ns but about 10 ns changes may sometimes occur. The respective values of the Rb2 are 5.5 ns and 2.7 ns. The values of the Rb2 can be considered as a measure of the noise level which are 5.5 ns and 2.7 ns for the phase and frequency stability, respectively.

All distributions are quite symmetric but the distributions of the TV residuals and differentials have quite flat top with the width of 10 - 14 ns. The halfwidths of distributions of the TV residuals and differentials are 27 ns and 28 ns, respectively and the same values of the Rb2 are 9 and 6 respectively. The flat tops refer to the drift and random walk noise process.

The MEM spectra point out that there are no strong periodic changes in the data of the residuals or differentials. Moreover, the behaviour of the spectra are quite correct because the spectra of the residuals have a negative slope in average and the spectra of differentials have a positive one.
The MEM spectra and the distributions of the residuals and differentials imply that the noise type of Finnish TV network follows seems to be a linear combination of the $1/f$ and $1/f^2$ laws.

The spectrum analysis points out that there are some periodic effects in short term data but they are smoothed in the spectra of the long term data. Moreover both the long and short sampling interval spectra support the thought of the random walk noise process instead of drift or white noise processes.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 8

AUTOMATIC AND FAST TELEVISION TIME COMPARISON FOR THE ZIMMERWALD SATELLITE LASER RANGING STATION

Paul Klöckler, Th. Schildknecht

Astronomical Institute, University of Berne
Sidlerstrasse 5, CH - 3012 Berne, Switzerland

Abstract. Although GPS timing now is standard at most satellite laser ranging stations, common view television time transfer still is attractive because of its independence from a foreign satellite system primarily dedicated to defense purposes.

Since 1988, this method has been applied routinely for near automatic time transfers between Zimmerwald SLR and OFMET, using dedicated time interval and data acquisition devices called TICOMT. The only operator interference concerns the electronic mailing of the data gathered at OFM to the Zimmerwald station computer once a week.

The system can be made really fast and fully automatic by accessing the remote TICOMT randomly via data-network to achieve time transfer with accuracy below 50ns in real time at very low investment costs.

1. INTRODUCTION

Zimmerwald Satellite Laser Ranging Station needs a local time scale with sub-microsecond precision for time-tagging the measured ranges to geodetic satellites.

For the comparison of local with national time scales, the line synchronisation pulses of TV transmissions have been used for more than 20 years. [Tolman,1967].

Because the method still is an attractive alternative to GPS timing receivers, we have conceived and built a number of automatic units called **TICOMT** (**T**ime **I**nterval **Co**mparation **M**odules using **T**elevision). Although commercial time-interval counters and microcomputers could have been used, a dedicated solution was chosen with the following main characteristics:

- Communication with a personal computer (or mini computer) by serial link - user-friendly software - runs automatically and autonomously - can be remotely programmed and monitored

2. DESCRIPTION OF THE SYSTEM CONNECTING ZIMMERWALD SLR TO UTC(CH)

2.1 Block Diagram of the Time Comparison

Figure 1 shows the general setup of such a time transfer. Two TICOMTs are placed near the clock under test and the reference. Each TICOMT is interrogated weekly by a local computer. Between the two computers, a data link allows electronic mailing of the raw data in the shown direction. One computer serves as a data-bank and allows remote access to the gathered data. A differencing program can be run that draws the data from the data-bank and places the result files back into the data-bank. A remotely usable graphics program allows inspection of the clock's drift from an external terminal.

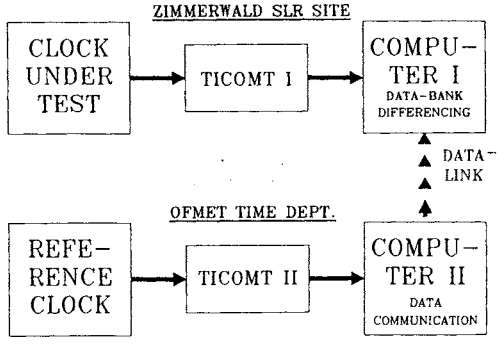


FIGURE 1: SET-UP OF TV-TIME-TRANSFER IN ZIMMERWALD SLR

Due to the open architecture of the system, several clocks can be measured and their respective data stored and compared to the reference.

The person responsible at OFMET has not only the weekly task to communicate the data to the data-bank, but he also has to supply an estimate of the delay between the reference and UTC(CH). For this reason mainly, a weekly data-exchange has been chosen.

2.2 Block Diagram of the Functional Parts

The TICOMT - as presented in Figure 2 - consist of a set of components, most of which are commercially available:

1. **Television Tuner**
   Receives the TV-channel and converts it to the composite video signal. A picture monitor optionally allows supervision of the reception quality.
2. **Line Selector**
   Extracts a preselected line-sync pulse. Has TTL output.
3. **Timecode Receiver**
   Provides automatic and maintenance-free time and date.
4. **Host Computer**
   PC or minicomputer for periodical data collection and exchange.
5. **Telephone Modem**
   Links the computer with the data-base

The only non-standard item is the TICOMT **measuring and data-logging unit** which performs the following tasks:

- Time-interval counter with 100 or 10 ns resolution
- Decoder of the timecode signal
- Data-logger with programmable measuring window
- Ring memory for the storage of one week's data
- Communication with the host-computer
- Status display and messaging

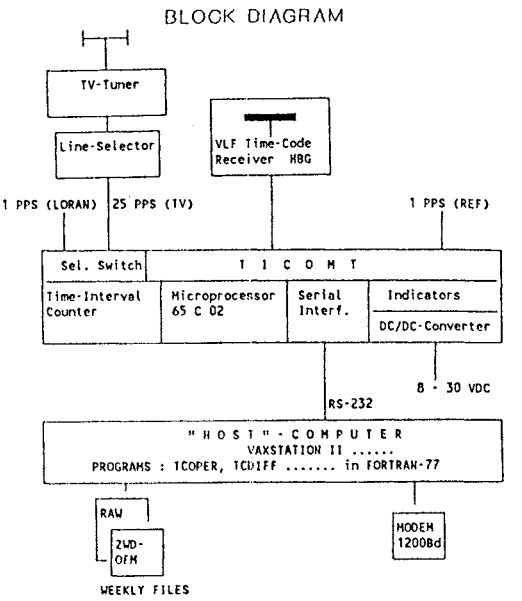


Figure 2

The associated software is contained in Read-Only Memory for the TICOMT measuring and data-logging unit. A set of FORTRAN-programs for the interaction between TICOMT and host computer allows automatic data transfer and estimation of time-scale difference.
Figures 3, 4 and 5 show some of the remote screens available for the user at the remote PC-Terminal. Note especially the immediate information on the measured time interval available in screen 3. Figure 5 plots three weeks' TV-transfer data as displayed on a remote terminal. The ASCII pseudo-graphics has a low resolution but can be displayed on non-graphic terminals. It serves well for remote supervision of clock behaviour.

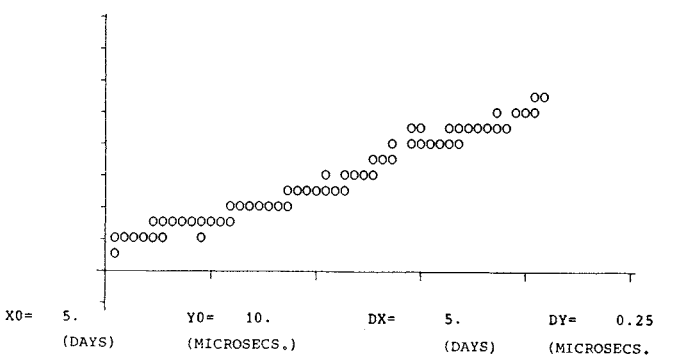
```
                    TCOPER
 1          INTERACTIVE OPERATION

- TICOMT synchronized to HBG signal          : YES
- HBG interrupt occurred between two meas. interr.  : YES
- meas. interrupt occurred between two HBG interr.  : YES
- measured signal                           : LORAN
- TICOMT in block measure process           : NO
- TICOMT data buffer full                   : NO
- TICOMT data buffer empty                  : NO

- 1 PPS REFERENCE signal                    : PRESENT
- HBG signal                                : PRESENT
- LORAN signal                              : PRESENT
- TV signal                                 : PRESENT

| 1 TEST | 2 STATUS | 3 MEAS. TV | 4 MEAS. LOR | 5 BACK |
```

Figure 3: Screen 2 (STATUS)

```
                    TCOPER
 1          INTERACTIVE OPERATION

RECORD ID     STATION NUMBER     HBG SIGNAL QUAL.     EPOCH
   I2               I                 GOOD            UT+1h

DD-MMM-YY   HH:MM:SS    TV    MEASUREMENT (MS)
09-MAR-90   09:24:09          30.5739

| 1 TEST | 2 STATUS | 3 MEAS. TV | 4 MEAS. LOR | 5 BACK |
```

Figure 4: Screen 3 (MEAS. TV)

FIGURE 5: PSEUDO-GRAPHIC SCREEN AT REMOTE TERMINAL



3. RESULTS

Figure 6 shows 18 days of two-hourly time comparison between Zimmerwald SLR and OFMET. The maximum deviation from the straight line is about 200 ns; the r.m.s of the single point, however, is about 50 ns. This shows that epoch errors of the satellite data can be corrected "a posteriori" with a precision much higher than the required microsecond. Yearly clock transports verify the absolute TV-delay between the two TICOMT on the 10 ns level.

Figure 7 displays one week of conventional LORAN-C tracking. Note that the noise of the TV-transfer in Fig.6 is four times lower than the one of the LORAN-transfer.

FIGURE 6: RESULTS OF 19 DAYS TV-COMPARISON BETWEEN ZIMMERWALD SLR AND OFMET.

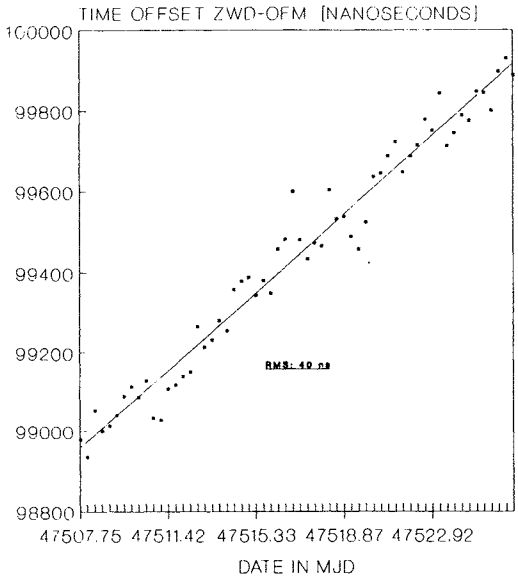
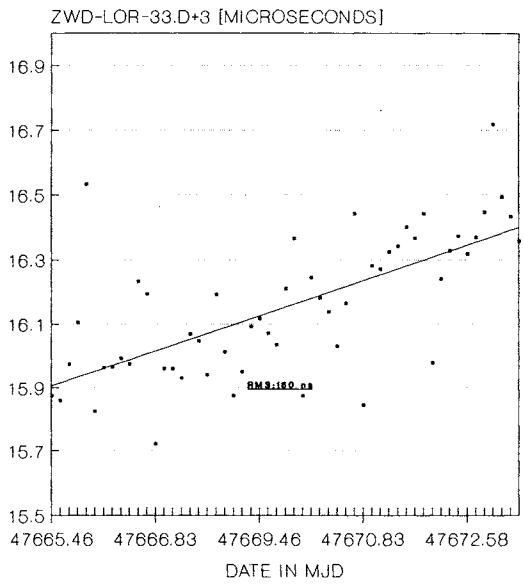


FIG.7: RESULTS OF ONE WEEK LORAN-C TRACKING AT ZIMMERWALD SLR



4. CONCLUSION

TV time transfer still is attractive for short distances, e.g. common view of the same TV-transmitter. It can be performed automatically. Since it is a relative method, an absolute calibration has to be performed. **Due to the inherent stability of the involved delays, precisions below 50 ns are routine and 10 ns feasible.**

One disadvantage in the operation is the data exchange, which for the moment is done weekly via telephone modems. This limits the application to "a posteriori" time-scale corrections. This latter limitation, however, can be overcome, provided the remote station can be accessed randomly via a data-network. Even via switched telephone line, **one instant time transfer per second** is possible. The concept sketched in Fig. 8 would allow "random access" of the remote TICOMTs with no local operator needed.

One final remark concerns the **Resolution** of the time/interval measurement. Prototypes exist with 10 ns resolution, which is about the limit for a simple counter built by fast TTL integrated circuits. The prototypes work, but the multiplication of the reference frequency yet has to be improved.

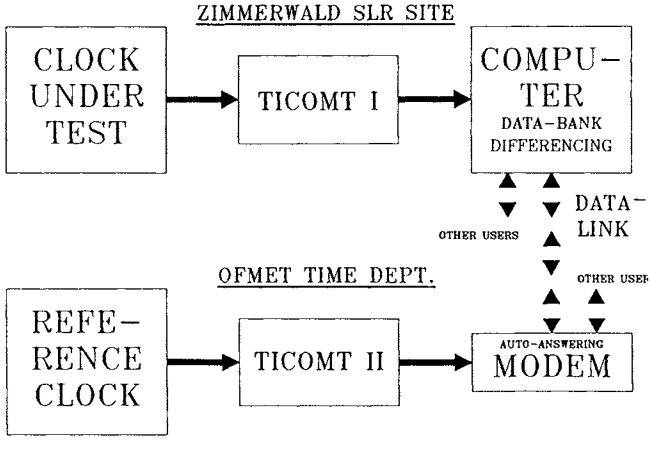


FIGURE 8: SET-UP OF PROPOSED REALTIME OPERATION

5. ACKNOWLEDGEMENTS

The Astronomical Institute of Bern University is indebted to Messrs. **Prost** at **OFMET**, **Schumacher** at **Observatoire de Neuchâtel** and **Melana** at **PTT** for the kind cooperation in this project.

Session PO-02 Number 9

Omega – VLF Signal Synthesizer

Enrico Rubiola
Politecnico di Torino — Dipartimento di Elettronica.

Introduction

Besides the GPS, that will be the the main system for time / frequency comparisons and radionavigation, the Omega system finds some fields of application not covered by the satellite based methods. Up till now, Omega is the unique system offering a frequency or positioning reference continuously available, with global coverage. Moreover, at the Omega frequencies the signals are pervasive and not easily disturbed by buildings or natural obstacles; a poor signal-to-noise ratio normally do not significantly affect the accuracy of the receivers.

Many details about this system can be found in the references [1,2,3] and [4], that reports also VLF time standard emissions.

Applications include maritime and aviation navigation, civil and military, land vehicle positioning, frequency monitoring and timekeeping. New receivers, capable of automatic switching between hyperbolic / range-range / differential Omega and VLF navigation modes are gaining acceptance as an alternative or as a supplement to the inertial navigators.

The aim of this paper is to present a general purpose Omega - VLF synthesizer designed for testing the receivers. The main feature is the high time stability and accuracy, obtained with the digital synthesis of all the signals from a single frequency reference. With this technique an high phase-shift resolution, about 12 ns, is easily achieved.

All these characteristics make the instrument suitable for testing metrological equipments, disciplined frequency standards and navigation receivers.

Signal	Frequency (kHz)	N	M
Omega f1	10.2	260	1
Omega f2	13.6	195	1
Omega f3	11.(3)	234	1
Omega f4	11.05	240	1
GBR	16	663	4
NAA	24	221	2

Table 1: parameters for the synthesis of Omega and VLF frequencies with the sampling frequency $f_s = 2.652$ MHz.

1 Frequency Synthesis

The digital synthesis of the Omega and some VLF signals can be obtained from a unique sampling frequency $f_s = 2.652$ MHz by direct division and multiplication. A look-up table, driving a digital to analog converter, is filled with N samples of M periods of the output sinusoidal waveform. The sampling algorithm take account and corrects the phase errors arising from the amplitude quantization. The output frequency is $f_0 = M/N f_s$. All the parameters for the six synthesizers used in the instrument are shown in table 1.

The block diagram of one synthesizer is reported in fig. 1, where the counters are modulo N; the leftmost one is the time reference from which the rightmost one is delayed by the *phase high* bits. The basic resolution is $t_s = 1/f_s \simeq$ 377ns.

The resolution is reduced to the value $\delta t = 2^{-k}/f_s$ by switching the memory mapped shape; this is done by the *phase low* bits (see fig. 1). The theoretical limit of k, keeping the phase accuracy within one bit, is not known. It depends on the response time, the resolution and the accuracy of the DAC, the sampling frequency and the ratio M/N. A practical limit has been found by computer simulation with some realistic models of the errors distribution on the error budget in the ladder network of the converter. The most significant bit does not give any con-

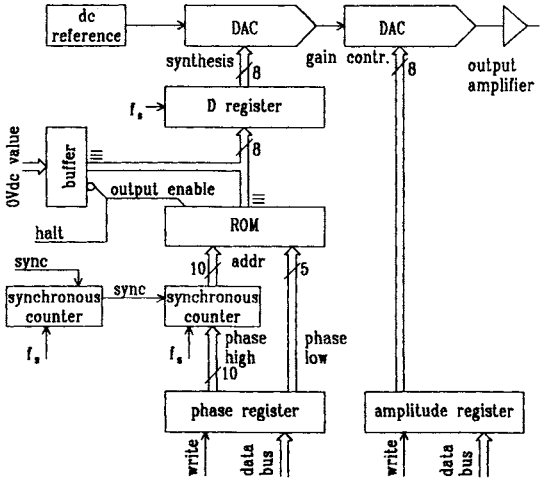


Figure 1: block diagram of frequency synthesizer cell.

tribution to the phase error, and different hypotesys on the error distribution of the other bits give similar results.

Using an 8 bit converter accurate within 0.25 LSB with a settling time of 135 ns, the limit is $k = 5$ that leads to a resolution $\delta t \simeq 11.8$ ns, i.e. 19 μrad at 10.2kHz. Measurements performed on the prototypes shows systematic errors, depending on the *phase low* value, distributed in the range ± 10 ns.

The amplitude control is performed with an auxiliary multiplying DAC. The amplitude can be forced to zero locking the main DAC (*halt* signal); this technique allows a rejection of -115 dB, probably limited by the parasitic capacitances of the wire-wrapped prototype.

The output amplifier includes a large bandwidth low pass filter (100 to 160 kHz, depending on the output frequency) for the suppression of the highest noise frequencies.

Some harmonics are present at the output, about 55 dB below the signal.

2 Accuracy and stability

The digital inputs of the converters are synchronized to the *clock* signal by D-type registers (fig. 1), which has a delay time of a few nanoseconds, stable within hundreds of picoseconds. Both of the accuracy and stability depend only on the delay from the clock of the register to the analog output, with a minimum contribution of the output amplifier because of its small total delay due to the bandwidth; the problems arising from the frequency refrence are not considered here. This is the same concept

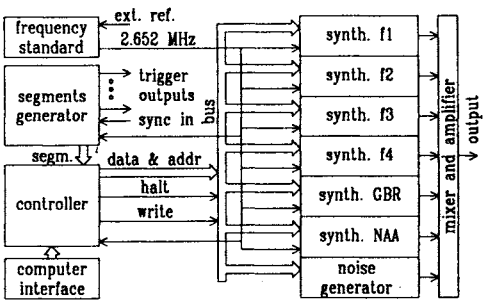


Figure 2: block diagram of the whole instrument.

each transmitter. All the phases and amplitudes of each station can be individually set, allowing the simulation of the various propagation phenomena, such as the self-interference of a signal.

Two auxiliary carriers, with characteristics close to the other signals and corresponding to the VLF stations GBR and NAA, can be generated for simulating coherent interferences or additional references. Noise also can be added.

All the phase and amplitude values, about 100 bytes, are stored inside the controller in a dual port memory which can be written by an external computer at any time.

of accuracy and stability explained in [5].

The delay from *clock* to the zero-crossing of the analog output of one of the synthesizers (10.2 kHz) has been measured in the -10 to $+80$ °C range. The variations from the $+22$ °C value never exceed ± 15 ns; the umidity, not monitored in the test chamber, is probably an important factor in these phase variations.

Slightly higher phase variations (± 20 ns) are due to amplitude changes in the gain control. They are repeatable and can be corrected by the external computer.

3 Description of the instrument

The whole instrument is built around 6 synthesizers — one for each Omega frequency and two additional ones for the simulated VLF stations — and a noise generator, as shown in the block diagram of fig. 2.

The segment generator provides the timing pattern for the Omega system, also available as trigger pulses at the front panel. The segment generator can be synchronized to the *pps* of an external time scale and provides the synchronization to the synthesizers. The leftmost counter of the fig. 1 is locked to the segment generator.

The controller provides the phase shifts and amplitude levels to all the synthesizers writing their control registers at a very low rate during the dead time between the segments.

The frequency synthesizers f1 to f4 are shared by the eight Omega stations. In this way the instrument can simulate the complete Omega pattern (eight stations, four frequencies each one), excluding the frequencies unique to

4 Final remarks

Phase accuracy and stability are due to the direct digital synthesis method.

An analog deglitcher (sample-and-hold) at the output of the main converter could probably increase the overhall performances.

The phase resolution can be improved proportionally to the number of bits of the converter; fast 18 bit DACs are commercially available.

The phase resolution obtained switching the memory mapped signal shape seems to find other application fields, such as the synthesis of the 5751 Hz signal for the phase detector in Hydrogen masers.

References

[1] Swanson E.R. "Omega" in *IEEE Proceedings* vol. 71 n. 10, oct. 1983, pp. 1140–1155.

[2] "Radionavigation System" published by U.S. Coast Guard, 1984, code G-NRN, part III — Omega.

[3] "Bibliography of Omega Publications". Published by the *International Omega Association*, P.O. Box 2324, Arlington, Va 22202, USA.

[4] "Standard Frequencies and Time Signals," CCIR report 267-6, 1986.

[5] Rubiola E. "Loran-C Signal Synthesizer," *Proc. 3rd European Time and Frequency Forum*, Besançon 1989, p. 313–318.

Session PO-02 Number 10

RENAL CALCULI AS NEW PIEZOELECTRIC MATERIALS

Ravinder Agarwal and V.R. Singh
National Physical Laboratory, New Delhi **(India)**

During the study of mechanical and ultrasonic properties, renal calculi are found piezoelectric in nature. The samples were used in the form of circular tablets made from renal calculi in their solid compact original form. These were also specially made after grinding the stones into the powdered form and then compressing then. Such tablets are made by using conventional processing techniques as used for piezoelectric ceramic materials. Electrical and dielectric parameters of these stone tablets exhibited that renal calculi are piezoelectric in nature.

INTRODUCTION

Renal calculi are complex composite, and crystalline in nature (1-2) with a variation in its composition due to presence of calcium oxlate monohydrate, calcium oxalate dihyrate, calcium carbonate, uric acid, cystine silicate and magnesium ammonium phosphate hexahydrate (3). There are a number of preurinary risk factors which help in stone formation, but diet factor appears to dominate. Various mechanical, chemical and ultrasonic properties have been studied in the recent past (4,6), while studing design parameters for an ultrasonic lithotripter (7). In the present study, electrical, dielectric and piezoelectric properties of the tablets of the calculi made from solid as well as from powdered stones are studied.

MATERIAL AND METHODS

Specimen Prepration

(i) From Solid Compact Renal Calculi

The tablet specimen of renal calculi were obtained from the local hospitals. After biopsy, calculi were allowed to dry for a month and were then machined. Electrical contacts were provided by using air-dry silver paste on either side of the stone. After proper silvering, the specimen were poled at DC electric field of 10 kV/cm to 12 kV/cm at elevated temperature of 150°C. The measurements of piezoelectric properties were made for both poled and unpoled samples.

(ii) Tablets from Powdered Renal Calculi

Several specimens collected from different hospitals were crushed into powdered form. The powder thus obtained was ground with the help of dry ball-milling for 12 hours. The stone powder was mixed with a few drops of polyvinyl alcohol (PVA) and compressed into the circular tablets having diameter of 19 mm and thickness 3 mm. These tablets were then allowed to dry for ten days. These tablets were then electroded with air-dry silver paste on either side by applying DC electric field of 8 kV/cm to 9 kV/cm at elevated temperature of 80 - 90°C. The measurements were then made both on poled and unpoled samples.

(iii) Measurements

The measurement of electrical and piezoelectric parameters at room temperature were made, after a gap of two days after poling of the samples for better stability due to ageing. The measurement of capacitance and dielectric loss factor (tan δ) at 1 KHz was made by using a precision impedance (LCR) bridge. The resistivity was measured by employing an ohm-meter (make BPL, India, Million Megohmmeter model RM 160 MK III A). The piezoelectric charge contant (d_{33}) was measured by using Aerotech ("d_{33}" Tester model A-2). The dielectric constant was then determined by using the relationship:

$$K_3^T = \frac{C.t}{0.0885.A} \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

where 't' is thickness (cm) of the sample, 'A' is the area (cm^2) and 'C' the capacitance (pf). Voltage constant g_{33} was calculated from the measured values of charge constant and dielectric constant by using.

$$g_{33} = \frac{d_{33}}{\varepsilon_o \, K_3^T} \quad \ldots \quad \ldots \quad (ii)$$

where 'ε_o' is the permittivity of free space, which is equal to 8.854×10^{-12} F/m.

RESULTS AND DISCUSSION

Variation in the electrical and dielectric parameters, viz., capacitance, resistivity, dielectric loss factor, dielectric constant, charge constant, voltage constant and curic temperature are found due to variation in chemical composition. The range of electrical and dielectric parameters are given in Table - I for various renal calculi in solid compact original form as well as for specially made samples after grinding them into powdered form. The resistivity is found to vary from 14.8×10^8 to 20.0×10^8 ohm-cm for solid shaped calculi and 12.0×10^8 to 23.0×10^8 ohm-cm for renal calculi tablets. The dissipation factor and dielectric loss factor of the solid calculi specimen are found to be relatively more than that for the powdered stone tablets. A range of variation in these values from minimum to maximum is also given for such samples in Table - I. There was no piezoelectric effect before poling as found by taking measurements on impoled samples. The curic temperature of the stone is found to be in the range of 235°C to 250°C. The results of d_{33}, g_{33} tan δ etc. demonstrate the property of piezoelectricity in renal calculi. The tablets with lower loss factor (tan δ) are suitable for piezoelectric devices such as medical therapy probes and those with relatively low value of g_{33} constant are suitable for hydrophones. Hence, renal calculi are another source of piezoelectric materials.

CONCLUSION

The value of electrical and dielectric parameters have been measured for solid shaped calculi as well as for tablets made from the powdered stones. The data obtained may be useful for reasearchers in material science and biomedical applications, due to piezoelectric nature (mainly due to presence of alpha-quartz).

REFERENCES

1. Consdale Kathleen, "Human Stones", Science, Vol.19, pp 1194-1207, 1968.
2. Singh V.R., Agarwal Ravinder and Sud S.P., "Ultrasonic propagation in kidney stones", Indian Journal of Pure and Applied Physics, Vol.6, pp 475-476, 1988.
3. Agarwal Ravinder, Singh V.R., Dhawan, U., Singh Ved and Nagpal K.C., "X-ray diffraction technique", J Mater. Science, 1989 (Communicated).
4. Agarwal Ravinder, Singh V.R. and Singh Ved, "Microhardness studies in kidney stones", To be presented at All India Conference on Biomedical Instrumentation (BMIC-90), Feb 23-24, 1990, Lucknow (India).
5. Johrde L.G. and Cocks F.H., "Fracture strength studies of renal calculi", J. Matter. Sci. Lett., Vol.4, pp 1264-1265, 1985.
6. Singh V.R. and Agarwal Ravinder, "A study of ultrasonic characteristics of renal calculi in vitro", J. Acous. Soc. Am., Vol.85, pp 962-963, 1989.
7. Agarwal Ravinder and Singh V.R., "Determination of design parameters for an ultrasonic kidney stone disintegrator", J. Acous. Soc. Am., Suppl.1, Vol.83, S-89, 1988.

TABLE - I

PIEZOELECTRIC CHARACTERSTICS OF RENAL CALCULI

Sr.No.	Parameters	Solid Renal Calculi		Powdered Renal Calculi	
		Minimum values	Maximum values	Minimum values	Maximum values
1.	Capacitance (pF)	11.5	12.0	15.4	18.9
2.	Dielectric loss factor (tan δ)	0.030	0.033	0.023	0.031
3.	Dielectric constant (K_3^T)	32.12	32.41	16.43	18.79
4.	Resistivity (10^8 ohm-cm)	14.00	20.0	12.0	23.0
5.	Charge constant (10^{-12} C/N)	27.0	27.0	32.0	37.0
6.	Voltage constant (10^{-3} Vm/N)	94.09	94.98	19.23	25.43

Session PO-02 Number 11

USE OF A TENSORIAL REPRESENTATION OF THE DISSOLUTION SLOWNESS TO PREDICT ETCHING SHAPES OF QUARTZ AND GaAs CRYSTALS

C.R. TELLIER and A. BRAHIM-BOUNAB

ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Laboratoire de Chronométrie, Electronique et Piézoélectricité
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE

In the past few years a lot of experimental works (1 - 7) have been devoted to the study of the anisotropic etching of crystals. Interest in the photolithographic process as an application of the anisotropic attack of semiconductor crystals such as Si (7) or GaAs (5, 6) for the fabrication of microelectronic circuits of high density or for the micromachining of mechanical devices (7) has been in continuous progress. But the photolithographic process can also be applied to the preparation of miniaturized quartz resonators and sensors (8). The design of such micromechanical devices requires generally a great accuracy in the geometrical definition of the structure and a rigourous control of the orientation. For this purpose it becomes very interesting to deal with theoretical models which allow us to follow the changes with the orientations of the mask and of the starting wafer in

(i) the shape of the section of etched grooves
(ii) the lateral etch under the mask (6, 9, 10)
(iii) the corner undercutting (11).

An interesting method to approach these problems consists to start from the tensorial method for the research of the dissolution slowness surface of crystals recently proposed by Tellier et al (12, 13, 14). Effectively as soon as the equation for the slowness surface is known we can track the various surface elements forming the initial structure and moving within the crystal during the chemical attack (13) and then by distinguishing deverging and converging trajectories (3, 12) we can cosntruct numerically the etched shape. In the following some results are given in the case of quartz and Gallium Arsenide crystals.

First of all let us recall that for the quartz crystal which possesses an axis x_1 of twofold symmetry and an axis x_3 of threefold symmetry the generalized equation for the slowness surface is written as (12)

$$L_{XYZ}(\varphi, \theta) = \sum_j A_j \cos^{2j}\theta + \sum_k B_k (\cos\theta)^{2k+1} \sin 3\varphi + \sum_l C_l (\cos\theta)^{2l+1} \sin\theta \cos 3\varphi \qquad (1)$$

whereas for the GaAs crystal which belongs to the cubic class $\bar{4}3m$ and then is a non-centrosymmetric crystal with three fourfold rotation-inversion axes x_1, x_2 and x_3 the equation for the slowness surface takes the form

$$L_{XYZ}(\varphi, \theta) = \sum_{k=0} a^*_k \cos^{2k}\theta \sin^{2k}\theta + \sum_{l=1} \sum_{m=0} b^*_{l,m} (\sin\varphi \cos\varphi \cos^2\theta)^{2l} (\cos\theta)^{2m} +$$

$$+ \sum_{r=0} c_r (\sin 2\varphi \sin 2\theta \cos\theta)^{2r+1} + \sum_{s=0} \sum_{t=0} d^*_{s,t} \sin\theta (\sin\varphi \cos\varphi \cos^2\theta)^{2s+1} (\cos\theta)^{2t} \qquad (2)$$

In eqns. (1) and (2) we adopt the convention specified in the I.E.E.E. Standard on Piezoelectricity (15) to define the angles of cut φ and θ. Moreover it should be noticed that to satisfy completely to the symmetry associated with the class 32 and $\bar{4}3m$ the unknown coefficients $\{A_j, B_k, C_l\}$ and $\{a^*_k, b^*_{l,m}, c_r, d^*_{s,t}\}$ are related together by additional equations characteristic to the respective point group symmetry.

Using these equations numerical simulations have been developped for the graphical representation of the etched shape (i) of the initially circular section of a crystal, (ii) of the cross-sectional profile of a groove. In the case (ii) the etched section is composed of potentially present elements whose orientation is detected by an angle α lying in the range (90°, 270°) (figure 1). If the inert mask deposited on a singly-rotated wafer is aligned along a direction which does not coïncide with a reference axis each planar element of the initial section corresponds to a doubly-rotated surface element of orientation (φ, θ) and the graphical simulation starts with a polar diagram of the dissolution slowness involving varying values of the angles φ and θ. The final etched profile (Figure 1) which is formed by the elements which do not disappear during the dissolution is characterized by the bordering plane whose intersection with the reference surface is visualized by means of the element LB and by the lateral etch corresponding to the distance OL.

In the case of quartz crystals we have chosen to present some results for various singly rotated quartz plates (φ = 0°, θ_0) with a groove lying parallel to the crystallographic axe x_1 so that only the polar graph $L_{YZ}(\theta)$ is needed for the simulation. The results displayed in Figures 2 to 4 clearly indicate that the shape of the etched section as well as the extend of the lateral etch markedly depend

on the orientation of the wafer. For some particular values of the angle of cut θ_o, we observe the development (Figure 3) of limiting planar elements. Because during the formation of an etched groove we are concerned only with concave intersections these elements correspond to limiting planes associated with maxima in the dissolution slowness.

For <100> GaAs slices with grooves aligned along a (011) direction the theoretical shape of the etched section is determined by the polarity of the reference surface (As or Ga character). The theoretical results for the (011) section and the orthogonal (01$\bar{1}$) section are respectively illustrated in Figures 5 and 6. For the (011) section the numerical simulation yields an etched profile with a relatively flat portion LD and a curved bottom DE whereas for the (01$\bar{1}$) section the simulation generates a profile with a single curved portion LD. Such cross-sectional profiles have been effectively encountered in experiments on GaAs slices **(2, 5, 6)**.

The few examples clearly show that a tensorial representation of the dissolution slowness combined with a graphical simulation is applicable to analyze the etched cross-sectional profiles resulting from localized dissolution in presence of a mask. The knowledge of the complete equation for the slowness surface remains essential but it is possible to estimate the unknown constants from experiments with a sufficient accuracy to obtain a satisfactory correlation between predicted and experimental profiles.

REFERENCES

(1) R.B. HEIMANN in J. GRABMAIER (ed.), ***Silicon Chemical Etching***, (Springer, Berlin, 1982)
(2) K. SANGWAL, ***Etching of Crystals***, (North-Holland, Amsterdam, 1987)
(3) C.R. TELLIER, N. VIALLE and J.L. VATERKOWSKI, Surface Coatings Technol., 34 (1988) 417
(4) B.A. IRVING in P.J. HOLMES (ed.), ***The Electrochemistry of Semiconductors***, (Academic Press, London, 1962) pp 256-289
(5) Y. TARUI, Y. KOMIYA and Y. HARADA, J. Electrochem. Soc., 118 (1971) 118
(6) D.W. SHAW, J. Cryst. Growth, 47 (1979) 509
(7) K.E. PETERSEN, Proc. IEEE, 70 (1982) 420
(8) W. ZINGG, **C.N.R.S. Seminar on Frequency Standards, Besançon, 6-7 mars 1985**, (Editions du C.N.R.S., Paris, 1985) pp N-1-15
(9) L. CSEPREGI, Microelectron. Engng., 3 (1985) 221
(10) G. DELAPIERRE, Sensors and Actuators, 17 (1989) 123
(11) X-P WU and W.H. KO, Sensors and Actuators, 18 (1989) 207
(12) C.R. TELLIER and J.L. VATERKOWSKI, J. Mater Sci., 24 (1989) 1077
(13) C.R. TELLIER, J. Cryst. Growth, (1989)
(14) C.R. TELLIER, J. Cryst. Growth, (1990) to be published
(15) ***Standard on Piezoelectricity***, (IEEE, New York, 1978) pp 15-27.

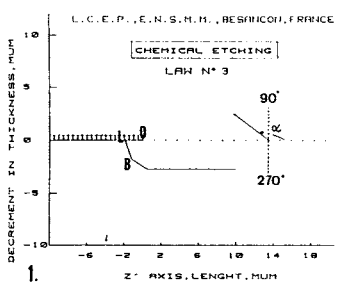


Figure 1 : The geometry of a numerically constructed cross sectional profile.

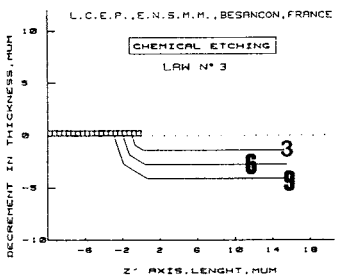


Figure 2 : The predicted cross-sectional profile for a quartz wafer with $\psi = 0°$, $\theta_o = -41°$

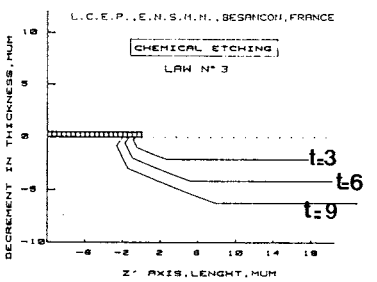


Figure 3 : The predicted cross-sectional profile for a quartz wafer with $\psi = 0°$, $\theta_o = -49°$.

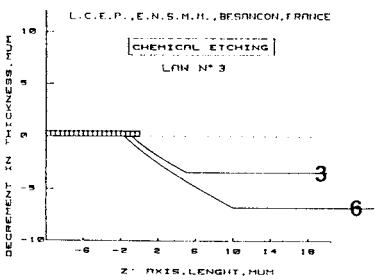


Figure 4 : The predicted cross-sectional profile for a quartz wafer with $\psi = 0°$, $\theta_o = 35°$

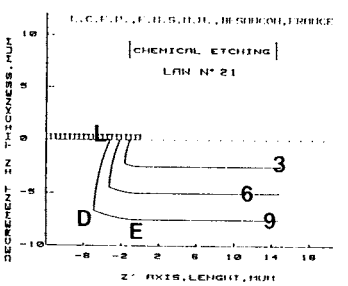


Figure 5 : The predicted (011) section through a (100) GaAs wafer

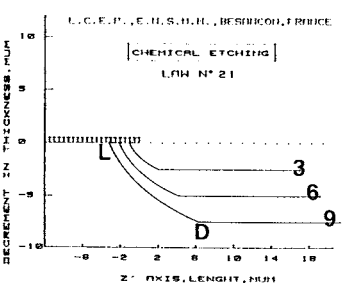


Figure 6 : The predicted (01$\bar{1}$) section through a (100) GaAs wafer.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 12

OSCILLATORS FOR REAL-TIME ON-BOARD APPLICATIONS OF NAVSTAR-GPS

Jean Paul LACROIX - SEXTANT AVIONIQUE

25 rue J. Védrines 26027 VALENCE - FRANCE

1. INTRODUCTION

The NAVSTAR-GPS is basically a worldwide radio-navigation system, which addresses a very large range of navigation applications : fixed and rotary-wing aircraft, ships, land vehicles, space, missiles, RPV's, in the civil and military domains. It is based upon time measurements performed on the signals provided by a satellite constellation which constitutes the space segment and theoretically ensures a continuous worldwide coverage. The system navigation performances range from 10 to 100 meters 3-D position accuracies, associated with a velocity accuracy that can be as good as 0.1 m/s, thanks to the ranging (PRN code correlation) and velocity measurement (L-band Doppler shift) techniques used in the receivers (these performances are consistent with the time transfer accuracy of the Navstar-GPS system).

These system principles (time and frequency, measurements) induce the need for a user time and frequency reference, which is part of the receiver, and has to perform well enough to make profit of the system capabilities. This paper details the requirements that may be considered for the user equipment reference clock, concerning :

- aging
- retrace effect
- medium-term stability
- short term stability
- phase noise
- g-sensilivity
- warm-up time
- other caracteristics

and also lists some of the technologies that can be used. This paper is mainly centered on aircraft and high-dynamic applications ; however, a similar analysis may be carried out for the other areas.

2. AGING

A reference clock frequency error, at equipment initialization, such as the effect of oscillator aging, is seen by the receiver like an unknown additional Doppler shift affecting all signals. This Doppler shift is easily evaluated by

$fd = f_{L1} \cdot \Delta fc/fc$ where f_{L1} = 1575.420 MHz and $\Delta fc/fc$ is the frequency error due to aging

(i.e roughly 1.5 KHz for 10^{-6})

The effect of an unknown Doppler shift is to increase the acquisition time (because of frequency search) as soon as fd becomes significant with respect to the acquisition predetection bandwidth (say, over half this bandwidth). The acquisition bandwith depends on receiver design ; 500 Hz could be an average value. With this assumption, it limits fd to 250 Hz and then $\Delta fc/fc$ to about $1.5\ 10^{-7}$.

Two cases have to be considered :

a) No "continuous" estimation of $\Delta f/fc$ is maintained (e.g in non volatile memory). In that case $\Delta fc/fc$ must remain below $1.5\ 10^{-7}$ between each frequency adjustment operation (preventive maintenance) or during the whole equipment life. Considering 10 years for this period induces an aging requirement of :
$\leq$ 1.5 10-8/year

b) Continuous estimation of $\Delta f/fc$ is maintained. Then $\Delta fc/fc \leq 1.5\ 10^{-7}$ is only necessary between each equipment utilization (say a few months for a majority of applications). It only leads to an aging requirement of 1 to 2 10^{-7}/year. Over the equipment life (or between frequency adjustments), $\Delta f/fc$ will only have to be low enough not to induce fd values exceeding the overall receiver Doppler shift range (at least a few KHz, or $\Delta fc/fc \simeq$ a few 10^{-6}). This requirement is met with an aging characteristic of 2.10^{-7}/year.

3. RETRACE EFFECT

We can directly derive from the aging analysis that an additional Doppler shift fd, due to the retrace effect of the reference clock, should not be over 250 Hz. A retrace effect requirement of $\leq 1.5\ 10^{-7}$ can then be derived.

4. MEDIUM TERM STABILITY

We define here the medium term stability as the stability over periods from 10 minutes to 1 day. It is linked to the capability the receiver may have to operate over the same periods of time with only three satellites. This mode of operation, also known as "clock-coasting" is also used :

- to operate with four satellites, but a bad geometry
- to derive integrity informations and system failure detection when using only four satellites.

In this mode, the receiver clock is used as an absolute time reference, the initial biais and frequency offset of which have been estimated, and the effect of its drift on time and frequency are directly equivalent to ranging and velocity errors. Considering an unforeseen frequency drift of $\Delta f/fc$ (assumed continuous for simplicity of evaluation) over a period at time ΔT, the equivalent ranging error may be expressed as :

$\Delta R = 0.5\ \Delta t.\Delta f/fc.C$ (C being the velocity of light)

For instance a 10^{-9} drift over 10 minutes is equivalent to an error of 90 meters.

In these conditions we may deduce medium term stability requirements from :

- the maximum time the receiver may have to operate in this mode

- the accepted ranging error.

Table 1 gives some examples of application with the corresponding stability requirements (of course, if this mode is not used, e.g for low-cost equipment, almost no medium term stability is required).

APPLICATIONS	MAXIMUM Δt	ACCEPTED ΔR	STABILITY REQUIREMENT
SPS receiver (173 m accuracy) Continuity of operation during short outages (masks, bad PDOP's)	10 mn	100 m	~ 1.10-9 over 10 minutes
Same, PPS receiver (18 m accuracy)	10 mn	10 m	~ 1.10^{-10} over 10 mn
SPS receiver. Continuity of operation through all constellation "black holes"	40 mn	100 m	~ $2.5\ 10^{-10}$ over 40 mn
Same, PPS receiver	40 mn	10 m	~ $2.5\ 10^{-11}$ over 40 mn
SPS receiver. Use of a time reference for integrity monitoring (during longest periods without 5 satellites, taking into account failures), or for circular mode navigation	4 hours (order of magnitude)	100 m	~ $5\ 10^{-11}$ over 4 hours
Same, PPS receiver	4 hours	10 m	~ $5\ 10^{-12}$ over 4 hours

TABLE 1 : MEDIUM TERM STABILITY

5. SHORT-TERM STABILITY

We call here short-term stability the stability over periods ranging from 1 second to 10 minutes

Short term stability has an influence on :

- satellite acquisition. During this phase where the carrier loop is not locked, the fd (doppler shift) induced by the clock drift should be maintained well inside the predetection bandwidth. If we assume (as in § 2) a maximum fd of 250 Hz and a maximum acquisition time of 100 seconds (which can be the case for poor signal/noise ratios), we derive a short term stability requirement of $1.5\ 10^{-7}$ over 100 seconds. A reasonable safety margin would require $10^{-8}/100$ seconds.

- navigation accuracy of sequential receivers. These receivers have to propagate the pseudorange measurements over the complete fix period and the clock drift induces errors, as in the "clock-coasting" mode, over this period. For a 5 seconds sequence time and a 10 m accepted error component, $a \simeq 1.10^{-8}/5$ s stability is necessary.

- operation of the receiver during short outages, (for instance antenna masks during aircraft turns) where only 3 satellites could be usable. The analysis in this case is the same as for the medium-term stability. For 200 seconds (long turns) and 10 m accepted errors (PPS), $\simeq 3.10^{-10}$ over 200 s would be required.

6. PHASE NOISE

Clock phase noise has some importance for the GPS receivers, since it is directly transformed in additive receiver noise within the local oscillator chain. We consider phase noise is limited between 1 Hz and 10 KHz, assuming that :

- the spectrum below 1 Hz is the domain of short-term stability,

- the spectrum over 10 KHz is considered filtered in the local oscillator generation and/or in the receiver IF's.

The noise floor requirements (concerning the higher part of the spectrum, roughly 100 Hz to 10 KHz) are determined from the signal carrier to noise density ratio characteristics... For the GPS satellites, on the L1 C/A signals, C/No ratios of 50 dB.Hz or more are frequently encountered, and one can consider that not to degrade these signals the effect of oscillators phase noise should not be worse than a 60 dB.Hz C/No. This is equivalent to a - 60 dBc/Hz phase noise density.

We also have to take into account that the reference clock phase noise is multiplied in the local oscillator generation, by a ratio which is $(fL1/fc)^2$.

Where fL1 = 1.575.420 KHz

fc = clock frequency

For a 10 MHz clock, it adds about 44 dB to its phase noise, if the contribution of the local oscillator hardware itself may be neglected (which is the case for good receivers).

Finally, the noise floor requirement becomes :

- 60 - 44 = - 104 dBc/Hz

Noise components between 1 Hz and 100 Hz are less easy to analyse, since they depend from both the clock noise form of spectrum and the receiver tracking loops transfer functions. Il comes in fact directly from measurement noise requirements imposed on code and phase. For a quartz oscillator, the noise density represented on fig 2, assuming f^{-3} components from 1 to 10 Hz and f^{-1} components from 10 Hz to 100 Hz, can be considered as a specification basis for good receiver performance.

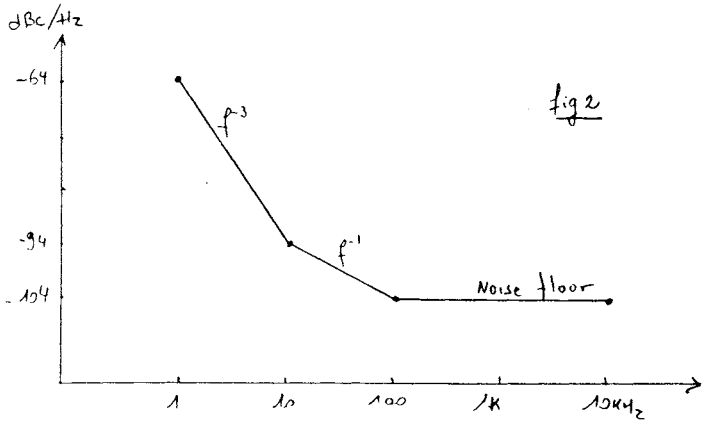


fig 2

7. GRAVITATIONAL SENSIBILITY

The clock gravitational sensitivity (g-sensitivity) can influence the receiver performance in several ways :

- its induces apparent Doppler shifts on all signals, in presence of vehicle accelerations, which may cause loop unlocks. In addition these shifts cannot be compensated by inertial aiding, in the general case. For instance, for Sg = 10^{-9}/g and γ = 10 g (fighter aircraft), a $\Delta fc/fc$ of 10^{-8} is seen, inducing a doppler shift of 15 Hz (to be compared with Costas loop bandwidths of 10 Hz or less),

- in addition, this Doppler shift may cause divergent navigation errors if the receivers operates in the clock-coasting (3 satellites) mode, if the acceleration experienced by the oscillator is not zero-mean (it is the case for γz on aircraft). Its an example, the same 10^{-9}/g with a mean γz acceleration of 0.2 g (possible for some aircraft trajectories) would be equivalent to a medium term drift of 2.10^{-10}. So, medium term stability requirements should be associated with an equivalent g-sensitivity requirement,

- mechanical vibrations, associated with Sg, cause the clock frequency to be modulated by coherent signals (sinusoidal vibration, or noise), degrading the signal to noise ratio, or being able to generate false locks.

Let us take two examples, coming from fighter aircraft environmental specifications.

1) Sine vibration : f = 100 Hz and γ = 10 g , Sg = 10^{-9}/g

   $\Delta fc/fc = 10^{-8}$, so $\Delta f \simeq 15$ Hz, or m = 0.15.

   This is equivalent to presenting at the receiver input two signal replicas $\pm$ 100 Hz apart, and only down about - 20 dB (then trackable).

2) With random vibrations (typically 0.05 g^2/Hz), we can find that Sg = 10^{-9} generates phase noise up to - 40 dBc/Hz, then limiting the signal C/No to 40 dBHz).

As a conclusion, Sg requirements could be :

- about 10^{-9} for low dynamic aircraft (or other applications)

- about 3.10^{-10}/g or less for high dynamic aircraft (it is actually limited by the state-of-the-art)

- roughly 1/5 times the medium term stability requirement if 3 satellites operation is required, with high dynamics.

8. WARM-UP TIME

Warm-up time requirements come from both operationnal needs (which are generally more or less defined) and from competition. A coarse analysis leads to average requirements of 3 to 10 minutes, depending on equipment class of performance.

9. OTHER CHARACTERISTICS

- Dimensions, weight

With the generalization of very integrated digital technologies (Ga As, ASIC's), GPS receivers tend to become smaller and smaller, and volumes of 1 liter or less for the complete receiver are now current. In these conditions, the oscillator dimensions are significant. Taking as an exemple, the new generation of aircraft GPS sensors, we think that the reference oscillator dimensions should not exceed 4 x 4 x 4 centimeters for medium-performance oscillators and 5 x 5 x 5 centimeters for high-performance units. Correlatively, the weight should be in the order of 100 g of less.

- Environmental conditions

Environnemental conditions are defined by standards. The most important requirements, concerning oscillators, are temperature, where 95°C can be reached inside aircraft equipement (for 70 to 75°C outside), and vibration, where 10 g's levels up to 2 KHz may be required for qualification.

10. USABLE TECHNOLOGIES

- TCXO's (Temperature Compensated Quartz Oscillators)

  TCXO's perform well concerning warm-up time and have some short term stability performance. However, their medium and long term stabilities are rarely better than 10^{-7}, and then high performance equipment requirements would not be fulfilled. They can however be used for some low-cost equipment.

- Standard OCXO's (Oven Controlled Quartz Oscillator)

  Standard OCXO's generally have some of the stability and phase noise characteristics required (except for the medium term stability requirements) but have poor Sg's and warm-up time. As the TCXO's, they can be used for low-cost equipment.

- High performance OCXO's

  Navstar GPS applications have generated the need for some developments among the OCXO's manufactuers, in order to improve warm-up time, g-sensitivity, as well as stability and dimensions. All the performance requirements given here can be nearly reached with these technologies : SC-cuts, BVA, QAS, vacuum-isolation, etc... The only remaining difficulties are the extreme medium term stability requirements, and of course the cost of these items.

- Atomic standards

  Atomic standards (e.g. Rubidium) can have all the characteristics detailed here, even for highest performance. In addition, the rubidium oscillators have been greatly miniaturized recently, and it begins to be possible to equip GPS receivers with them. However, the major remaining problems are still size, weight, power consumption, and cost.

Session PO-02 Number 13

MAGNETIC SENSITIVITY OF QUARTZ OSCILLATORS*

F. Deyzac

Office National d'Etudes et de Recherches Aérospatiales,
BP 72, 92322 Châtillon Cedex, France

INTRODUCTION

Quartz crystal oscillator are used on board satellites in order to provide frequency or time references that are very useful to the following space applications: localization, observation, telecommunication...

The increase of the required performances leads to take into account the influence of some parameters that were negligible until now, such as magnetism.

In fact, magnetic field fluctuations on board satellites may reach several times the level of the Earth magnetic field; this is in particular due to the use of magneto torquers for the attitude control.

The corresponding fractionnal frequency change may reach 10^{-10}. That represents more than one order of magnitude than the present specifications and leads to positionning error of 2 m by minutes.

PREVIOUS STUDIES

The possible magnetic sensitivity of quartz crystal resonator has been already pointed out in 1981 [1]. At this time, the magnetic field was not the major disturbance for USO_s and thus was negligible in comparison with the other parameters.

While CNES was introducing in 1986 new specifications on the USO magnetic sensitivity (10^{-11}/gauss), ONERA studied the same requirement on the atomic frequency standards [2].

In spite of the more stringent requirements (10^{-12}/gauss) [4], ONERA pointed out in 1988 the source of USO magnetic sensitivity and proposed how to deal with [3].

Next year, CNRS and CNES analyse the resonator sensitivity and evaluate its contribution to 10^{-12}/gauss [5].

QUARTZ CRYSTAL OSCILLATOR

A quartz crystal oscillator is a closed loop system (Fig. 1) in which the oscillation condition is:

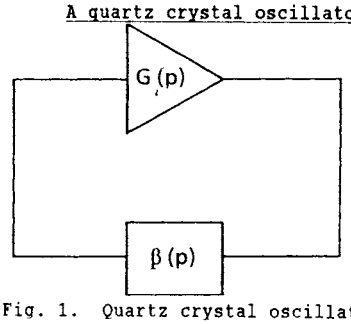


Fig. 1. Quartz crystal oscillator.

$$G(p) \cdot \beta(p) = 1 \quad \text{i.e.} \quad \begin{cases} |G(p) \cdot \beta(p)| = 1 \\ \text{and} \\ \arg(G(p)) + \arg(\beta(p)) = 0 \end{cases}$$

Quartz resonators are capable of resonating in many different modes and in their overtones. Thus it is necessary to insert into the loop (Fig. 2) a "mode selector" circuit in order to only conserve the chosen mode.

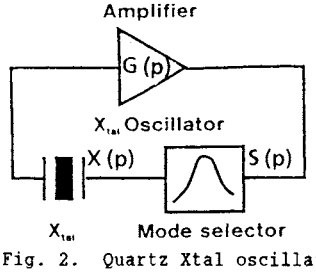


Fig. 2. Quartz Xtal oscillator with mode selector.

The phase condition becomes:

$$\arg G(p) + \arg X(p) + \arg S(p) = 0 \qquad (1)$$

where X denotes the quartz crystal and S the mode selector.

(*This work was performed under DRET contract)

MODE SELECTOR

The mode selector generally consists of a series tank inductor/capacitor. This circuit is used either in series or in parallel according to the following aims: selection of a mode (Fig. 3) or rejection of an other (Fig. 4).

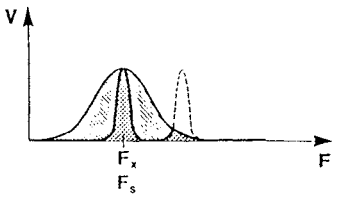


Fig. 3 Series mode selector.

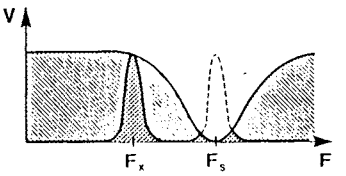


Fig. 4 Parallel mode selector.

As it is demonstrated hereafter, the oscillator frequency F_{USO} is defined like the mean value of the crystal frequency F_x and the selector frequency F_s, weighted by their respective quality factor Q_x and Q_s (Fig. 5).

A resonant circuit that is tuned to the frequency f exhibits at the frequency $f + \Delta f$ a phase rotation φ:

$$\varphi = \tan^{-1}\left(2Q\,\frac{\Delta f}{f}\right) \tag{2}$$

Q denoting the quality factor of the circuit.

When Δf is weak, the first approximation of (2) leads to:

$$\varphi \simeq 2Q\,\frac{\Delta f}{f} \tag{3}$$

the phase condition (1) can be written as follows for the phase variations:

$$\delta\varphi_x + \delta\varphi_s = 0$$

$$\text{i.e.} \quad \delta\left(2Q_x\,\frac{F_{USO} - F_x}{F_{USO}}\right) + \delta\left(2Q_x\,\frac{F_{USO} - F_s}{F_{USO}}\right) = 0$$

$$\text{and} \quad Q_x\,\delta(F_{USO} - F_x) + Q_s\,\delta(F_{USO} - F_s) = 0$$

that leads to:

$$\delta F_{USO} \simeq \delta F_x + \alpha\,\delta F_s \qquad \alpha = Q_s/Q_x \tag{4}$$

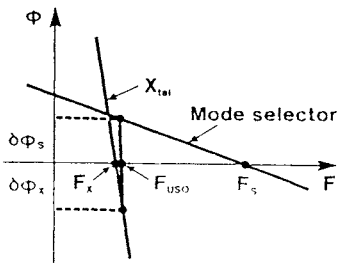


Fig. 5 Phase equilibrium for a quartz crystal oscillator with mode selector.

SERIES OR PARALLEL MODE SELECTOR

The approximation (3) of the phase relation (2) is only possible if Δf is close to null, i.e. F_s is close to F_{USO}; that is the case for series mode selector.

With the parallel mode selector, the tank circuit is tuned to a different frequency than F_{USO}. The variationnal equilibrium relating to the phase condition is now:

$$\delta\left(2Q_x\,\frac{F_{USO}-F_x}{F_{USO}}\right)+\delta\left[\tan^{-1}\left(2Q_s\,\frac{F_{USO}-F_s}{F_{USO}}\right)\right]=0$$

$$\text{i.e. } Q_x\,\delta(F_{USO}-F_x)+Q_s'\,\delta\,(F_{USO}-F_s)=0$$

$$\text{with } Q_s'=\beta\,Q_s \text{ and } \beta=\frac{1}{1+\left(2Q_s\,\frac{F_{USO}-F_s}{F_{USO}}\right)^2}$$

$$\text{and } \delta F_{USO}\simeq\delta F_x+\alpha\,\delta F_s\ ;\ \alpha=Q_s'/Q_x$$

NA: For a 10 MHz, third overtone SC cut crystal: $Q_x\simeq 2\cdot10^6$

- Series mode selector tuned to C mode - F_s = 10 MHz: $Q_s\simeq 20$; $\Rightarrow$ $\alpha\simeq 10^{-5}$,
- parallel mode selector tuned to B mode - $F_s\simeq 11$ MHz: $Q_s\simeq 20$; $\beta=1/17$ $\Rightarrow$ $\alpha\simeq 10^{-6}$.

But the use of parallel mode selector is not efficient when there are too many interfering modes.

MAGNETIC SENSITIVITY OF MODE SELECTOR

Experimental sensitivity

The magnetic characterization of 10 MHz LC tank circuit (with a 2.2 µH inductor and 100 pF inductor) exhibits (Fig. 6) a sensitivity of:

$$\frac{\Delta f}{f}\simeq -\,7.5\times10^{-6}\,(B-0.63)^2 \qquad (B \text{ in gauss})$$

when the field is applied along the inductor axis.

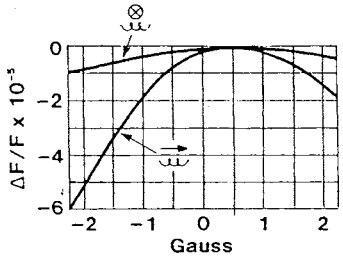


Fig. 6 Magnetic sensitivity of a 10 MHz LC oscillator (L = 2.2 µH, C = 100 pF).

Theoretical sensitivity

The commercial inductors are made with ferromagnetic cores (iron powder or ferrite) taking advantage of their higher permeability μ_r. But these materials present non linear behaviors versus the external applied magnetic field (Fig. 7) that modifies the frequency by:

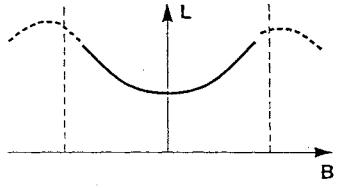


Fig. 7 Susceptibility of inductor with iron core.

$$\frac{\Delta f}{f}=\frac{1}{2}\,\frac{\Delta L}{L}.$$

USO MAGNETIC SENSITIVITY

The following experimental results are obtained with a spatial oscillator using a series mode selector and a 10 MHz SC P3 crystal.

The USO is lightly sensitive in the plane normal to inductor axis, and is very sensitive along this axis:

$$\text{(measured value)} \qquad \frac{\Delta F_{USO}}{F_{USO}}\simeq -\,7.5\times10^{-11}\,B^2 \qquad (B \text{ in gauss}).$$

Sensitivity due to mode selector:

- the magnetic sensitivity of the implemented mode selector (L = 2.2 µH and C = 100 pF) is:

(measured value) $\frac{\Delta F_s}{F_s} \simeq -7.5 \times 10^{-6} B^2$ (B in gauss).

- the impact on the USO frequency according to the theoretical formula $\left(\frac{\Delta F_{uso}}{F_{uso}} \simeq \alpha \frac{\Delta F_s}{F_s} \text{ with } \alpha \simeq 10^{-5}\right)$ is:

$$\frac{\Delta F_{uso}}{F_{uso}} \simeq -7.5 \times 10^{-11} B^2 \quad \text{(B in gauss)}.$$

This result is in very good agreement with the measured value and the order of magnitude estimated for the value of α.

SOLUTION

In order to confirm this approach, the same space USO has been equiped with non ferromagnetic inductor and the magnetic sensitivity of this USO has been measured in the ONERA laboratory.

The sensitivity is reduced from $4 \times 10^{-10}/2$ gauss with a "ferro" inductor to $0.05 \times 10^{-10}/2$ gauss with a "non ferro" inductor.

Fig. 8 Magnetic sensitivity of USO using "ferro" or "non ferro" inductor.

The quadratic sensitivity is become linear in particular when the magnetic field is applied along quartz Y axis:

$$\frac{\Delta f}{f} \simeq 1 \times 10^{-12} B \quad \text{(B in gauss)}.$$

CONCLUSION

This interesting result gives on a first step a low cost solution for the magnetic sensitivity of the space USO.
Moreover, it is pointing out that the weak contribution of the mode selector in the value of the USO frequency (10 ppm) can be sufficient to limit the USO performances: even if the mode selector characteristics do not vary of more than 1 ppm, the stringent USO specifications of 10^{-11} are not met.
The instability of the mode selector must be re-analyzed and its contribution re-evaluated for the noise and the aging of the oscillator in particular in presence of temperature and humidity perturbations.

REFERENCES

[1] - J.J. Gagnepain, J.G. Theobald
Influence d'un champ magnétique sur la fréquence propre d'un résonateur à quartz.
C.R. Acad. Sci., Paris, Vol. 292, série II, Janvier 1981.

[2] - F. Deyzac
Caractérisation en environnement perturbé des horloges atomiques au césium.
L'Onde Electrique, Vol. 66, n° 6, Novembre 1986.

[3] - F. Deyzac
Oscillateurs à quartz de bord en environnement severe. Sensibilité magnétique.
ONERA Technical Report n° 37/7132 PY, 1988.

[4] - J.B. Laporte
Specifications and performances of ultra-stable oscillators in the DORIS orbitography and positionning system.
Proceedings 2nd European Frequency and Time Forum, Neuchâtel, 1988.

[5] - R. Brendel, C. El Hassan, M. Brunet, E. Robert
Influence of magnetic field on quartz crystal oscillators.
Proceedings 43rd Annual Symposium on Frequency Control, 1989.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session PO-02 Number 14

LINEAR PHASE FILTERING

B. d'ALBARET - L. BIDART - G. SILLIOC

Q K PIEZOELECTRONIQUE - 30 Boulevard Galliéni -F92390 VILLENEUVE LA GARENNE

SUMMARY

Digital communication processing is placing the importance of the linearity information in a prominent position.

Today reliability as well as the rapidity of transmissions widely depend on the phase linearity or on the constant group delay in filtering systems.

Concerning classical filters, it appears that linear phase is incompatible with selectivity. Most of the time, it is necessary to find the best compromise between these requirements.

INTRODUCTION

This presentation draws up the inventory of the different types of linear phase Xtal filters and the ways in which they can improve, control, correct the phase linearity so that they obtain a group delay adapted to needs of up to date Electronic.

Theorical possibilities and practical limitations afforded by different polynomial or elliptical filters will be discussed as well as the correction which can be made by equalizers.

COMPARISON BETWEEN DIFFERENT TYPES OF FILTERS

With classical approximations it appears that linear phase is incompatible with selectivity. In order to compare the different approximations, we have chosen a 6 poles Xtal filter with 5 types of transfer function :

1. Butterworth
2. Tchebycheff with 0.1 dB of ripple
3. Bessel
4. Linear phase with equiripple 0.5°
5. Semi linear

Fig. 1. shows the 5 types of attenuation.

Fig. 2. shows the 5 types of bandwidth and group delay.

The following table compares the different attenuations and group delay variations :

Function type	K 60/3	$d\tau/\tau$ 60%	$d\tau/\tau$ 100%
Butterworth	3.2	20	66
Tchebycheff 0.1 dB	2.3	25	150
Bessel	5.4	0	0
Linear equiripple 0.5°	4.4	4	4
Semi linear	2.6	2	125

K is the shape factor, i.e. the ratio : 60 dB BW/3 dB BW

$d\tau/\tau$ is the relative variation of the groupe delay for 60% or 100% of the 3 dB bandwidth.

If we observe the 4 first lines, we constate that the shape factor is bad when the group delay variation is good, so it was necessary to find a solution between both.

COMPROMISE

In order to preserve a constant groupe delay in part of the bandwidth, while preserving relatively high selectivity, special transfer function have been developed : these filters called semi linear filters present a constant group delay in part of the 3 dB bandwidth, this is called the usefull bandwidth. It is like a Bessel filter for this part of band, but with an out of band attenuation like a Butterworth filter.

As we can see on the table, the "semi linear" filter is sharper than the other 2 linear filters. Its group delay is completely constant on the usefull bandwidth (60% of the total bandwidth).

Fig. 3 shows the real curves of a 6 poles 90 MHz Bessel filter usering 6 quartz resonators. The groupe delay is constant over all the 3 dB BW $\pm$ 14 KHz ; we observe a variation of 3 μs for an absolute group delay of 32 μs, i.e. 10%.

Fig. 4 presents the practical realization.

Fig. 5 shows a "semi linear" 6 poles filter at 21.380 MHz (6 dB BW $\geq \pm$ 13 KHz) using 3 monolithic resonators. The group delay is constant over 60% of the bandwidth : $d\tau / \tau \leq 3$ μs over $\pm$ 8.5 KHz i.e. a variation $\leq$ 5%. Fig. 5 and 6 show a shape factor 60/6 dB of 2.3.

FEASIBILITY

For a Xtal filter, the maximum realizable relative bandwidth is approximatively for quartz : 0.3 % But this value is only available for non linear Filters.

For a linear filter, it is necessary to take into account twice the real bandwidth (in fact the resonators frequencies are twice far apart as usually).

Therefore, filters which were realizable in quartz with a Butterworth or Tchebycheff function are no longer feasable with linear phase function if the relative bandwidth is greater than 0.15%.

In this case, the Lithium Tantalate can be used. It allows us to go up to 5% of relative bandwidth (2.5% for linear filters).

EQUALIZERS

Sometimes, attenuation conditions are so hard that it is impossible to use a traditional function. A typical application is the Single Side Band Filter where 2 schemes are employed :

- a scheme in ladder with resonators in series and in parallel (see Fig. 7).
- a scheme in bridge with several resonators per arm (see fig. 8).

All these filters present several poles of infinite attenuation and are very sharp, i.e. a shape factor K 60/3 dB less than 1.5. Of course, for such filters, the groupe delay is basically distorted.

For certain applications, it is in fact necessary to obtain a relatively constant group delay.

In this case, we use a second filter called equalizer which follows the first. It presents a very flat attenuation in the bandwidth in order to preserve the attenuation curve of the first filter.

Furthermore it has a group delay opposite to the first in order to obtain by the sum the flattest possible variation of group delay.

The typical curves of attenuation and group delay of a network "pass all" , i.e. a network which changes the phase but not the attenuation, with the use of Xtal resonators is given Fig. 9.

To obtain such a network, it is necessary to make the synthesis of a non minimum phase function so called, if it has zeros in the right half part of the complex plan.

A synthesis program gives us all the values of the elements for a bridge network (Jaumann scheme) with 2 Xtals in parallel in one arm (see Fig. 10).

If the correction of the group delay is not enough sufficient, of course, 2 cells in series can be used.

Ex. 1. : Picture 11 shows a 10 MHz filter S.S.B. made with a ladder scheme using 13 quartz resonators.

The specification was 3 dB BW $\geq \pm$ 6 KHz
Group delay variation in the BW $\pm$ 5 KHz $\leq$ 200 μs.

We can observe on the curve 1 of Fig. 12 and 13 the theoritical simulation. The variation of the group delay makes 280 μs, therefore it is necessary to correct a minimum of 100 μs with an equalizer.

One cell has been calculated and the theoritical curve can be seen on curve 2 of Fig.13. Curve 3 shows the results of the two filters in series after simulation on computer.

Fig. 14 and 15 show the practical results with and without equalizer that the specification achieves.

Ex. 2 : Fig. 16 presents the scheme of a 10 MHz filter calculated with a Cauer synthesis Jaumann type using 6 $LiTaO_3$ resonators. In fact, the 3 dB bandwidth of $\pm$ 50 KHz was too wide for quartz.

The specification for the group delay was :

Maximum variation over $\pm$ 45 KHz : 2 μs.

Curve 1 of Fig. 17 and 18 shows the theoritical simulation of the filter which makes 5 μs of variation at extremities of the bandwidth.
An equalizer has been calculated with a correction of 5 μs (Curve 2 of Fig. 18).

Curve 3 shows the result of the two filters in series, the group delay variation is less than 1.5 μs.

CONCLUSION

In this presentation, we have considered in detail the three possibilities which exists to obtain a filter with a good phase linearity i.e. with a constant group delay :

- filters linear in phase
- Filters "semi linear" in phase
- Filters with equalizer

At the time of the design of such filter, it is necessary to know which performances are the most usefull in order to make the right choice.

Following table establishes the inventory of the performances :

FILTER TYPE	Phase linerity or constant group delay	Selectivity or shape factor	Difficulty of design
Phase Linear	Very good over 100% of BW	Bad	Medium
Semi linear	Very good over 60% of BW	Good	Medium
Filter with Equalizer	Medium good over 100% of BW	Very Good	Difficult

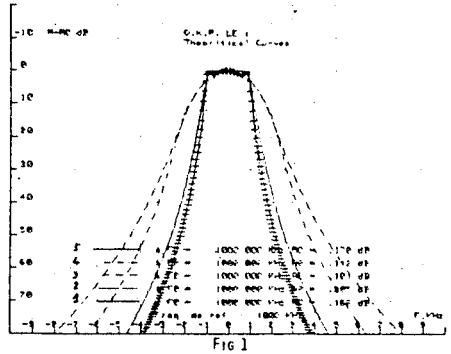


FIG 1

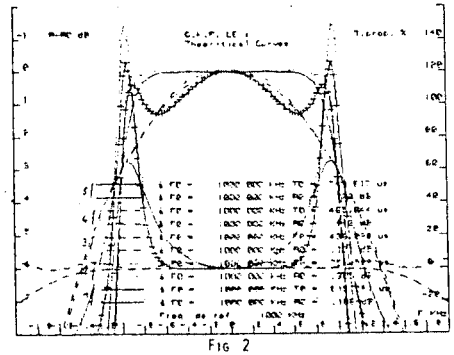


FIG 2

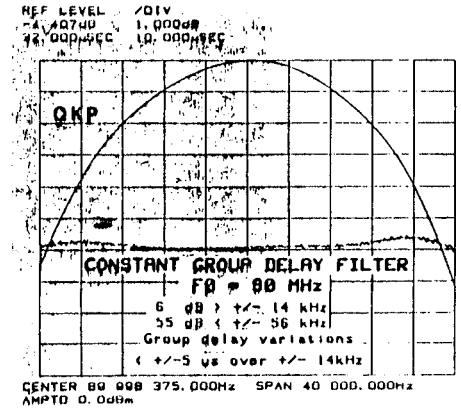


FIG 3

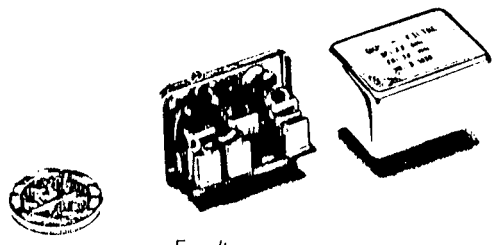


FIG 4

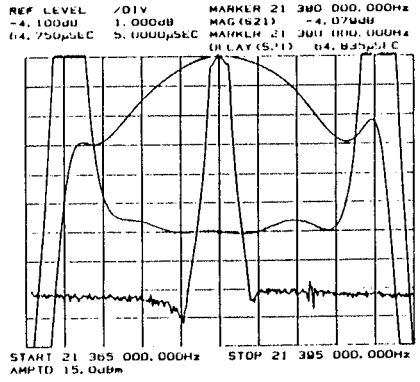


FIG 5

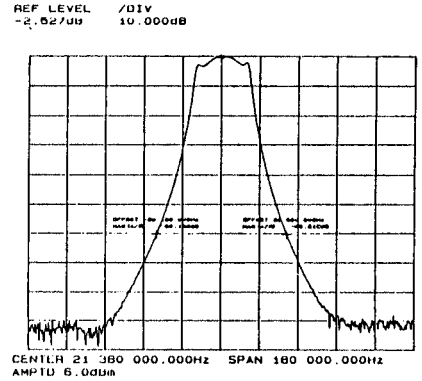


FIG 6

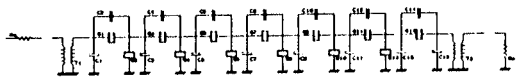


FIG 7

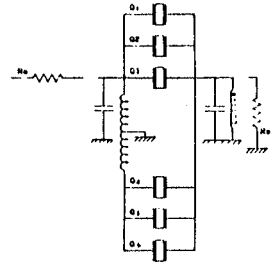


FIG 8

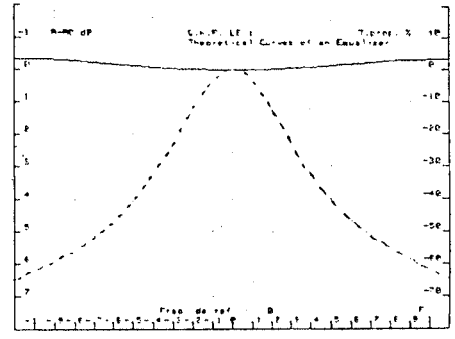


FIG 9

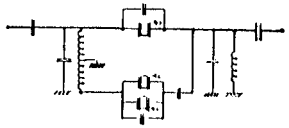


FIG 10

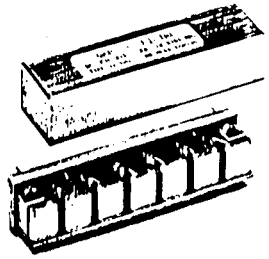


FIG 11

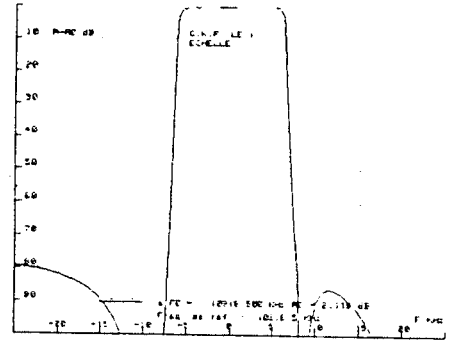


FIG 12

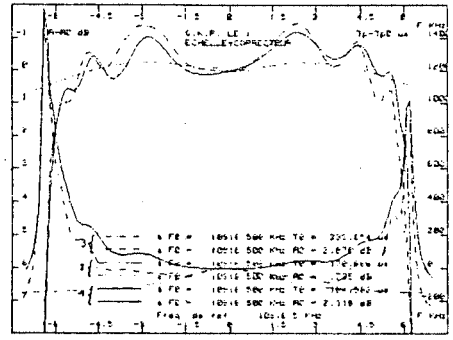


FIG 13

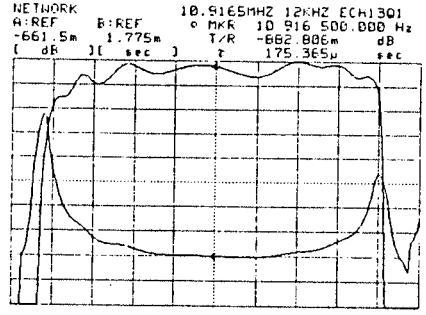


FIG 14

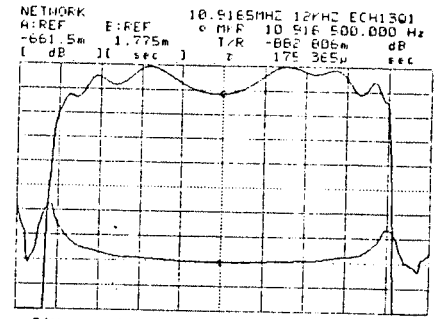


FIG 15

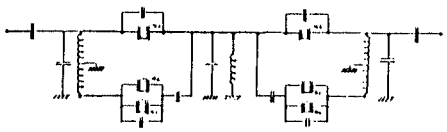


FIG 16

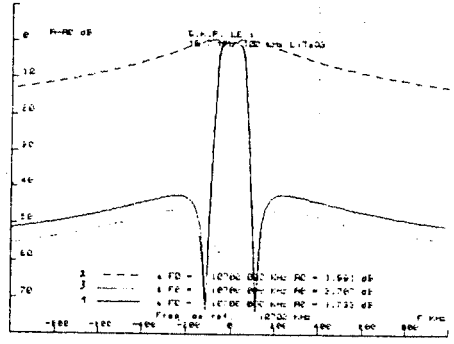


FIG 17

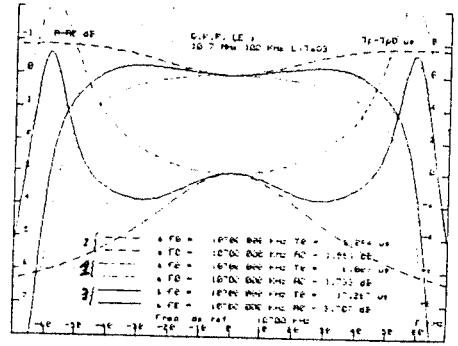


FIG 18

Frequency Standards

FS-03 Future Prospects

Chairman: H. Hellwig

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-03 Number 1

PROGRESS AT NIST ON ABSOLUTE FREQUENCY STANDARDS USING STORED IONS*

D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano,
D. J. Heinzen, S. L. Gilbert, C. H. Manney, M. G. Raizen, and C.S. Weimer
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado 80303

Abstract

Experiments at NIST, whose goal is to realize frequency standards of high accuracy using stored ions, are briefly summarized. In one experiment, an rf oscillator is locked to a nuclear spin-flip hyperfine transition (frequency $\simeq 3.03 \times 10^8$ Hz) in $^9Be^+$ ions which are stored in a Penning trap and sympathetically laser-cooled. Stability is better than $3 \times 10^{-12}\tau^{-\frac{1}{2}}$ and uncertainty in Doppler shifts is estimated to be less than 5×10^{-15}. In a second experiment, a stable laser is used to probe an electric quadrupole transition (frequency $\simeq 1.07 \times 10^{15}$ Hz) in a single laser-cooled $^{199}Hg^+$ ion stored in a Paul trap. The measured Q value of this transition is approximately 10^{13}.

Introduction

One goal of the ion storage group at NIST (formerly the National Bureau of Standards) has been to realize a stored ion frequency standard whose absolute inaccuracy is less than 1 part in 10^{14}. Primary design considerations in these experiments have been influenced by the assumptions that (1) for many ions in the trap, the uncertainty in the measurement of the second-order Doppler shift will ultimately be the largest contribution to inaccuracy and (2) the magnitude of the second-order Doppler shift decreases as the number of ions in a particular trap decreases.

*Contribution of the U.S. Government; not subject to copyright.

These assumptions are supported both by experiments and theoretical analysis [1-6]. For a given number of ions N in the trap the second-order Doppler shift is minimized when the secular motion in the Paul trap [2-5] or the cyclotron and axial motion in the Penning trap [1-3] are reduced by some means to negligible values (for example, by the use of buffer gas collisions [4] or laser cooling [6,7]). For the Paul trap, the second-order Doppler shift is then dominated by the velocity in the rf micromotion [1-5]. The basic idea is as follows: in the quadrupole Paul trap the rf micromotion velocity increases with the distance of the ion from the center of the trap. As the number of trapped ions increases, space charge repulsion holds ions farther from the center of the trap thereby increasing the ion's micromotion speed and second-order Doppler shift. For a spherical cloud of ions in a quadrupole Paul trap, the number of ions N is proportional to the fractional second-order Doppler shift, averaged over the cloud, $\langle\Delta\nu_{D2}/\nu_o\rangle$ [2-4]:

$$N = -2.16 \times 10^{16} r_{c\ell} M \langle\Delta\nu_{D2}/\nu_o\rangle / Z^2 \quad \text{(rf trap)} \qquad (1)$$

where $r_{c\ell}$ is the cloud radius in centimeters, M is the ion mass in atomic mass units and Z is ion charge in units of the proton charge. The number of ions for a given second-order Doppler shift can be increased by using a nonspherical ion cloud geometry in an elongated trap [5]. However, the basic idea still holds that as the number of ions increases so does $|\langle\Delta\nu_{D2}/\nu_o\rangle|$. The second order Doppler shift is minimized for a single ion [6] or for more than one ion in a trap with linear geometry.

In the Penning trap, when the cyclotron and axial kinetic energies are reduced to small values, the second order Doppler shift is dominated by the velocity in the rotation of the ion cloud. In this case, N and $\langle\Delta\nu_{D2}/\nu_o\rangle$ are related by [1-3]

$$N = 3.10 \times 10^{13}\ B \langle -\Delta\nu_{D2}/\nu_o\rangle^{\frac{1}{2}}\ z_{c\ell}(r_{c\ell} - r_c')/Z \quad \text{(Penning trap)} \qquad (2)$$

where B is the trap magnetic field strength in teslas, $2z_{c\ell}$ and $r_{c\ell}$ are the cloud height and radius in centimeters, $r_c' = (5\langle -\Delta\nu_{D2}/\nu_o\rangle)^{\frac{1}{2}} c/\Omega_c$ (Ω_c is the ion's cyclotron frequency and c is the speed of light), and Z is the ion charge in units of proton charge. Equations (1) and (2) are valid for one species of ion in the trap. For the special case where the ions are confined along the z axis, Eq. 2 no longer applies and the second order Doppler shift is given by the motion in the cyclotron and axial degrees of freedom.

Therefore, for both traps, $|\langle\Delta\nu_{D2}/\nu_o\rangle|$ typically increases as N increases and we are faced with a compromise in design. For good signal-to-noise ratio and therefore good frequency stability, we desire large N. However this increases $|\langle\Delta\nu_{D2}/\nu_o\rangle|$ and therefore decreases accuracy because of our inability to measure precisely the velocity distributions needed to determine $\langle\Delta\nu_{D2}/\nu_o\rangle$. This trade-off between stability and accuracy has resulted in different approaches. In the work of one group[4,8], a stored $^{199}Hg^+$ ion standard with excellent stability has been realized. In these experiments $N \simeq 2 \times 10^6$ and $\langle\Delta\nu_{D2}/\nu_o\rangle \simeq$

-2×10^{-12}, so an accuracy of 10^{-14} would require a knowledge of $\langle v^2 \rangle$ to 0.5% precision. This problem can be reduced in the elongated trap geometry of [5] but, independent of the trap geometry, the accuracy can typically be improved by using smaller numbers.

At NIST the primary goal of stored-ion frequency standard work has been high accuracy. The preceding arguments have forced us to use small numbers of ions (approximately 10^4 or less) which are laser-cooled. The loss in short-term stability due to reduced numbers can be partially regained by going to very long interrogation times and high detection efficiency. These trade-offs are apparent from the expression for stability if we assume the Ramsey method of interrogation, 100% detection efficiency and zero dead time between Ramsey periods. For these conditions [3],

$$\sigma_y(\tau) = (\tau N T_R \omega_o^2)^{-\frac{1}{2}} \quad (\tau > T_R) \tag{3}$$

where τ is the averaging time, T_R is the Ramsey interrogation time and ω_o is the clock transition frequency (in radians per second). From this expression, we also see the importance of making ω_o large.

Of course, to achieve high accuracy, we must also account for the perturbations due to static and time-varying multipole interactions for electric, magnetic, and gravitational fields. These include atom-trap field interactions, collisions, shifts due to uncontrolled electric and magnetic fields, and gravitational red shifts. However for more than one ion in the trap, it appears that the dominant uncertainty will be caused by the uncertainty in the second-order Doppler shift.

Below, we briefly discuss experiments at NIST on stored-ion frequency standards. More detailed accounts are forthcoming.

^{9}Be$^+$ Hyperfine Clock

In this experiment, an oscillator has been locked to the (m_I = -1/2, m_J = 1/2) $\leftrightarrow$ (-3/2, 1/2) nuclear spin-flip hyperfine "clock" transition ($\omega_o/2\pi \simeq$ 303 MHz) in the ground state of ^{9}Be$^+$ (Fig. 1). The basic idea of this experiment has been described previously [9-12]; the current experiment works as follows [12]: Between 5000 and 10 000 ^{9}Be$^+$ ions and 50 000 to 150 000 ^{26}Mg$^+$ ions were simultaneously stored in a cylindrical Penning trap [8] with B $\simeq$ 0.8194 T under conditions of high vacuum ($\lesssim 10^{-8}$ Pa). At a magnetic field B of 0.8194 T the clock transition depends only quadratically on magnetic field fluctuations, and therefore the accuracy is not limited by field fluctuations. To minimize second order Doppler shifts of the clock transition, the ^{9}Be$^+$ ions were cooled to less than 250 mK in the following manner: The ^{26}Mg$^+$ ions were directly laser-cooled and compressed by a narrow-band (~ 1 MHz) laser radiation source at 280 nm. The ^{9}Be$^+$ ions were then sympathetically cooled [13] by their Coulomb interaction with the cold Mg$^+$ ions (see Appendix A). A narrow-band 313 nm radiation source was used to optically pump and detect the ^{9}Be$^+$ ions [9-11]. With the 313 nm source tuned to the 2s $^2S_{\frac{1}{2}}$ (m_I = 3/2, m_J = 1/2) to 2p $^2P_{3/2}$(3/2, 3/2) transition, 94% ot the ^{9}Be$^+$ ions were optically pumped into the 2s $^2S_{\frac{1}{2}}$(3/2, 1/2) ground state. The 313 nm source was then turned off to avoid optical pumping and ac Stark shifts while the clock transition was driven.

The clock transition was detected by the following method: After the 313 nm source was turned off, the ions in the (3/2, 1/2) state were transferred to the (1/2, 1/2) state and then to the (-1/2, 1/2) state by two successive rf π pulses. Each pulse was 0.2 s long and was resonant with the appropriate transition frequency (around 321 MHz and 311 MHz respectively). The clock transition was then driven by Ramsey's method of separated oscillatory fields in the time domain with rf pulses of about 1 s duration and a free precession time on the order of 100 s. This transferred some of the ions from the (-1/2, 1/2) state to the (-3/2, 1/2) state. Those ions remaining in the (-1/2, 1/2) state were then transferred back to the (3/2, 1/2) state by reversing the order of the two rf π pulses. The 313 nm source was then turned back on, and the population of ions in the (-3/2, 1/2) state was registered as a decrease in the ^{9}Be$^+$ fluorescence, relative to the steady-state fluorescence, during the first second that the 313 nm source was on. (The optical repumping time of the ions from the (-3/2, 1/2) state to the (3/2, 1/2) state was about 10 s.) The sympathetic cooling of the ^{9}Be$^+$ ions by the continuously cooled Mg$^+$ ions is necessary if long interrogation times are to be used, since otherwise the ^{9}Be$^+$ ions would slowly heat up while the 313 nm laser is off.

The Ramsey signal was used to steer the frequency of a synthesized rf source [9-12]. Ramsey signal measurements were taken near the frequencies corresponding to the half minimum points on both sides of the center frequency. The difference in the measured signal strengths on either side of the line center was used to electronically steer the average frequency of the synthesizer to ω_o. Most runs were taken with a commercial cesium beam clock (fractional frequency stability $\sigma_y(\tau) \simeq 6 \times 10^{-12}\tau^{-\frac{1}{2}}$ for measurement time τ in seconds) as the reference oscillator, but a few runs were taken with a passive hydrogen maser ($\sigma_y(\tau) \simeq$ $2\text{-}3 \times 10^{-12}\tau^{-\frac{1}{2}}$) as the reference oscillator. Stabilities of the ^{9}Be$^+$ clock are measured to be better than $3 \times 10^{-12}\tau^{-\frac{1}{2}}$ for the number of ions used, which is within a factor of 4 of the theoretical maximium stability given by Eq.3. The systematic error of our measurement of ω_o due to uncertainty in the second-order Doppler frequency shift is about 5×10^{-15}.

An apparent pressure shift with an unexpectedly large value was discovered when the background gas pressure was increased. The background gas pressure could be increased by moving the magnet of the sputter ion pump which evacuated the trap region so that it overlapped fewer pumping cells and reduced the pumping speed. (We checked to make sure the magnetic field at the site of the ions was not disturbed.) The composition of the gas was not known since the pressure was measured with a Bayard-Alpert gauge. However, when the vacuum vessel containing the trap was initially evacuated the dominant background gases were H_2 and He as determined by a quadrupole mass analyzer. Therefore we expect that the background gas during the frequency standard measurements was either H_2 or He, or both. If the background gas was dominated by He, the fractional pressure shift was measured to be about $-3\text{x}10^{-6}\text{Pa}^{-1}$; if the background was dominated by H_2, the pressure shift was measured to be about $-9\text{x}10^{-6}\text{Pa}^{-1}$. Atomic ion hyperfine pressure shifts have previously been measured in ^{137}Ba$^+$ [14] and in ^{199}Hg$^+$ [8] to be $5\text{x}10^{-11}\text{Pa}^{-1}$ and $4\text{x}10^{-11}\text{Pa}^{-1}$ respectively. The authors of ref. 14 show that the charge induced multipole interaction between the Ba$^+$ and the noble gas

atoms used in that study should give an important contribution to the pressure shift. Since this interaction depends primarily on the polarizability of the neutral we would expect that the pressure shift for He atoms on $^9Be^+$ ions would not be significantly different than that for Ba^+ or Hg^+. Similarly, since the polarizability of H_2 is midway between Ar and Ne, we might expect the pressure shift for H_2 on Be^+ to be near those for Ar and Ne on Ba^+, which were measured to be $-6x10^{-10}Pa^{-1}$ and $-6x10^{-9}Pa^{-1}$ respectively.

The apparent large discrepancy between our data and other measured pressure shifts is not understood at this time. One possible explanation is suggested by studies [15] of radiative association of C^+ with H_2 to form CH_2^+. In the models of this process, it is assumed that the H_2 can stick to the C^+ for a long enough time to allow the C^+-H_2 complex to radiatively stabilize. This sticking is possible because the collision energy can be taken up by the internal degrees of freedom in the H_2 molecule. The sticking time can be orders of magnitude longer than the interaction time during a simple elastic collision. If a similar sticking mechanism is active in H_2 - Be^+ collisions, it may account for the apparent large pressure shift. One way to check this hypothesis is to measure the pressure shifts of Be^+ on Ne and Ar, where the sticking would not be expected to occur, and on H_2, where sticking might occur.

We are continuing to search for other causes of systematic errors such as electronic offsets; however, we feel the accuracy could be improved beyond 5 parts in 10^{15} in the future. If the apparent collisional shift is real, it may be necessary to use liquid He cryopumping to reduce the background pressure.

$^{199}Hg^+$ Optical Clock

The velocity in the micromotion for an ion in a quadrupole rf trap is proportional to the distance of the ion from the center of the trap. For two or more laser-cooled ions in the trap, the Coulomb repulsion between ions holds them away from the trap center, and the second-order Doppler shift is dominated by the velocity of micromotion. However, a single ion can be held near the trap center if sufficiently cooled. In this case the kinetic energy in the micromotion can be equal to that of the secular motion [2,3,16,17]. If the ion is laser-cooled, resulting Doppler shifts can be extremely small; uncertainties can be less than 1 part in 10^{20} in some cases [17,18]. However with N = 1, stability is marginal unless we make ω_o high enough (see Eq.3). One way to accomplish this is to let ω_o correspond to an optical transition. Dehmelt suggested this idea in 1973 [6]. The reasons that a clock based on an optical transition in an ion has not been realized yet are: (1) it took several years to isolate and manipulate single ions in the traps, (2) local oscillators (lasers) with the desired spectral purity are still not available, and (3) accurate comparison of laser and microwave frequencies is extremely difficult and remains an important problem. Nevertheless the potential accuracy of single-ion optical frequency standards is extremely high [6,19] ($\simeq 10^{-18}$) so it is important to pursue this research.

At NIST we have investigated the use of the $5d^{10}6s\ ^2S_{1/2} \rightarrow 5d^9 6s^2\ ^2D_{5/2}$ electric quadrupole transition ($\omega_o/2\pi \simeq 1.07 \times 10^{15}$ Hz) in $^{198}Hg^+$ (see Fig. 2) as an optical frequency standard [20]. The single mercury ion is confined in a miniature rf trap that has internal dimensions of $r_o \simeq 466$ μm and $z_o \simeq 330$ μm [20,21]. The amplitude of the trapping field (frequency $\Omega/2\pi \simeq$ 21-23 MHz) can be varied to a peak of 1.2 kV. The ion is laser-cooled to a few millikelvins by a few microwatts of cw laser radiation that is tuned below the $^2S_{1/2}$ - $^2P_{1/2}$ first resonance line near 194 nm. In order to cool all motional degrees of freedom to near the Doppler cooling limit [7] ($T = \hbar\gamma/2k_B \simeq 1.7$ mK) the 194 nm radiation irradiates the ion from 2 orthogonal directions, both of which are at an angle of 55° with respect to the symmetry (z) axis of the trap. The 282 nm radiation that drives the narrow $^2S_{1/2}$ - $^2D_{5/2}$ transition is obtained by frequency-doubling the radiation from a narrowband cw ring dye laser. The frequency of the laser is stabilized by locking it to a stable Fabry-Perot cavity. The frequency of the laser is scanned by an acousto-optic modulator that is driven by a computer controlled synthesizer. Up to a few microwatts of 282 nm radiation could be focussed onto the ion in a direction counterpropagating with one of the 194 nm light beams.

Optical-optical double resonance was used to detect transitions driven by the 282 nm laser from the ion's $^2S_{1/2}$ ground state to the metastable $^2D_{5/2}$ state [19]. The 194 nm fluorescence rate from the laser-cooled ion is high when the ion is cycling between the $^2S_{1/2}$ and $^2P_{1/2}$ states (Fig. 2) and nearly zero when the ion is in the metastable $^2D_{5/2}$ state. The $^2S_{1/2}$ - $^2D_{5/2}$ resonance spectrum was obtained by probing the S-D transition at a particular frequency for the 282 nm radiation for 20 ms, then turning off the 282 nm radiation and turning on the 194 nm radiation to look for the presence or absence of scattered photons at 194 nm. (The two radiation fields are alternately applied to avoid light shifts and broadening of the narrow S-D transition by the 194 nm radiation.) If there was no fluorescence at 194 nm, a transition into the metastable D state had occurred; the presence of 194 nm fluorescence indicated that the ion was in the ground state and no transition was recorded for this frequency of the 282 nm laser. The frequency of the 282 nm radiation was then stepped and the measurement cycle repeated. As the frequency was swept back and forth each new result at a particular frequency of the 282 nm radiation was averaged with the previous measurements at that frequency. Normalization (or digitization) of the signal was obtained by assigning a 1 to each measurement of high fluorescence and a 0 to each measurement of no fluorescence. The high fluorescence level made it possible to determine the state of the atom with almost no ambiguity in a few milliseconds. Thus, it is possible to reach the shot noise limit imposed by the single atomic absorber [21].

The quantized fluorescence signal obtained from an 8 MHz scan of the 282 nm laser through the $^2S_{1/2}(m_J = -1/2) \rightarrow {}^2D_{5/2}(m_J = 1/2)$ Zeeman component of the electric quadrupole transition in $^{198}Hg^+$ is shown in Fig. 3. The recoilless-absorption resonance (carrier) and the motional sidebands due to the secular motion in the harmonic well of the rf trap are completely resolved [21].

To avoid broadening of the quadrupole transition due to magnetic field fluctuations, we have recently[22] performed the same experiment on the $^2S_{1/2}(F=0, m_F=0) \rightarrow {}^2D_{5/2}(F=2, M_F=0)$ transition in $^{199}Hg^+$ which becomes field independent as $B \rightarrow 0$. The carrier is now observed with a linewidth $\Delta\nu \leq 100$ Hz (limited by laser spectral purity), which gives a line Q of about 10^{13}, the highest reported in atomic or molecular spectroscopy. Current efforts are devoted to improving the 282 nm laser spectral purity by locking it to a

more stable reference cavity. If the laser's spectral purity can be made high enough, then when the laser is locked to the ion transition, stabilities are anticipated to be better than $10^{-15}\tau^{-\frac{1}{2}}$ and accuracies could be 1 part in 10^{18} or better [1,3,17-21].

Conclusion

Progress on two stored ion frequency standard experiments at the National Institute of Standards and Technology has been reviewed. Although current results indicate that accuracies of 1 part in 10^{14} or better will eventually be achieved, certain technical problems must be overcome. An apparent pressure shift discovered in the $^{9}Be^{+}$ ion experiments must be further investigated and eliminated. Unless it is an effect unique to $^{9}Be^{+}$, it may be an important systematic effect to consider in other stored ion frequency standard experiments.

In the Hg^{+} optical frequency standard experiments, the limitation to resolution is caused by the spectral impurity of the laser local oscillator. Efforts are currently devoted to better mechanical isolation of the optical reference cavity to which the local oscillator is locked.

Some future experimental possibilities at NIST are discussed in Ref. 23. A cylindrical column of sympathetically cooled ions in a Penning trap, as was the case for the $^{9}Be^{+}$ experiments, may provide high accuracy for certain ion species. For an elongated trap geometry where the ions are confined near the z axis [5,24], the second order Doppler shift can be minimized while using many ions, therefore increasing signal-to-noise ratio and short term stability. At NIST, a linear rf trap for $^{199}Hg^{+}$ ions is currently being studied. This apparatus can be used for a clock based on the 40.5 GHz hyperfine transition [4,8] and the 1.07×10^{15} Hz quadrupole transition.

Acknowledgements

We gratefully acknowledge the continued support of the Office of Naval Research and the Air Force Office of Scientific Research. We thank R. Drullinger, H. Robinson and M. Young for helpful comments on the manuscript.

References

[1] D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, Proc. 35th Ann. Symp. Freq. Control, Philadelphia, PA, May 1981, p. 602-611. (Copies available from: Annual Frequency Control Symposium, c/o Electronic Industries Assoc., 2001 Eye St., Washington, D. C., 20006.)

[2] D. J. Wineland, Proc. Conf. on Precision Measurements and Fundamental Constants; Gaithersburg, MD, June 1981; Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, eds. National Bureau of Standards (US) Spec. Publ. 617 (1984), pp. 83-92.

[3] D. J. Wineland, W. M. Itano, J. C. Bergquist, J. J. Bollinger, F. Diedrich, and S. L. Gilbert, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 71.

[4] L. S. Cutler, R. P. Giffard and M. D. McGuire, Appl. Phys. B36, 137 (1985); L. S. Cutler, C. A. Flory, R. P. Giffard, and M. D. McGuire, Appl. Phys. B39, 251 (1986); H. Schaaf, V. Schmeling, and G. Werth, Appl. Phys. 25, 249 (1981).

[5] J. Prestage, G. J. Dick and L. Maleki, J. Applied Physics, 66, 1013 (1989).

[6] H. G. Dehmelt, IEEE Trans. Instrum. Meas. IM-31, 83 (1982), and references therein.

[7] D. J. Wineland and W. M. Itano, Physics Today, vol. 40, no. 6, p. 34 (1987).

[8] L. S. Cutler, R. P. Giffard, P. J. Wheeler and G. M. R. Winkler, Proc. 41st Ann. Symp. Freq. Control, 1983, (copies available from IEEE, 445 Hoes Lane, Piscataway, NJ 08854), p. 12.

[9] J. J. Bollinger, S. L. Gilbert, W. M. Itano and D. J. Wineland, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 319.

[10] J. J. Bollinger, W. M. Itano, and D. J. Wineland, Proc. 37th Ann. Symp. Freq. Control, 1983, p. 37-41 (copies available from IEEE, 445 Hoes Lane, Piscataway, NJ 08854.)

[11] J. J. Bollinger, J. D. Prestage, W. M. Itano and D. J. Wineland, Phys. Rev. Lett. 54, 1000 (1985).

[12] J. J. Bollinger, D. J. Heinzen, W. M. Itano, S. L. Gilbert and D. J. Wineland, Phys. Rev. Lett. 63, 1031 (1989).

[13] D. J. Larson, J. C. Bergquist, J. J. Bollinger, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. 57, 70 (1986).

[14] J. Vetter, M. Stuke, and E. W. Weber, Z. Phys. A273, 129 (1975).

[15] E. Herbst, Astrophys. J. 252, 810 (1982), and references therein.

[16] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1967).

[17] D. J. Wineland, W. M. Itano, J. C. Bergquist and R. G. Hulet, Phys. Rev. A36, 2220 (1987).

[18] F. Diedrich, J. C. Bergquist, W. M. Itano and D. J. Wineland, Phys. Rev. Lett. 62, 403 (1989).

[19] D. J. Wineland, Science 226, 395 (1985).

[20] J. C. Bergquist, F. Diedrich, W. M. Itano, and D. J. Wineland, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy (Springer-Verlag, Berlin, Heidelberg) p. 287.

[21] J. C. Bergquist, W. M. Itano, and D. J. Wineland, Phys. Rev. A36, 428 (1987).

[22] J. C. Bergquist, F. Diedrich, W. M. Itano, and D. J. Wineland, in Laser Spectroscopy IX, ed. by M. S. Feld, J. E. Thomas, and A. Mooradian (Academic, San Diego, 1989), p. 274.

[23] Proc. 43rd Ann. Symp. Freq. Control, Denver, June, 1989; IEEE Cat. no. 89CH2690-6, (Nov., 1989), p. 143; IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, submitted.

[24] H. G. Dehmelt, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy (Springer-Verlag, Berlin, Heidelberg) p. 286.

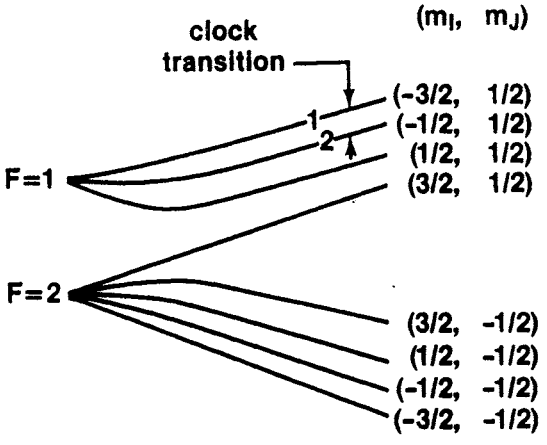


Fig. 1. Hyperfine energy levels (not drawn to scale) of the $^{9}Be^{+}$ 2s $^{2}S_{\frac{1}{2}}$ ground state as a function of magnetic field. At B = 0.8194 T the 303 MHz clock transition is independent of magnetic field to first order.

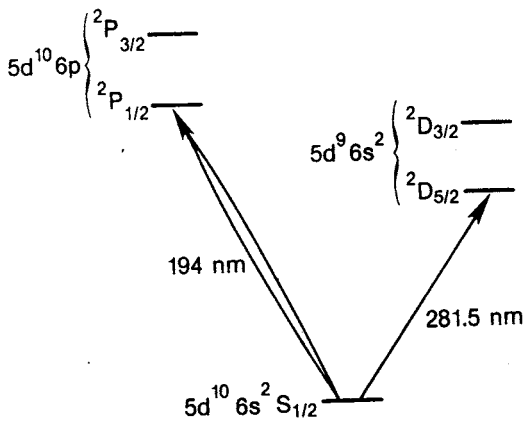


Fig. 2. Simplified optical energy-level diagram for Hg^+. The lifetime of the $^2D_{5/2}$ level is about 0.1 s which would give a linewidth of approximately 2 Hz on the electric quadrupole $^2S_{1/2} \rightarrow {}^2D_{5/2}$ transition. By observing the presence of or lack of fluorescence from the $^2S_{1/2} \rightarrow {}^2P_{1/2}$ transition, the quadrupole "clock" transition can be detected with 100% efficiency.

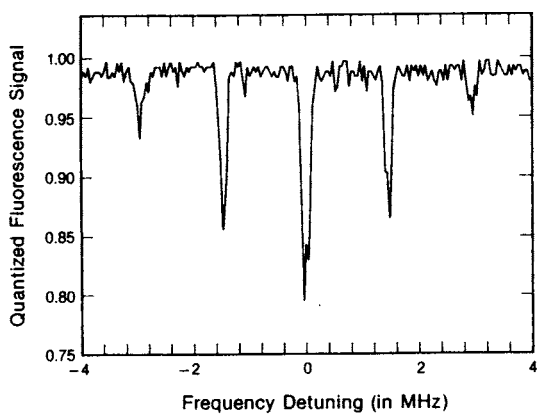


Fig. 3. Quantized signal showing the electric-quadrupole-allowed $5d^{10}6s\ ^2S_{1/2}(m_J=-1/2)$ - $5d^9 6s^2\ ^2D_{5/2}(m_J=1/2)$ transition in a single, laser-cooled $^{198}Hg^+$ ion. On the horizontal axis is plotted the relative detuning from line center in frequency units at 282 nm. On the vertical axis is plotted the probability that the fluorescence from the $6s\ ^2S_{1/2}$ - $6p\ ^2P_{1/2}$ first resonance transition, excited by laser radiation at 194 nm, is on. The electric-quadrupole-allowed S-D transition and the first-resonance S-P transition are probed sequentially in order to avoid light shifts and broadening of the narrow S-D transition. Clearly resolved are the recoilless absorption resonance (carrier) and the Doppler sidebands due to the residual secular motion of the laser-cooled ion. Each point is the average of 230 measurement cycles. (From ref. 21)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-03 Number 2

ATOMS IN OPTICAL MOLASSES: APPLICATIONS TO FREQUENCY STANDARDS

W. D. Phillips, P. D. Lett, S. L. Rolston, C. E. Tanner, R. N. Watts, C. I. Westbrook
Center for Atomic, Molecular and Optical Physics, National Institute of Standards and Technology
PHY B160, Gaithersburg, Maryland 20899, U. S. A.

C. Salomon, J. Dalibard, Laboratoire de Spectroscopie Hertzienne de l'Ecole Normale Supérieure
[Laboratoire Associé au Centre National de la Recherche Scientifique [LA18] et à l'Université Paris VI]
24, Rue Lhomond, F-75231 Paris Cedex 05, France.

A. Clairon, Laboratoire Primaire du Temps et des Fréquences
61 avenue de l'Observatoire, 75014 Paris, France

S. Guellati, Laboratoire Aimé Cotton, Bat. 505, 91405 Orsay, France

Kinetic motion of atoms is, directly or indirectly, one of the major sources of imprecision and inaccuracy in atomic frequency standards. This fact has been a major motivation for laser cooling of atoms and ions. Trapped ion frequency standards have already been demonstrated [1] and the characteristics and prospects of such standards have been discussed in some detail [2]. A major advantage of trapped ion standards is that they offer nearly unlimited storage time and thus very narrow linewidths, On the other hand, they normally operate with small numbers of ions because of the strong ion-ion coulomb interaction. Neutral atom standards are not so limited in useable density, and thus may have significantly better S/N ratio. Recent advances in temperatures obtainable with laser cooling offer interesting possibilities for neutral atom frequency standards

Early theoretical analyses of the laser cooling process, seen as being due to a Doppler-effect - induced velocity dependence of the radiation pressure, indicated that the lowest kinetic energy achievable by laser cooling was about equal to the energy width of the transition excited by the laser [3]. This "Doppler cooling limit" is on the order of a millikelvin for typical allowed transitions.

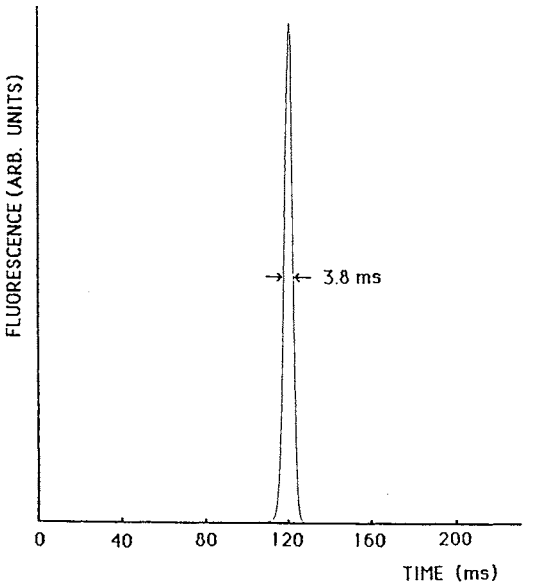
Recent experiments have shown that for the case of multilevel atoms, temperatures much lower than the Doppler cooling limit can in fact be achieved. The effect was first seen in our laboratory at NIST-Gaithersburg [4] for sodium atoms in "optical molasses", a configuration of three mutually orthogonal pairs of counterpropagating laser beams [5]. The effect has since been confirmed for cesium atoms in our laboratory at ENS [6], and for Na atoms by a group at Stanford [7]. The lowest temperatures we have measured at NIST for Na atoms are about 25 μK, while the Doppler cooling limit is 240 μK [8]. At ENS we have measured temperatures for cesium atoms lower than 3 μK, where the Doppler limit is 125 μK [9]. Theoretical explanations [6,10,11] of these results show that mechanisms not involving Doppler shifts are responsible for the ability to cool below the Doppler-cooling-limit. These mechanisms rely on differential light shifts of ground-state sublevels and spatially dependent optical pumping between these levels. The new theory indicates, and the experiments confirm, that the lowest rms velocities achievable are only a few times the recoil velocity $\hbar k/M$, the velocity change given to the atom from the recoil of a single photon. The recoil velocities for Na and Cs are 3 cm/s and 0.35 cm/s respectively, corresponding to temperatures of 2.4 μK and 0.20 μK respectively. The rms velocities (along a vertical axis) corresponding to the lowest temperatures we have observed for these atoms are 10 cm/s and 1.3 cm/s. Thus in each case the lowest observed rms velocity is between three and four times the recoil velocity. These extremely low temperatures offer new opportunities for frequency standards.

Experiments with optical molasses generally start with a thermal atomic beam which is decelerated to near zero velocity using the radiation pressure from a counterpropagating laser beam [12]. In the experiments with Na at NIST this is done continuously using the "Zeeman tuning" technique, while for the Cs experiments at ENS [13] it is done in a pulsed manner using the "chirping" technique. The very slow atoms produced by laser deceleration are introduced into the optical molasses, which further cools them, as well as providing viscous confinement. The confinement allows the atomic density to build up with a time constant determined by the escape of atoms out of the molasses. In our experiments at both NIST and ENS the time constant is about a second, and we collect around 10^8 atoms in about a cubic centimeter.

We measure the temperature of the atoms by a time-of flight (TOF) method. The atoms confined in the molasses are released by suddenly turning off the molasses laser beams. They then travel ballistically to a pair of counterpropagating probe laser beams located below the molasses. Here the atoms fluoresce, scattering many photons as they pass through the probe. This provides efficient detection of the atoms. The fluorescence as a function of time gives the TOF spectrum from which the temperature is determined. This TOF technique was introduced [4] to provide better control of errors due to geometric uncertainties than was possible with the earlier "release and recapture" method [5,6,8,14]. In the NIST Na experiments the typical separation of the probe from the molasses is 1.25 cm and the information about the temperature is conveyed mainly by the position of the peak of the TOF spectrum: hotter atoms arrive earlier. In the ENS Cs experiments, the separation is 7 cm and the temperature information is in

the width of the TOF spectrum, the mean arrival time being nearly independent of temperature. The difference between the behavior of the TOF spectrum in the two cases is due to both the difference in separation between molasses and probe and the difference in the rms velocity in the two cases. These differences also result in much smaller absolute temperature uncertainties in the measurements performed on Cs at ENS compared to those at NIST on Na.

At the lowest observed Cs temperatures the vertical spread in initial atomic positions adds significant "instrumental" width to the TOF spectrum. We have nearly eliminated this effect by "slicing" the molasses. Just after the molasses beams are turned off, we irradiate the atoms with a horizontal, unidirectional beam in the middle of which is a sharp, 2 mm high shadow. After 7 ms of irradiation, all of the atoms except those in a 2 mm high slice are given enough transverse momentum that they fall to the outside of the probe. With this reduction in vertical position spread, and with a 2 mm high probe, the instrumental TOF width is reduced below the intrinsic width due to even the smallest velocity



Time-of-flight spectrum for Cs atoms released from optical molasses at ENS. The width of the TOF spectrum implies an rms velocity spread of 1.3 cm/s along a vertical axis. This corresponds to a temperature of 2.7 μK, and is believed to be the lowest kinetic temperature yet measured for a three dimensional sample of atoms. Such a low temperature offers new possibilities for improved frequency standards.

spreads. We have measured temperatures as low as 2.7 ± 1.0 μK for Cs atoms (see Figure). Down to this lowest temperature, the TOF spectrum is consistent with an initial Maxwellian velocity distribution, a fact which justifies the assignment of a temperature. 2.7 μK corresponds to a measured rms velocity of 1.3 cm/s along the vertical axis, and is to our knowledge the lowest kinetic temperature ever measured for a 3-dimensional sample.

The lowest temperatures in optical molasses are achieved at low laser intensity and large detuning from resonance. This temperature is achieved at a specific ratio of intensity to detuning. In units of Rabi frequency squared divided by detuning times linewidth, this ratio is about 10^{-2} for Cs. Unfortunately, these conditions are not optimal for capturing the atoms from the decelerated atomic beam. In order to have both high molasses density and low temperatures, the atoms are first captured with the molasses laser beams at high intensity and detuned only two linewidths from resonance. After the molasses is loaded with atoms, both the intensity and detuning of the molasses laser are switched to conditions for optimally low temperature.

We intend to use this dense, ultracold sample of atoms in an atomic fountain configuration to make an atomic cesium hyperfine frequency standard. The idea for such an atomic fountain originated with Zacharias[15] and has been studied by other authors [16]. The idea is that the atoms would be launched upwards, passing through a microwave cavity, then falling back again through the same cavity. The two passes would constitute the two interactions needed to obtain a Ramsey fringe spectrum. Related procedures have already been used to observe narrow hyperfine resonances using atoms laser cooled in molasses. At JILA, cooled Cs atoms were dropped through a microwave field [17], and at Stanford cooled Na atoms were launched upward into a cavity, where they were subjected to two separated pulses as they rose and fell, always staying inside the cavity [18].

An important consideration for a fountain experiment is the manner in which the atoms are launched upward. In the Stanford experiment [18] they were pushed up with the radiation pressure from a single laser beam. While this is a very simple procedure, it has the disadvantage of heating the atoms due to the random variation in the number of absorbed photons and the random directions of the radiated ones. In the Stanford experiment this heating did not seriously affect the measurement, but for the extremely low temperatures we have achieved with Cs, it can be very important. Consider launching the atoms with a vertical velocity of 2.5 m/s. This requires scattering about 700 photons. Assuming isotropic scattering, this would lead to about 4.7 cm/s additional rms transverse velocity, or a temperature increase of a factor of 17 over the initial 2.7 μK. A more favorable alternative is to release the atoms from a molasses moving at the required upward velocity. This can be accomplished by shifting

the frequencies of the beams in the vertical beam pair. Increasing the frequency of the upward travelling wave by $\Delta\nu$ and decreasing that of the downward wave by $\Delta\nu$ creates a "moving standing wave" rising at a velocity of $\lambda\Delta\nu$. The atoms will equilibrate with this moving molasses so as to have zero mean velocity in the moving frame. Atoms released from such a moving molasses should remain as cold as the atoms released from static molasses.

Let us consider what would be the performance of an atomic clock based on a Cs optical molasses with the density and temperature we have already achieved. We will assume that we start with a molasses containing 10^8 atoms having an rms velocity of 1.3 cm/s along a given axis. The vertical beams will be frequency-switched by $\pm$2.9 MHz so as to establish a vertical velocity of 2.5 m/s, after which the molasses laser beams will be turned off. This will launch the atoms to a height of 30 cm, with a roundtrip time to return to the launch point of 0.5 sec. The linewidth $\delta\nu$ of the Ramsey signal for a time between resonance regions of 0.5 sec is $\approx$1 Hz. During the 0.5 s of free-fall, the atoms will spread radially by 0.7 cm. This will keep the transverse size of the atomic sample on the order of a half wavelength of the 9 GHz hyperfine radiation. (We envisage a high finesse, Fabry-Perot type of cavity, which will allow a sample of this size to pass through). We assume that of the 10^8 atoms launched that 1/9 are in the F=4, $m_F = 0$ state [19], and, considering the spreading of the atomic sample, and the time for loading and launching the molasses, that 10^6 atoms per second make the hyperfine transition at the peak of the Ramsey signal. The atoms which make the transition will be detected optically with high efficiency. We emphasize that there would be no lasers on during and between the times of interaction with the microwaves so that there will be no perturbation from light shifts. We can also detect those atoms which do not make the transition, allowing the normalization of the resonance signal to the total number of atoms arriving at the detector after the second passage through the cavity. With high detection efficiency the signal to noise ratio should be determined by the shot noise of the atoms making the transition. For 10^6 atoms/s this S/N will be 1000 for 1 second. The square root of the Allan variance for the fountain is given by $\sigma(\tau) = \delta\nu/[\pi\nu(S/N)(\tau/t)^{1/2}]$. S/N is the signal to noise ratio for one cycle of the fountain and t is the time required for a complete molasses loading and fountain cycle. For our case, this gives $\sigma(\tau) = 3.5 \times 10^{-14}$ at one second, or nearly 1×10^{-16} in one day. If the clock performance were limited by this estimate based only on S/N and linewidth, it would outperform any present day atomic clock at one day. We note that this estimate is based on currently achieved molasses densities, which might be improved by more than a factor of 10 with improvements in the flux of atoms from the laser decelerated beam.

The ultimate accuracy of any clock will be limited by systematic errors. For present day Cs standards, these errors are mainly related to atomic motion. In the clock envisaged here, the uncertainty due to second order Doppler shifts would become entirely negligible, less than 10^{-20}. The fact that we employ only a single cavity for the Ramsey resonance eliminates phase shift errors due to unequal electrical lengths between the microwave source and the two cavities used in a traditional Cs standards. The distributed cavity phase shift uncertainty will be reduced, compared to present clocks, by the ratio of the linewidths, in this case, 1 Hz, compared to tens of hertz in typical primary Cs standards. Use of a single cavity also makes easier the evaluation or reduction of the distributed cavity phase shift error by reversing the direction from which microwaves are injected into the cavity. Another way of reducing the phase shift error is to pulse the microwaves to as to restrict the interaction of the atoms with the microwave field to the center of the Fabry-Perot cavity where the wavefronts are nearly flat. We emphasize that this is only possible because the ultra cold, free-falling sample of atoms remains localized to better than a microwave wavelength throughout the fountain cycle. The problem of stray magnetic fields should be much easier to handle in a fountain of 30 cm height than in a several meter long beam apparatus. The problem should be similar to that of shielding a hydrogen maser, where homogeneities better than 5 nT are routinely achieved. With such shielding the related uncertainty in the fountain standard should be able to be reduced below the 10^{-16} level. Whatever the ultimate source of limiting systematic errors, the excellent linewidth and statistical uncertainty of the fountain standard should allow unprecedentedly accurate evaluation of those errors.

In summary, it seems reasonable that neutral atom frequency standards based on atoms that are laser cooled in optical molasses may achieve a performance orders of magnitude better than realized with present Cs standards. We would like to point out that all the technology needed to produce microkelvin Cs in optical molasses is well developed and not much more difficult than that required for an optically pumped Cs standard. In particular, diode lasers can be used for cooling and controlling the atoms. Therefore, it may be that not only primary frequency standards but also commercial ones could make use of this technology. This will be particularly likely if it becomes possible to make ultra-cold molasses in a vapor cell as has been done for a magneto-optic trap [20] eliminating the need for a laser decelerated atomic beam. Finally, even better performance may be obtained in the nearly zero gravity environment of space orbit. There the observation time would not be limited by the height available for a fountain, and, albeit with some loss in S/N, significant improvement in accuracy could be achieved.

The work at NIST was partially supported by the U. S. Office of Naval Research. The work at ENS was partially supported by Collège de France and by DRET. W. D. P. expresses his thanks to Claude Cohen-Tannoudji, Alain Aspect, Laboratoire de Spectroscopie Hertzienne de l'École Normale Supérieure, and the CNRS for their hospitality and support during his stay in Paris.

References:

[1] D, J. Wineland et al., these proceedings.
[2] See, for example, W. M. Itano, J. C. Bergquist, and D. J. Wineland, Science, **237**, 612 (1987).
[3] D. J. Wineland and W. M. Itano, Phys. Rev. A **20**, 1521 (1979).
[4] P. Lett, et al., Phys. Rev. Lett., **61**, 169 (1988).
[5] S. Chu et al., Phys. Rev. Lett., **55**, 48 (1985).
[6] J. Dalibard et al. in Atomic Physics 11, S. Haroche, J. C. Gay and G. Grynberg, eds. (World Scientific, Singapore, 1989) p.199.
[7] Y. Shevy, D. Weiss, and S. Chu in Spin Polarized Quantum Systems, S. Stingari ed. (World Scientific, Singapore, 1989) p. 287; D. Weiss et al., J. Opt. Soc. Am. **B6**, 2072 (1989).
[8]. P. D. Lett et. al., J. Opt. Soc. Am. **B6**, 2084 (1989).
[9] C. Salomon, J. Dalibard, A. Clairon, W. Phillips, and S. Guellati, to be published.
[10] J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. **B6**, 2023 (1989).
[11] S. Chu et al in Atomic Physics 11, S. Haroche, J. C. Gay and G. Grynberg, eds. (World Scientific, Singapore, 1989) p.636; P. J. Ungar et al., J. Opt. Soc. Am. **B6**, 2058 (1989).
[12] For a review of atomic beam deceleration, see for example, W. Phillips, J. Prodan, and H. Metcalf, J. Opt. Soc. Am. **B2**, 1751 (1985).
[13] C. Salomon and J. Dalibard, C. R. Acad. Sci. Paris **306**, 1319 (1988).
[14] D. Sesko, C. Fan, and C. Wieman, J. Opt. Soc. Am. **B5**, 1225 (1988).
[15] see for example the discussion in N. F. Ramsey, Molecular Beams (Oxford University Press, Oxford, 1985).
[16] A. DeMarchi, Metrologia, **18**, 103 (1982); G. Beausoleil and T. Hänsch, Phys. Rev. **A33**, 1661 (1986).
[17] D. Sesko and C. Wieman, Opt. Lett. **14**, 269 (1989).
[18] M. Kasevich et al., Phys. Rev. Lett., **63**, 612 (1989).
[19] Nearly all of the atoms in the molasses are in the F=4 ground state when the molasses beams are turned off. One might imagine optically pumping all the atoms into the $m_F = 0$ state. Unfortunately, the schemes for doing this (see for example, E. deClercq and P. Mangin, in the proceedings of the European Time and Frequency Forum, 21-23 March 1989, Besançon.) involve scattering enough photons to significantly heat the atoms.
[20] C. E. Wieman, University of Colorado-JILA, private communication, 1989.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-03 Number 3

FUTURE PROSPECTS FOR THE COLD HYDROGEN MASER

Robert F.C. Vessot and Edward M. Mattison
Smithsonian Astrophysical Observatory, Cambridge, Massachusetts 02138

Ronald L. Walsworth and Isaac F. Silvera
Harvard University, Cambridge. Massachusetts 02138

Michael L. Dingus
Alabama Cryogenic Engineering, Huntsville, Alabama 35804

1. INTRODUCTION

Since its invention in 1960 by Goldenberg, Kleppner and Ramsey,[1] the hydrogen maser has been developed to the point where frequency stabilities of order parts in 10^{16} have been achieved. In 1978 Crampton et al.[2] pointed out that a cryogenic hydrogen maser (CHM) might have improved stability due to the reduced spin-exchange cross-sections at low temperature and higher atomic density.

In 1980 the possibility of a CHM became a reality when Silvera and Walraven[3] demonstrated the stabilization of atomic hydrogen using special techniques including walls covered with superfluid helium to suppress condensation and subsequent recombination of gaseous atomic hydrogen. Vessot et al.[4] first analyzed the advantage of operation at low temperature and later Berlinsky and Hardy[5] made a detailed theoretical analysis of the CHM with helium covered walls. Comparison to the room temperature maser showed that improvements in frequency stability by a factor as high as 1000 might be achieved. In 1986 three groups,[6-8] including ours, succeeded in demonstrating oscillation of the CHM. To date there have been no long-term experimental studies of the CHM, nor has it been seriously exploited in the measurement of fundamental atomic interactions. Recent theoretical studies by Verhaar, Crampton[9] and coworkers have pointed to previously overlooked effects due to the hyperfine interaction of hydrogen that become important at low temperature and may possibly limit the stabilities predicted by Berlinsky and Hardy for the CHM.

The SAO/Harvard CHM is shown schematically in Figure 1. The basic principles of its operation are the same as that of the room temperature maser. However, a number of important advantages are available at low temperature. In the conventional maser wall coatings are of Teflon, which is neither a reproducible nor a time-invariant surface. Thus the behavior of the room-temperature maser can be time dependent or can vary from maser to maser depending on the wall covering preparation. The cryogenic maser, on the other hand, uses superfluid helium surfaces, which do not change their nature with time, can cover all subsurface impurities, and allow recombined hydrogen molecules to migrate away from the surface to the underlying substrate. Magnetic shielding can be greatly improved at low temperatures by using superconducting shields, and long-term dimensional stabilities of materials are greatly increased, promising greater long-time stability. The cross-section for spin-exchange collisions between atoms is substantially reduced from room temperature, as shown in Figure 2. Furthermore, the velocity of the atom is reduced by a factor of 24, resulting in a reduction by a factor of about 2000 in the spin-exchange relaxation rate.

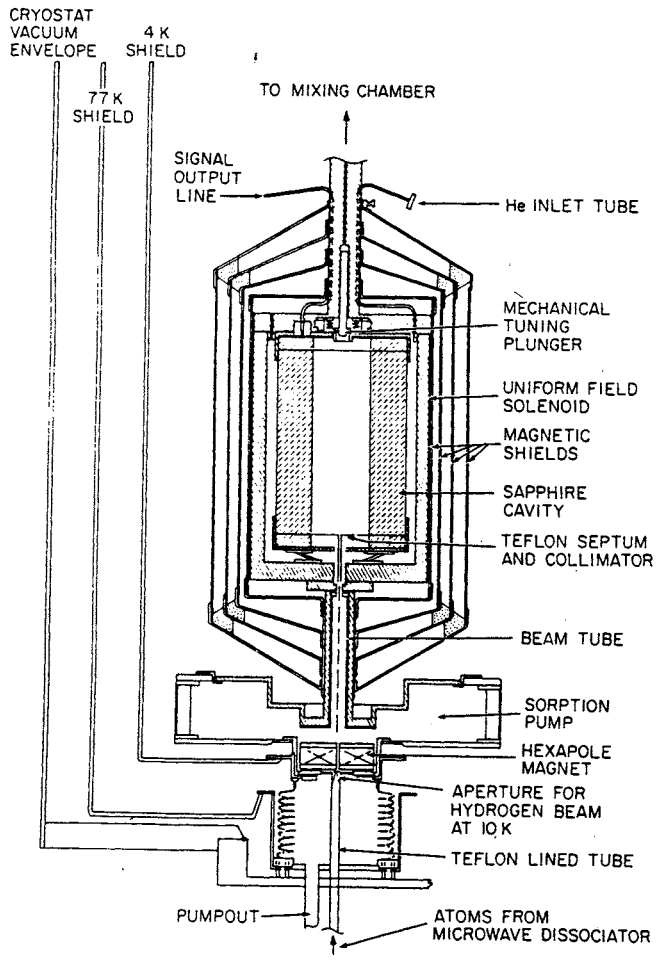


Figure 1. The Harvard/SAO Cryogenic Hydrogen Maser

This paper describes a program of research and some of the development efforts required to realize a cryogenic hydrogen-maser with extremely high frequency stability.

2. THE DESIGN OF THE SMITHSONIAN ASTROPHYSICAL OBSERVATORY-HARVARD UNIVERSITY CRYOGENIC HYDROGEN MASER

The design of this maser, which has been operated in a dilution refrigerator, is focussed on obtaining extremely high mechanical stability and the capacity of obtaining well-controlled magnetic fields and temperature. To obtain a wide range of operating conditions we have provided independent control of all possible parameters.

The maser has already been operated at low temperatures.[8,10] Its innovative design begins with a source of hydrogen atoms from a room temperature discharge that enters the cryostat, is cooled to 12°K, effuses from a hole to form an atomic beam, and passes through a 6-pole magnet. The atoms in the two upper hyperfine magnetic sublevels are focussed into the entrance aperture of the storage volume, which is the interior of a thick-walled, cylindrical dielectrically-loaded TE_{011} cavity made of sapphire, coated on its exterior with a conducting layer of silver. Surrounding the cavity is a 3-section, longitudinal magnetic field solenoid. The cavity is housed in a copper vacuum tank thermally connected to the mixing chamber of a dilution refrigerator.

The isothermal copper tank is enclosed by a 4-layer magnetic shields made of Cryoperm*. Superfluid helium flown into the cavity through a controlled leak, coats the interior surfaces, and emerges through a neck tube that connects the cavity to a very high capacity sorption pump. The vapor pressure in the neck is kept low by providing a large surface area of copper foil in the neck so that the helium film is kept below its saturation thickness. A longitudinal field solenoid is wound on the neck and a transverse RF magnetic field coil in the neck allows us to alter the populations of the hyperfine states of the atoms before they enter the storage volume.

The measurable quantities available with this design of hydrogen maser are 1) Output frequency; 2) Output power level; and 3) Linewidth of oscillation signal. The controllable parameters are 1) Temperature, and thus the relative population of surface and gas state of atoms; 2) Helium flow rate (film thickness); 3) Beam flux; 4) Atomic state distribution entering the storage volume (by inducing resonance transitions in

* ® Vacuumschmeltz.

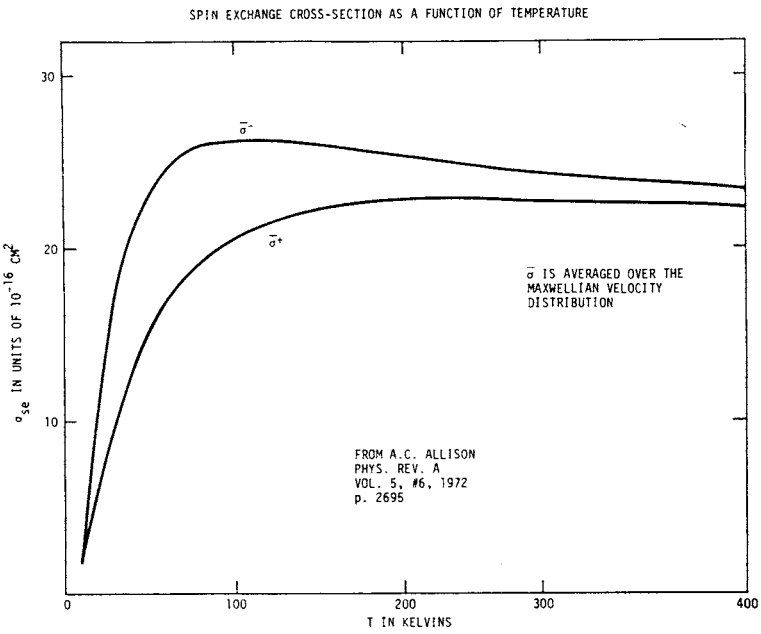


Figure 2. Spin exchange cross-section of atomic hydrogen between 4K and 400K

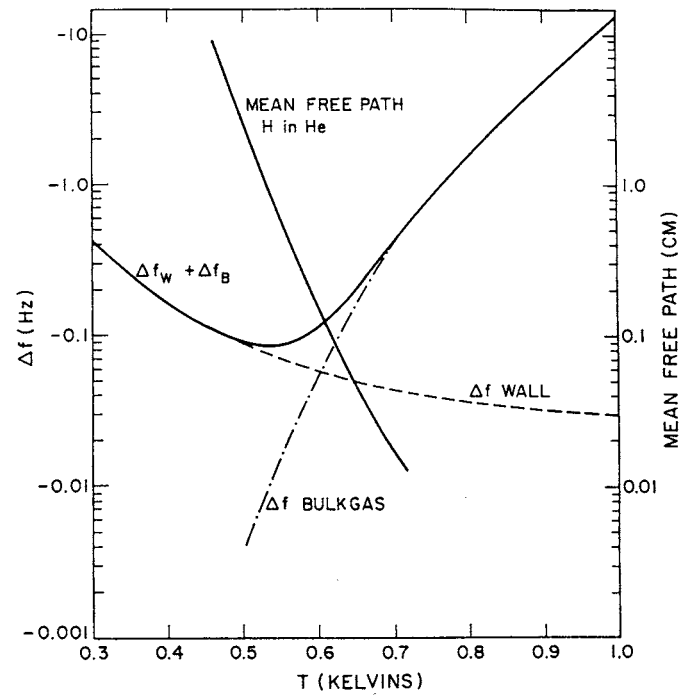


Figure 3. Hydrogen collision frequency shifts with gaseous ^{4}He and superfluid surfaces of ^{4}He

the neck with the transverse RF field coil; 5) Atomic state distribution within the storage volume (by means of a separately excited transverse field coil surrounding the cavity); 6) Magnitude and homogeneity of the axial d.c. magnetic field in the storage volume; and 7) Resonance frequency of the cavity.

It has been shown[3,11] that superfluid films of helium can provide storage volume surfaces capable of confining an ensemble of hydrogen atoms so that their phase coherence is maintained for many hours. Due to the résidency time of the hydrogen atoms on the surface, this cold storage process causes a substantial average frequency shift in the oscillating atoms; the residency time, and thus the frequency, is sensitive to the temperature of the superfluid helium surface. There is also a frequency shift associated with hydrogen atoms colliding with helium atoms in the gas phase.[11] The liquid and gas collision shifts are shown as Δf_{wall} and $\Delta f_{bulk\ gas}$ bulk gas in Figure 3; we see that the combined shift has a minimum at about 0.53K for saturated helium pressure, which allows us some latitude in the control of temperature so as to obtain a stable oscillation frequency. In the existing maser we expect to be able to control temperature within 20μK so that the resulting temperature-induced frequency shifts will be in the 10^{-18} domain. Also shown in Figure 3 is the mean free path of H in helium. At 0.53K the mean free path is the storage volume is 4 cm. In the neck tube it is considerably larger.

3. PLANNED PROGRAM OF RESEARCH

We plan two general areas of experimental investigation. First we will measure low temperature interactions of atomic hydrogen with surfaces of liquid helium, with gaseous helium and with other hydrogen atoms. These interactions will be measured with high precision and with accuracy traceable to the National Institute of Standards and Technology (NIST) by means of the Global Positioning System (GPS). Second, we will make frequency stability measurements with the low temperature maser against state of the art room temperature atomic hydrogen masers to assess the potential limits of stability and accuracy capability of the cold maser.

We will begin by making frequency shift and line Q measurements relating to hydrogen-hydrogen interactions and hydrogen-helium interactions both on liquid He surfaces and in the gas phase. The results from hydrogen-hydrogen measurements will be compared to the theoretical analysis of the proton-proton, electron-electron 4-body collision process made by Verhaar and others.[9] The hydrogen-hydrogen collisional spin exchange-hyperfine cross-sections resulting in line shifts and broadening, calculated by Verhaar and others, have a considerable level of uncertainty due to uncertainty in the hydrogen-hydrogen singlet-triplet potentials and to the sensitivity of the cross-sections to these potentials. These measurements are fundamental to understanding the interaction potentials of hydrogen. Present predictions indicate the possibility of a strong, systematic frequency shift that depends on the length of time the atoms are confined in the maser interaction region and not simply on the collision rate of atoms as has been previously believed. The new experimental data will guide developments of the low temperature hydrogen maser as a frequency standard of unprecedented stability and accuracy.

To make proper use of the extraordinarily high frequency stability of the signals from the cryogenically cooled maser in measuring the collisional effects of hydrogen we require frequency measuring equipment with excellent noise qualities and long-term phase stability, and traceability to the definition of the second of time. Our electronics system is designed to perform the following three functions: 1) to measure the resonance frequency of the maser cavity, 2) to apply a pulse of microwave energy of variable amplitude and duration to stimulate resonant transitions between the F=1, m_F=0 and F=0 m_F=0 levels of the hydrogen atom and measure the amplitude and frequency of the "ring-down" or signal decay as the atoms release energy to the maser cavity resonator, and 3) to measure the amplitude and frequency of the oscillating maser with the greatest possible precision with respect to two state of the art masers using a 3-way cross-correlation technique and to relate the frequency of the cold maser to the best available primary frequency standards. Figure 4 shows the system we have in operation.

At present we can measure variations of the oscillating maser line Q, $Q_\ell = 2\pi f_o/\gamma_2 \approx 10^{10}$ with a precision of 0.3 percent, (where f_o is the output frequency and γ_2 is the rate of dephasing of the atoms) and with an accuracy of 5 percent limited by our knowledge of the loaded cavity Q, Q_L. The accuracy of the measurement of cavity resonance frequency is about ± 3.5 parts in 10^8. With $Q_\ell \approx 10^{10}$ and $Q_L = 64 \times 10^3$ we are able to make reproducible measurements of output frequency with accuracy of $\pm 2.3 \times 10^{-13}$, which is well within the stability of the reference maser monitored by the GPS system.

Measurements of frequency and decay rate under maser conditions that will not permit self-oscillation are made by applying pulses of microwave energy at 1,420,405,751 Hz to the 20 dB directional coupler to stimulate the F=1, m_F=0$\rightarrow$F=0, m_F=0 "clock" transition of the atoms in the storage volume of the maser cavity. This system is useful for getting the cold maser started, adjusting the superfluid helium surface film thickness and the atomic beam flux density, trimming the magnetic fields in the cavity, etc. A precise evaluation of decay rates and ringing frequency will be made by conversion of the output voltage ringing signal to digital form and analysis by a computer.

As seen in Figure 4, a coaxial line from the hydrogen maser cavity is led out of the top of the dilution refrigerator by a coaxial vacuum connector and then directly through the 20 dB directional coupler and on to the low noise (10K, noise temperature 35 dB gain) cooled FET preamplifier. The receive itself is a quadruple heterodyne system. The last mixing stage is a phase sensitive detector, whose output controls a 100 MHz voltage-controlled crystal oscillator (VCXO) with good short-term (τ<0.1 sec) phase stability. The control bandwidth is adjusted to optimize the overall spectral density of the phase fluctuations in the output signal. A plot of the spectral density of phase variations is shown in Figure 5, based on the use of a commercially-available, 100 MHz VCXO. From the intersection of the phase spectrum of the cold maser and that of the VCXO, we see that the loop bandwidth must be chosen at about 100 Hz to get the lowest overall spectral density under the best operating conditions of the maser. The technique of using simultaneous frequency intercomparisons between 3 (or more) oscillators is often referred to as the "3-cornered-hat method.[12] Because of the near simultaneity of the three measuring processes, correlations result and we can invoke these

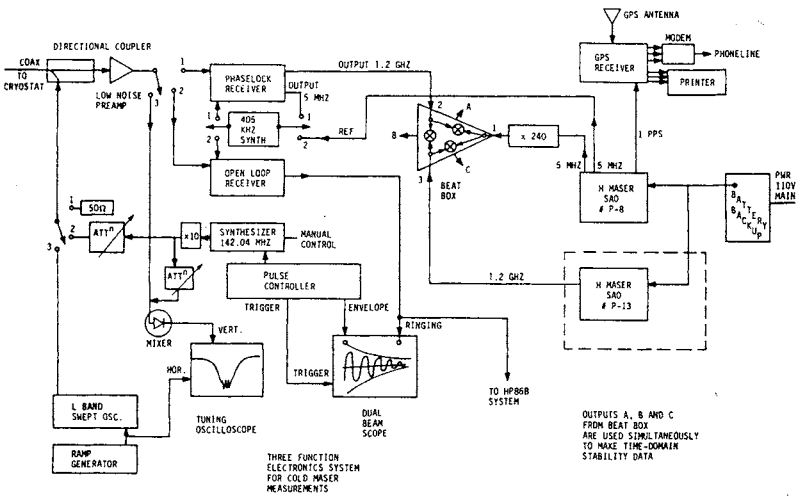


Figure 4. Three function electronics system for cold maser measurements

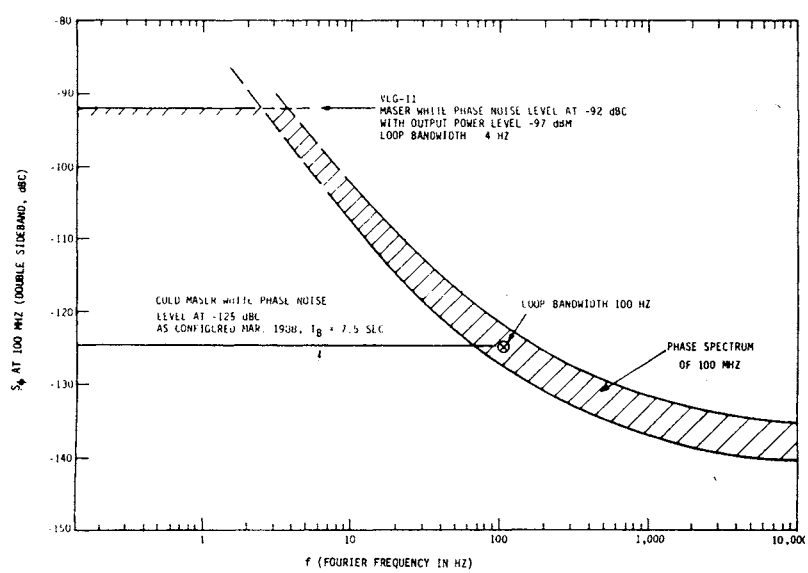


Figure 5. Spectral density of phase fluctuations at 100 MHz

correlations to calculate the variance associated with each individual oscillator.

An oscillator's frequency stability over an averaging interval τ can be expressed in terms of the Allan variance, $\sigma(\tau)$. For oscillators a, b and c, being measured simultaneously,

$$\sigma_a^2 = \frac{\sigma_{ab}^2 + \sigma_{ca}^2 - \sigma_{bc}^2}{2}$$

$$\sigma_b^2 = \frac{\sigma_{bc}^2 + \sigma_{ab}^2 - \sigma_{ca}^2}{2}$$

$$\sigma_c^2 = \frac{\sigma_{ca}^2 + \sigma_{bc}^2 - \sigma_{ab}^2}{2}$$

where $\sigma_i^2(\tau)$ is the variance ascribable to the performance of the i^{th} oscillator and $\sigma_{jk}^2(\tau)$ is the variance measured from the joint performance of oscillators j and k, during the same time interval τ, at the same epoch of time.

The success of this method depends on how well we can estimate the true variances σ_{jk}^2 and the extent to which the frequency measurements can be made simultaneously. The measurement involves a statistical process whose error converges as $N^{-\frac{1}{2}}$, where N is the number of sample variances. We must be careful to avoid the many hazards that occur in making the 3 independent simultaneous measurements associated with signal cross-talk between channels. We expect to be able to measure the frequency stability of the CHM to a factor of perhaps 10 to 20 better than the stability of the room temperature masers by using this process. Until we can operate two or more cold masers, we will have to rely on the 3-corner hat technique to characterize the CHM's frequency stability.

4. NOISE ANALYSIS AND FREQUENCY STABILITY

The stability of the oscillating hydrogen maser can be expressed in terms of maser oscillation parameters.[5,4] Describing the stability in terms of the Allan deviation we identify contributions associated with two distinctly different spectral distributions.[13]

1. The additive, white phase noise (constant spectral density of phase fluctuations) associated with the thermal noise, kT, per unit bandwidth; and

2. A fundamental limit associated with the white frequency noise (constant spectral density of frequency fluctuations) associated with thermal noise kT within the linewidth of the ensemble of oscillating hydrogen atoms.

The Allan deviation associated with the first process, when the maser is operated at optimum flux[5] is

$$\sigma_A(\tau) = \left[\frac{k(T_N + T)B\sigma\bar{v}}{\hbar\omega^3 V_b} \frac{Q_{ext}}{Q_L} \frac{I_{tot}}{I}\right]^{\frac{1}{2}} T_b \tau^{-1} \qquad (1)$$

The Allan variance associated with the second process when the maser is operated at optimum flux[5] is

$$\sigma_F(\tau) = \left[\frac{8kT\sigma\bar{v}I_{tot}}{\hbar\omega^3 V_b\ I}\right]^{\frac{1}{2}} \tau^{-\frac{1}{2}} \qquad (2)$$

where k is Boltzmann's constant, $\hbar$ is Planck's constant divided by 2π, T is the maser's physical temperature, T_N is the noise temperature of the signal processing system, $\sigma\bar{v}$ is the spin exchange cross-section times the average relative velocity of the atoms, $\omega = 2\pi f$ is the angular frequency of maser oscillation, V_b is the volume of the atomic storage volume and T_b is its storage time constant, B is the measuring system's noise bandwidth, Q_L is the loaded cavity resonator quality factor, and Q_{ext} is the quality factor associated with power loss in the external circuit coupled to the cavity. I_{total}/I is the ratio of total atomic hydrogen flux entering the storage volume to the flux in the desired F=1, m_F=0 hyperfine sublevel.

Figure 6 shows the Allan standard deviation, $\sigma_y(\tau)$, of the reference masers and that predicted for the cold maser, plotted against averaging time interval, τ. The upper curve is the measured performance of SAO VLG-11 masers operating at optimum flux. These data follow very closely the behavior given by equations 1 and 2. However, for $\tau > 4{,}000$ sec, $\sigma(\tau)$ increases with τ, indicating that other processes are affecting the frequency. The principal component of these processes are thought to be identified with a frequency pulling effect owing to the instability of the cavity resonator frequency. A change in cavity resonance frequency causes a change in output frequency proportional to the ratio of the loaded cavity Q to oscillation line Q, Q_L/Q_ℓ. If there exists a $1/f$ spectral density to the instability of the cavity resonator frequency we would see a flattening of the Allan standard deviation, while a $1/f^2$ spectrum of cavity resonance would produce a $\tau^{\frac{1}{2}}$ slope, and a linear drift of the cavity resonance with time would produce a τ^1 trend in $\sigma(\tau)$.

It is important to note that the fundamental limit discussed in equation 2 is the special case for optimum beam flux, which occurs when the spin-exchange line broadening due to hydrogen-hydrogen collisions has the same magnitude as the rate of loss of hydrogen atoms from the oscillating region. The general case is given by

$$\sigma(\tau) = \frac{1}{Q_\ell}\sqrt{\frac{kT}{2P\tau}}.$$

where P is the power delivered to the cavity by the oscillating atoms.

In the statistical representation of the maser performance we have not yet included the most important feature of the maser, its very narrow oscillation linewidth. The

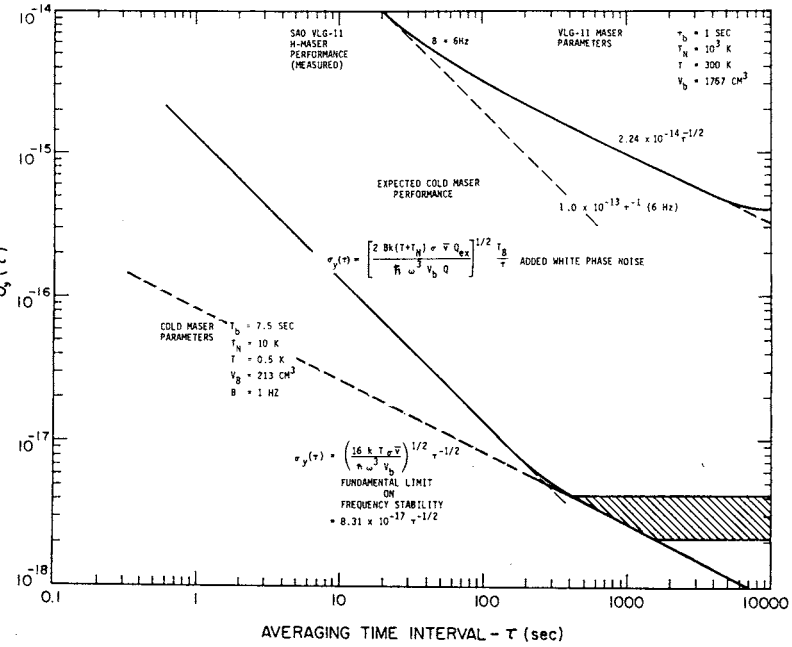


Figure 6. Expected performance of CHM and existing SAO VLG-11 masers. Frequency stability is expressed in terms of the Allan standard deviation, $\sigma_y(\tau)$.

linewidth is related to the duration of time over which atoms are stored without significantly perturbing the phase of their oscillating magnetic dipole moments. The total rate of change of dephasing, γ_2, is related to the oscillation linewidth used to define the atomic line Q, Q_ℓ, as follows:

$$Q_\ell = \frac{\omega_o}{2\gamma_2}.$$

The importance of having a high line Q relates to the precision of the output frequency. The accuracy of this frequency depends on our knowledge of systematic processes that shift the hydrogen frequency. The processes resulting in the relaxation rate γ_2, which include all mechanisms that result in loss of phase coherence, also result in systematic frequency shifts associated with a number of collision processes. One can consider the phase of an individual hydrogen atom as having the probability both of being, on average, systematically shifted as well as being spread out over a range of values when it collides with another atom. A high line Q is equivalent to a limitation on the spread of average phase that the atoms experience.

The lower solid curve in Figure 6 shows the cold maser performance expected under optimum flux conditions. The underlying fundamental limit is shown by the lower dashed line, sloping as $\tau^{-\frac{1}{2}}$. For optimum flux the maser is now configured to have $Q_\ell \approx 10^{10}$. For stability of 10^{-18} at 10,000 seconds we require that the cavity resonator must be stable to 2×10^{-4} Hz. The great mechanical stability of sapphire, along with its very small coefficient of dielectric constant with temperature at 0.5K, gives us reasonable optimism that we will reach stabilities below 1×10^{-17} at about 1,000 seconds and allow us to observe the $\tau^{-\frac{1}{2}}$ region of the variance plot.

5. PRELIMINARY DEVELOPMENT OF A SPACE-WORTHY RECIRCULATING HELIUM-3 REFRIGERATOR

A significant part of the SAO/Harvard cryogenic hydrogen maser program is the development of a ^{3}He refrigerator specifically designed for the cryomaser. One of the ground rules for this development is to use technology that could, at least in principle, be suitable for long-term application in space. This of course, means recirculating device, rather than a one-shot system. We also, at least at the outset, require 5 milliwatts of cooling power at 0.5K, which is, by today's standards, a high level of cooling power. The hot bath for this system is a constantly replenished ^{4}He reservoir kept below 1.2K by being pumped. Depending on the anticipated lifetime of a space mission, the hot bath could be a large dewar containing several thousand liters of liquid helium vented to outer space, or a far more costly and complicated recirculating ^{4}He system. The anticipated consumption of ^{4}He at 1.2K to operate at a 5 milliwatts load is approximately 0.5 liters per day; in comparison the evaporative loss of presently available large space dewars holding a few thousand liters of ^{4}He is about three times this amount.

Figure 7 shows a layout of the system. The circulating pump is a specially modified Danielson Tribodyne-500 series pump, which is an oil-free device capable of operating in any orientation. It has a pumping speed of about 27 liters per sec at inlet pressures from 10^{-2} to 10^{-6} torr. The compressed ^{3}He is led through a heat exchanger incorporated in the exhaust line, liquified under pressure at about 1.2K, and led to the ^{3}He evaporator. To avoid the problems associated with a free liquid surface the ^{3}He is evaporated by passing through a porous silver plug, which is connected thermally to the copper stem of the cryomaser.

The cryostat includes the capability of pre-cooling the ^{4}He pot and rapidly cooling the maser with an exchange gas to enable initiation of the ^{3}He cooling process. The liquid ^{4}He tank holds 30 liters and we expect the system to operate for about 7 days without refilling ^{4}He. The 77K outer shield is thermally connected to an annular 30 liter liquid nitrogen tank and will be replenished daily by an automatic system.

6. MEASUREMENT OF HYDROGEN-HYDROGEN INTERACTIONS

The success of the cryomaser depends very much on the nature of the hydrogen-hydrogen interactions at temperatures below 1K. Theoretical studies of these interactions[9] have led to startling results that require experimental verification. At very low temperatures the frequency shift owing to hydrogen-hydrogen collisions and cavity frequency pulling is no longer strictly proportional to the density-dependent spin exchange collision process. In room temperature masers it is possible to tune the cavity resonator at a frequency offset from the true hyperfine frequency so that the cavity pulling exactly compensates for the collisional frequency shifts when the hydrogen density is varied, conventional masers are tuned in this way to avoid frequency shifts owing to beam flux variations. In cryomasers no such cancellation seems available.

From Verhaar, et al.,[9] the expression describing the frequency pulling and collisional shifts is

$$\delta\omega = \left[\Delta + \alpha\lambda_o(1+\Delta^2) - \Omega\right]\gamma_2 + \Omega\gamma_o$$

where $\alpha = \left(\bar{v}/\mu_o\right)\left(\hbar/\mu_B{}^2\right)\left(V_c/\eta Q_L V_b\right)$, a constant that is determined by the temperature and the construction of the maser (here μ_o is the permeability of free space and μ_B is the Bohr magneton), $\Delta = 2\frac{\Delta f_c}{f} Q_L$ where Δf_c is the cavity mistuning, f is the oscillators frequency, γ_o is the relaxation rate resulting from all causes not due to spin exchange collisions, and γ_2 is the total relaxation rate that defines the maser oscillator's line Q, $Q_\ell = \omega/2\gamma_2$.

The parameter Ω includes three velocity-averaged collisional spin exchange cross-sections, $\lambda_{0,1,2}$, which cause frequency shifts, and two velocity-averaged cross-sections, $\bar{\sigma}_{1,2}$, which cause line broadening. These cross-sections are shown as functions of temperature

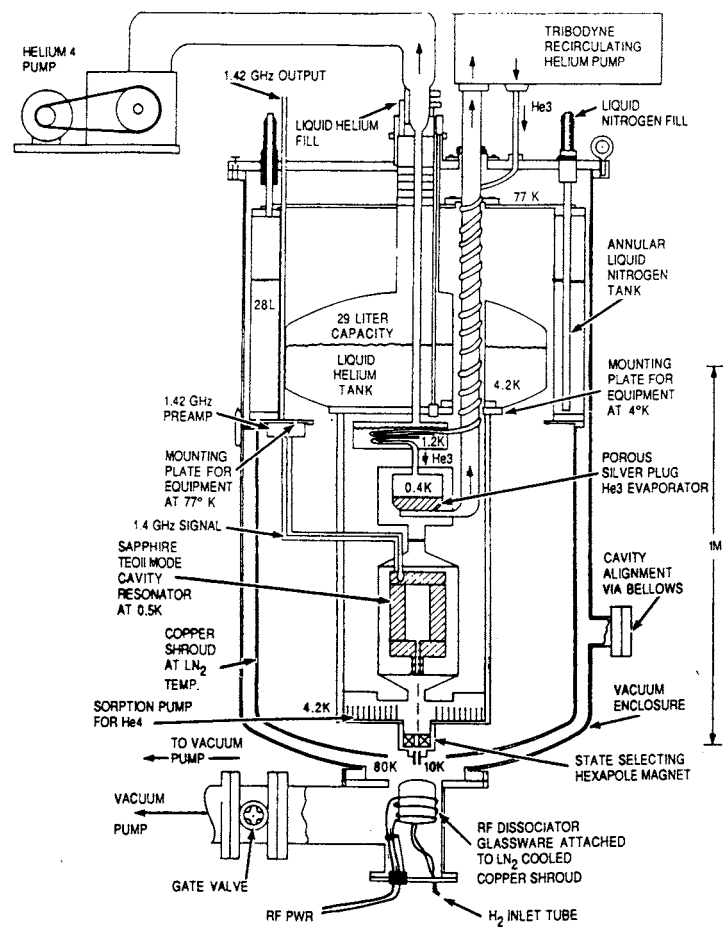


Figure 7. ^{3}He refrigerator for the cryogenic hydrogen maser

in Figure 8, which also shows the magnitude of Ω as a function of temperature for the two possible extremes of hyperfine state distribution among the atoms in the bulb.[9] The lower curve is for atoms entering entirely in the F=1, m_F=0 state, while the upper curve is for atoms entering at equal rates in both the F=1, m_F=0 and F=1, m_F=1 states. Here the parameter α has been chosen as 10^3. The parameter Ω decreases further as temperature increases, and is calculated to have little effect on the stability of room temperature hydrogen masers.[14]

In order to obtain frequency stability at the 2×10^{-18} level at 0.5K, where $\Omega \sim 7 \times 10^{-2}$, we must maintain γ_o constant to within 3×10^{-7} sec.$^{-1}$, assuming that atoms enter the cavity only in the desired F=1, m_F=0 state (the $\rho_{cc'}, \rho_{cc} = 1$ condition). Considering that γ_o depends on all the non-spin exchange effects such as magnetic gradient relaxation, constancy of the ^{4}He film thickness, and dimensional stability of the storage volume, the 2×10^{-18} level may be difficult to achieve.

7. CONCLUSION

For the cryogenic hydrogen maser to progress as a frequency standard it is now up to experimenters to measure the value of Ω and λ_o and thus determine the validity of the cross-sections described in Figure 8A. These measurements are a principal goal of the SAO/Harvard CHM program. It may turn out that operation at a somewhat higher temperature, where Ω is smaller, will give better results. Such temperatures will require wall coatings of other materials than ^{4}He films. Crampton[15,16] and his associates at Williams College are preparing a maser to operate with surface coatings of frozen neon at temperatures near 10K. Operation at this temperature will provide an additional set of data for Ω and $\bar{\lambda}_0$ in a different regime.

Although it is still too early to know whether the cryogenic hydrogen maser can realize the two to three orders of magnitude improvement in frequency stability that has been predicted, the CHM is certain to probe a rich new area of physics.

8. ACKNOWLEDGEMENTS

This work is supported by the Smithsonian Institution's Scholarly Studies Program, the NASA Jet Propulsion Laboratory's Science Coordinating Committee for Experimental Gravitation, and the U.S. Department of Energy. R.L.W. is supported by a NASA Graduate Student Researchers Program Fellowship.

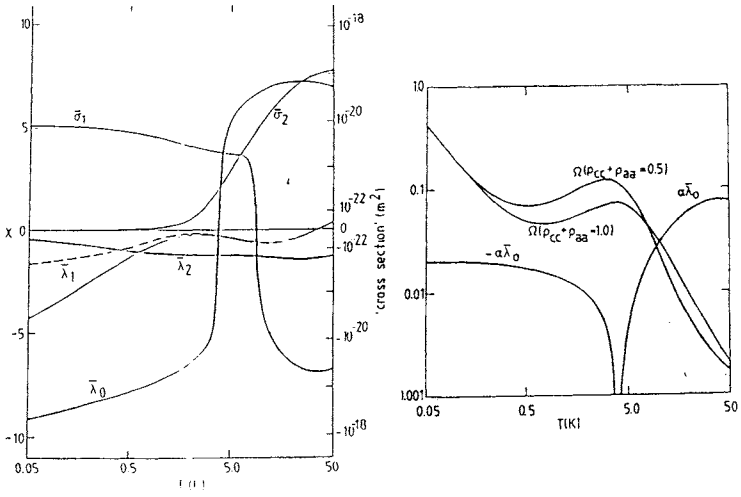


Figure 8. Collisional frequency shift and relaxation cross-sections λ_0, λ_1, λ_2, $\bar{\sigma}_1$, and $\bar{\sigma}_2$ (from reference 9)

REFERENCES

1. H.M. Goldenberg, D. Kleppner and N.F. Ramsey, *Phys. Rev. Lett.*, **5** (1960) p. 361, also *Phys. Rev. Lett.* **126** (1962) p. 603.
2. S.B. Crampton, W.D. Phillips and D. Kleppner, *Bull. Am. Phys. Soc.*, **23** (1978) 86.
3. I.F. Silvera, and J.T.M. Walraven, *Phys. Rev. Lett.* **44**, (1980) 164.
4. R.F.C. Vessot, M.W. Levine and E.M. Mattison, In *Proc. 9th Annual Precise Time and Time Interval Conference*, NASA Technical Memorandum No. **78104** (1978), p. 549.
5. A.J. Berlinsky and W.N. Hardy, In *Proc. 13th Annual Precise Time and Time Interval Conference*, NASA Conference Publication, No. **2220**, 1982, p. 547.
6. G.P. Kochanski, J.M. Doyle, T.J. Greytak and D. Kleppner, *Phys. Rev. A* **34** (1986) p. 1602.
7. M.D. Hurlimann, W.N. Hardy, A.J. Berlinsky and R.W. Cline, *Phys. Rev. A* **34** (1986) p. 1605.
8. R.L. Walsworth, I.F. Silvera, H.P. Godfried, C.C. Agosta, R.F.C. Vessot, and E.M. Mattison, *Phys. Rev. A*, **34** (1986) p. 2550.
9. B.J. Verhaar, J.M.V.A. Koelman, H.T.C. Stoof, O.J. Luiten, and S.B. Crampton, *Phys. Rev. A*, **35** (1987) 3825;
10. R.F.C. Vessot, E.M. Mattison, R.L. Walsworth, I.F. Silvera, H.P. Godfried and C.C. Agosta, *IEEE*, IM-36, No. **2**, p. 588 (1987).
11. M. Morrow, R. Jochemson, A.J. Berlinsky and W.N. Hardy, *Phys. Rev. Lett.* **46** (1981) 195, also *Phys. Rev. Lett.* **47** (1981) 455 (E)
12. J. Groslambert, D. Fest, M. Olivier, J.J. Gagnepain, *Characterization of Frequency Fluctuations by Cross-correlations and by Using Three or more Oscillators*, In *Proc. 35th Ann. Freq. Control Symposium*, USAERDCOM, Ft. Monmouth, N.J. 07703. May 1981, p. 458.
13. R.F.C. Vessot, L. Mueller and J. Vanier, In *Proc. IEEE*, **54**, No. **2**, 1966, p 199.
14. J.M.V.A. Koelman, S.B. Crampton, H.T.C. Stoof, O.J. Luiten and B.J. Verhaar, "Spin-Exchange Frequency Shifts in Cryogenic and Room Temperature Hydrogen Masers," *Phys. Rev. A* 38, 3535 (1988).
15. S.B. Crampton, "Hydrogen Atom Sorptions on Molecular and Solid Surfaces," *American Phys. (Paris)*, Vol. 10, pp. 893 (1985).
16. S.B. Crampton, K.M. Jones, G. Nunes and S.D. Souza, "Hydrogen Maser Oscillation at 10K," *Proc. of the 16th Annl. Precise Time and Time Interval Meeting*, Nov. 27-29, 1984, Goddard Space Flight Center, Greenbelt Maryland, pp. 339-350.

Resonators, Oscillators and Materials

ROM-01 Resonator Fundamentals

Chairman: R.C. Smythe

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-01 Number 2

3-d nonlinear finite element analysis; Determination of the static force induced frequency perturbation of quartz-crystal resonators

K. Trümpy
ETA S.A. , Abt. feintechnisches Labor , CH-2540 Grenchen , Schweiz

Summary

A triangular-prismatic isoparametric element has been developed with quadratic interpolation in the x-y-plane and ranging from linear up to quintic interpolation in the z-direction. For example, cubic interpolation in z makes possible to calculate the fundamental thickness-shear mode. Also the program is used to calculate the tangent-stiffness matrix which is used in the Newton iterative procedure, to treat both geometric- and material-nonlinearity.
These instruments allow to determine the eigenfrequency first-order perturbation of vibrating structures, which is caused by static forces. The following effects caused by static deformation have been considered : the change in the resonator geometry and elastic constants,the influence of initial deformation and stress.
Computed results are represented for AT and SC-cut quartz crystal plates vibrating in the fundamental and the third harmonic overtone thickness-shear mode. The resonators are exposed to perturbation forces acting within or perpendicular to the plate. Also the method is used to calculate the force sensitivity of a double-ended tuning fork quartz resonator used as a force sensor. The values are compared to experimental and analytical values obtained by other authors.

Introduction

Nonlinear finite-element analysis is often carried out with the Newton iterative procedure, requiring a tangent-stiffness matrix at each step of deformation. The tangent-stiffness matrix, which relates incremental nodal deformations to incremental nodal forces, is appropriate to analyse the superposition of small vibrations on initial static deformations. This matrix generally consists of the normal stiffness matrix with a nonlinear constitutive relation, the initial-deformation matrix, and the initial-stress matrix.
The present paper describes the calculation of the tangent-stiffness matrix (and mass matrix) of a triangular-prismatic isoparametric element with quadratic displacement interpolation in the triangular-plane and ranging from linear up to quintic order in the normal direction. The tangent-stiffness matrix, which depends on the initial deformation (calculated with a linear static finite-element analysis), and the unperturbed eigenvalue and eigenvector of the resonator mode (finite element solution of the eigenproblem obtained with the subspace iteration method), make it possible to apply a perturbation method in order to calculate the frequency change in first-order. Four different effects caused by the initial static deformation have been considered, namely the change in the resonator geometry, the change of the elastic constants, the nonlinearity in the strain-displacement relation, and the influence of initial stress.
The paper also gives values regarding the changes of the thickness-shear mode-c frequencies for AT and SC-cut quartz disks which have been subjected to either a pair of diametrical forces or to transverse cantilever bending forces . Cubic normal displacement interpolation permits to calculate the fundamental thickness-shear mode, while quintic interpolation is used to calculate the third harmonic overtone thickness-shear mode.The thickness/diameter ratio of the quartz disks is less than 0.02. By using a fine-mesh discretisation with a thick-plate model, it has been seen that, for such ratios, there are nearly no transverse bending deformations superposed on the thickness-shear oscillations. Consequently, a discretisation rather rough compared to the plate thickness is sufficient. The number of degrees of freedom is below nine thousand.
Another type of resonator analysed with our method is the double-ended tuning fork, which is used as a force sensor. Here the initial-stresses are predominant, while for the thickness-shear mode they do not contribute to the frequency perturbation.
All results obtained are in good accord with experimental and analytical values of other authors.

Finite element equations of incremental elastic deformation

The objective in this section is to present the nonlinear continuum-mechanics equations for the finite element formulation. We use the principle of virtual work with the possibility that the elastic body undergoes large displacements and rotations, large strains, and that the stress-strain relation is nonlinear.
The principal of virtual work states that the equilibrium of the deformed body requires that for any small virtual displacements imposed onto the body of volume V, the total internal virtual work δU is equal to the total external virtual work δW

$$\delta U = \int_V \vec{\sigma}^T \delta\vec{\varepsilon}\, dV \quad = \quad \vec{f}^T \delta\vec{a} = \delta W$$

The actual stresses $\vec{\sigma}$ and the actual forces $\vec{f}$ are related to the initial state of deformation. The virtual strain $\delta\vec{\varepsilon}$ is expressed in terms of the imposed virtual displacement $\delta\vec{a}$:

$$\delta\vec{\varepsilon} = \hat{\mathbf{B}}\,\delta\vec{a} \tag{1}$$

This leads to the variational equation

$$\delta\vec{a}^T\left(\int_V \hat{\mathbf{B}}^T\vec{\sigma}\,dV - \vec{f}\right) = 0$$

from which we have the equation of equilibrium

$$\vec{f} = \int_V \hat{\mathbf{B}}^T\vec{\sigma}\,dV \tag{2}$$

To calculate strain displacement relation (1), we use the general nonlinear description of strain, the Green-Lagrange strain tensor, defined with respect to the cartesian coordinate system x,y,z and the displacements u,v,w :

$$\varepsilon_x = \frac{\partial u}{\partial x} + \frac{1}{2}\frac{\partial u}{\partial x}\frac{\partial u}{\partial x} + \frac{1}{2}\frac{\partial v}{\partial x}\frac{\partial v}{\partial x} + \frac{1}{2}\frac{\partial w}{\partial x}\frac{\partial w}{\partial x} \tag{3a}$$

$$\gamma_{yz} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} + \frac{\partial u}{\partial y}\frac{\partial u}{\partial z} + \frac{\partial v}{\partial y}\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\frac{\partial w}{\partial z} \tag{3b}$$

The other components of the strain tensor $\vec{\varepsilon} = (\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{yz}, \gamma_{zx}, \gamma_{xy})$ are obtained by cyclic permutation.

The total variation of the equations (3) give :

$$d\varepsilon_x = d\left(\frac{\partial u}{\partial x}\right) + \frac{\partial u}{\partial x}d\left(\frac{\partial u}{\partial x}\right) + \frac{\partial v}{\partial x}d\left(\frac{\partial v}{\partial x}\right) + \frac{\partial w}{\partial x}d\left(\frac{\partial w}{\partial x}\right) \tag{4a}$$

$$d\gamma_{yz} = d\left(\frac{\partial v}{\partial z}\right) + d\left(\frac{\partial w}{\partial y}\right) + \frac{\partial u}{\partial z}d\left(\frac{\partial u}{\partial y}\right) + \frac{\partial u}{\partial y}d\left(\frac{\partial u}{\partial z}\right) + \frac{\partial v}{\partial z}d\left(\frac{\partial v}{\partial y}\right) + \frac{\partial v}{\partial y}d\left(\frac{\partial v}{\partial z}\right) + \frac{\partial w}{\partial z}d\left(\frac{\partial w}{\partial y}\right) + \frac{\partial w}{\partial y}d\left(\frac{\partial w}{\partial z}\right) \tag{4b}$$

Equations (4) and the corresponding terms can be summarized in the following formula :

$$d\vec{\varepsilon} = \left(\mathbf{A}_0 + \mathbf{A}(\vec{\theta})\right)d\vec{\theta} \tag{5}$$

where the (6x9)-matrices $\mathbf{A}_o$ and $\mathbf{A}(\vec{\theta})$ are defined as follows :

$$\mathbf{A}_0 = \begin{bmatrix} \vec{e}_x^T & o & o \\ o & \vec{e}_y^T & o \\ o & o & \vec{e}_z^T \\ o & \vec{e}_z^T & \vec{e}_y^T \\ \vec{e}_z^T & o & \vec{e}_x^T \\ \vec{e}_y^T & \vec{e}_x^T & o \end{bmatrix}, \quad \mathbf{A}(\vec{\theta}) = \begin{bmatrix} \vec{\theta}_x^T & o & o \\ o & \vec{\theta}_y^T & o \\ o & o & \vec{\theta}_z^T \\ o & \vec{\theta}_z^T & \vec{\theta}_y^T \\ \vec{\theta}_z^T & o & \vec{\theta}_x^T \\ \vec{\theta}_y^T & \vec{\theta}_x^T & o \end{bmatrix}, \quad \begin{aligned} \vec{e}_x^T &= (1,0,0), \ldots. \\ \vec{\theta}_x^T &= \left(\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial w}{\partial x}\right), \ldots. \\ \vec{\theta}^T &= (\vec{\theta}_x^T, \vec{\theta}_y^T, \vec{\theta}_z^T) \end{aligned} \tag{5a}$$

In the finite element displacement-method, the displacement at a point inside the element is expressed in terms of the element nodal-displacement vector $\vec{a}_e$, using interpolation functions which depend on the specific element type. The deformations $\vec{\theta}$ are then calculated from the displacements. We define the following interpolation relation :

$$\vec{\theta} = \mathbf{G}\,\vec{a}_e \tag{6}$$

The equations (5) and (6) finally lead to the the matrix $\mathbf{B}$ analogous to $\hat{\mathbf{B}}$ in equation (1) :

$$d\vec{\varepsilon} = (\mathbf{A}_0\,\mathbf{G} + \mathbf{A}(\vec{\theta})\,\mathbf{G})\,d\vec{a}_e = (\mathbf{B}_0 + \mathbf{B}_L(\vec{\theta}))\,d\vec{a}_e = \mathbf{B}\,d\vec{a}_e \tag{7}$$

Following the notation utilized in [1] we now deduce the element tangent-stiffness matrix $\mathbf{K}_{T_E}$, which relates the increment of the element nodal-force vector $\vec{f}_e$ to the increment of the element nodal-displacement vector $\vec{a}_e$.

The variation of the equation of equilibrium (2) gives :

$$d\vec{f}_e = \int_{Ve} d\mathbf{B}^T\vec{\sigma}\,dV + \int_{Ve}\mathbf{B}^T d\vec{\sigma}\,dV = \mathbf{K}_{T_E}d\vec{a}_e \tag{8}$$

where Ve is the element volume.

The variation of stress $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z, \tau_{yz}, \tau_{zx}, \tau_{xy})$ is obtained from the variation of strain by the nonlinear incremental constitutive relation :

$$d\vec{\sigma} = \mathbf{C}_T(\vec{\varepsilon})d\vec{\varepsilon} = (\mathbf{C} + \mathbf{C}_D(\vec{\varepsilon}))d\vec{\varepsilon} \tag{9}$$

Equations (7) and (9) lead to the transformation of the right-hand integral in equation (8)

$$\int_{Ve}\mathbf{B}^T d\vec{\sigma}\,dV = \left(\int_{Ve}\mathbf{B}^T\mathbf{C}_T\mathbf{B}\,dV\right)d\vec{a}_e = \left(\int_{Ve}(\mathbf{B}_0 + \mathbf{B}_L)^T\,\mathbf{C}_T\,(\mathbf{B}_0 + \mathbf{B}_L)\,dV\right)d\vec{a}_e = (\mathbf{K}_{0_E} + \mathbf{K}_{L_E})\,d\vec{a}_e \tag{10}$$

where $\mathbf{K}_{0_E}$ stands for the normal element stiffness matrix and $\mathbf{K}_{L_E}$ for the element initial-deformation matrix :

$$\mathbf{K}_{0_E} = \int_{Ve} \mathbf{B}_0^T \mathbf{C}_T \mathbf{B}_0 \, dV \quad , \quad \mathbf{K}_{L_E} = \int_{Ve} (\mathbf{B}_0^T \mathbf{C}_T \mathbf{B}_L + \mathbf{B}_L^T \mathbf{C}_T \mathbf{B}_L + \mathbf{B}_L^T \mathbf{C}_T \mathbf{B}_0) \, dV \tag{10a}$$

The left-hand integral in equation (8) takes into account initial stress $\vec{\sigma}$. The matrix $\mathbf{G}$ in equation (6) depends only on constant geometric element properties, so we obtain with equation (7) :

$$\int_{Ve} d\mathbf{B}^T \vec{\sigma} \, dV = \int_{Ve} d(\mathbf{B}_L(\vec{\theta}))^T \vec{\sigma} \, dV = \int_{Ve} \mathbf{G}^T d(\mathbf{A}(\vec{\theta}))^T \vec{\sigma} \, dV \tag{11}$$

We set :

$$d(\mathbf{A}(\vec{\theta}))^T \vec{\sigma} = \mathbf{S} d\vec{\theta} \tag{12}$$

It is not difficult to verify the following relation concerning the (9x9)-matrix $\mathbf{S}$, using the equation (5a) :

$$\mathbf{S} = \begin{bmatrix} \sigma_x \mathbf{E}_3 & \tau_{xy} \mathbf{E}_3 & \tau_{zx} \mathbf{E}_3 \\ & \sigma_y \mathbf{E}_3 & \tau_{yz} \mathbf{E}_3 \\ sym. & & \sigma_z \mathbf{E}_3 \end{bmatrix} \quad , \quad \mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12a}$$

Equations (11), (12) and (6) finally lead to a symmetric matrix, the element initial-stress matrix $\mathbf{K}_{\sigma_E}$:

$$\int_{Ve} d\mathbf{B}^T \vec{\sigma} \, dV = \int_{Ve} \mathbf{G}^T \mathbf{S} \, d\vec{\theta} \, dV = \left(\int_{Ve} \mathbf{G}^T \mathbf{S} \mathbf{G} \, dV \right) d\vec{a}_e \quad , \quad \mathbf{K}_{\sigma_E} = \int_{Ve} \mathbf{G}^T \mathbf{S} \mathbf{G} \, dV \tag{13}$$

With the normal element stiffness matrix $\mathbf{K}_{0_E}$ and the element initial-deformation matrix $\mathbf{K}_{L_E}$ (both defined in equation (10a)) and with the element initial-stress matrix $\mathbf{K}_{\sigma_E}$ (defined in equation (13)), we have found an expression for the symmetric element tangent-stiffness matrix $\mathbf{K}_{T_E}$, which was defined in equation (8) :

$$d\vec{f}_e = (\mathbf{K}_{0_E} + \mathbf{K}_{L_E} + \mathbf{K}_{\sigma_E}) \, d\vec{a}_e = \mathbf{K}_{T_E} \, d\vec{a}_e \tag{14}$$

In a next step we elaborate a specific element type in order to get the interpolation matrix $\mathbf{G}$ defined in equation (6), and we also have to execute the integration over the element volume V_e .

Isoparametric triangular-prismatic element

The right prism shown in Figure 1 is the starting point for generating the three-dimensional isoparametric triangular-prismatic element defined with respect to the local coordinates ξ, η, ζ. We execute an isoparametric transformation on this right prism to obtain, with respect to the global coordinates x, y, z, an arbitrary prism with warped surfaces, shown in Figure 2.

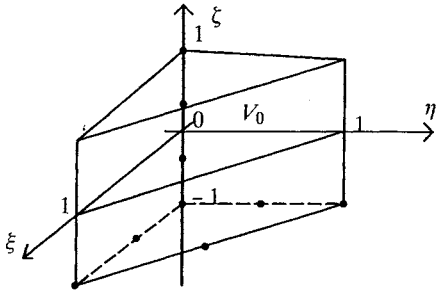


Fig. 1 : Right prism with unit-triangle base area and height 2

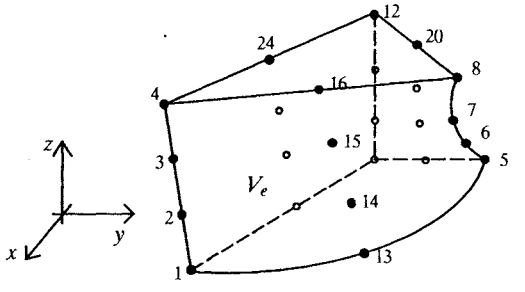


Fig. 2 : Arbitrary prism with warped surfaces and 24 nodes

The 24 nodes marked on Figure 2 enable a quadratic interpolation in the triangular domain and a cubic interpolation in the normal direction. Quadratic interpolation functions D_i belonging to the unit-triangular domain (see e.g. [2]) and the more general quintic Lagrange normal interpolation functions L_i for the $(-1, +1)$ domain are given below

$$D_1 = (1-\xi-\eta)(1-2\xi-2\eta)$$
$$D_2 = \xi \, (2\xi - 1)$$
$$D_3 = \eta \, (2\eta - 1)$$
$$D_4 = 4\xi \, (1-\xi-\eta)$$
$$D_5 = 4\xi\eta$$
$$D_6 = 4\eta \, (1-\xi-\eta)$$

$$L_1 = 625 \, (1-\zeta)(\zeta^2 - 9/25)(\zeta^2 - 1/25)/768$$
$$L_2 = 3125 \, (\zeta - 3/5)(\zeta^2 - 1)(\zeta^2 - 1/25)/768$$
$$L_3 = 3125 \, (1/5 - \zeta)(\zeta^2 - 1)(\zeta^2 - 9/25)/384$$
$$L_4 = 3125 \, (1/5 + \zeta)(\zeta^2 - 1)(\zeta^2 - 9/25)/384$$
$$L_5 = 3125 \, (\zeta + 3/5)(1-\zeta^2)(\zeta^2 - 1/25)/768$$
$$L_6 = 625 \, (1+\zeta)(\zeta^2 - 9/25)(\zeta^2 - 1/25)/768$$

Lagrange interpolation polynomials L_i of order p other than five $(i = 1, \ldots, p+1)$ are found straight away, considering the fact that the value of $L_i(\zeta = (2j-p-2)/p)$ equals 1 for $i = j$ and 0 otherwise.

Volume interpolation functions N_i for the right prism involve the product of the D_i and L_i :

$$N_i(\xi,\eta,\zeta) = D_n(\xi,\eta) \cdot L_m(\zeta) \quad , \quad i = 1, \ldots, k \quad , \quad k = 6(p+1) \quad , \quad p = 1,2,3,4,or\ 5 \tag{15}$$
$$n = 1 + Int\,((i-1)/(p+1)) \quad , \quad n = 1, \ldots, 6$$
$$m = 1 + Mod((i-1),(p+1)) \quad , \quad m = 1, \ldots, p+1$$

The coefficients n and m result from the numeration of nodes as shown in Figure 2. The number of nodes is k. Again the value of $N_i(\xi,\eta,\zeta)$ equals 1 for coordinates ξ,η,ζ corresponding to the node i, and 0 for all other nodes.

Three interpolation $(3k)$ -vectors are defined with equation (15) :

$$\begin{aligned}
\vec{N}_1^T &= (N_1,0,0,N_2,0,0,\ldots\ldots & ,N_k,0,0) \\
\vec{N}_2^T &= (0,N_1,0,0,N_2,0,\ldots\ldots & ,0,N_k,0) \\
\vec{N}_3^T &= (0,0,N_1,0,0,N_2,\ldots\ldots & ,0,0,N_k)
\end{aligned} \tag{16}$$

We further define the element nodal-displacement $(3k)$ -vector $\vec{a}_e$ and the element nodal-coordinate $(3k)$ -vector $\vec{c}_e$:

$$\begin{aligned}
\vec{a}_e^T &= (u_1,v_1,w_1,u_2,v_2,w_2,\ldots\ldots & ,u_k,v_k,w_k) \\
\vec{c}_e^T &= (x_1,y_1,z_1,x_2,y_2,z_2,\ldots\ldots & ,x_k,y_k,z_k)
\end{aligned} \tag{17}$$

With the interpolation functions (16) and with (17) we get the displacement functions (18) and with the same interpolation functions (isoparametric transformation) the substitution of variables (19) :

$$u(\xi,\eta,\zeta) = \vec{a}_e^T \vec{N}_1(\xi,\eta,\zeta) \quad , \quad v(\xi,\eta,\zeta) = \vec{a}_e^T \vec{N}_2(\xi,\eta,\zeta) \quad , \quad w(\xi,\eta,\zeta) = \vec{a}_e^T \vec{N}_3(\xi,\eta,\zeta) \tag{18}$$

$$x(\xi,\eta,\zeta) = \vec{c}_e^T \vec{N}_1(\xi,\eta,\zeta) \quad , \quad y(\xi,\eta,\zeta) = \vec{c}_e^T \vec{N}_2(\xi,\eta,\zeta) \quad , \quad z(\xi,\eta,\zeta) = \vec{c}_e^T \vec{N}_3(\xi,\eta,\zeta) \tag{19}$$

Equation (19) makes possible to calculate, with the inverted Jacobi-matrix $\mathbf{J}$, the differentiations with respect to the global coordinates x,y,z of the interpolation functions defined in (15) :

$$\begin{bmatrix} \frac{\partial N_i}{\partial x} \\ \frac{\partial N_i}{\partial y} \\ \frac{\partial N_i}{\partial z} \end{bmatrix} = \mathbf{J}^{-1} \begin{bmatrix} \frac{\partial N_i}{\partial \xi} \\ \frac{\partial N_i}{\partial \eta} \\ \frac{\partial N_i}{\partial \zeta} \end{bmatrix} \quad , \quad \mathbf{J} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} & \frac{\partial z}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} & \frac{\partial z}{\partial \eta} \\ \frac{\partial x}{\partial \zeta} & \frac{\partial y}{\partial \zeta} & \frac{\partial z}{\partial \zeta} \end{bmatrix} = \begin{bmatrix} \vec{c}_e^T \frac{\partial \vec{N}_1}{\partial \xi} & \vec{c}_e^T \frac{\partial \vec{N}_2}{\partial \xi} & \vec{c}_e^T \frac{\partial \vec{N}_3}{\partial \xi} \\ \vec{c}_e^T \frac{\partial \vec{N}_1}{\partial \eta} & \vec{c}_e^T \frac{\partial \vec{N}_2}{\partial \eta} & \vec{c}_e^T \frac{\partial \vec{N}_3}{\partial \eta} \\ \vec{c}_e^T \frac{\partial \vec{N}_1}{\partial \zeta} & \vec{c}_e^T \frac{\partial \vec{N}_2}{\partial \zeta} & \vec{c}_e^T \frac{\partial \vec{N}_3}{\partial \zeta} \end{bmatrix} \tag{20}$$

From equation (16) we can define, with equation (20), nine interpolation $(3k)$ -vectors $\partial\vec{N}_1/\partial x$, ... , $\partial\vec{N}_3/\partial z$, and, according to equation (18), we have nine differentiated displacement functions

$$\frac{\partial u}{\partial x}(\xi,\eta,\zeta) = \vec{a}_e^T \frac{\partial \vec{N}_1}{\partial x} \quad , \quad \frac{\partial u}{\partial y}(\xi,\eta,\zeta) = \vec{a}_e^T \frac{\partial \vec{N}_1}{\partial y} \quad , \ldots \quad , \quad \frac{\partial w}{\partial z}(\xi,\eta,\zeta) = \vec{a}_e^T \frac{\partial \vec{N}_3}{\partial z}$$

Thus, finally, the $(9x\,3k)$ -matrix $\mathbf{G}$ in equation (6) is :

$$\mathbf{G} = \left[\frac{\partial \vec{N}_1^T}{\partial x} \; \frac{\partial \vec{N}_2^T}{\partial x} \; \frac{\partial \vec{N}_3^T}{\partial x} \; \frac{\partial \vec{N}_1^T}{\partial y} \; \frac{\partial \vec{N}_2^T}{\partial y} \; \frac{\partial \vec{N}_3^T}{\partial y} \; \frac{\partial \vec{N}_1^T}{\partial z} \; \frac{\partial \vec{N}_2^T}{\partial z} \; \frac{\partial \vec{N}_3^T}{\partial z} \right]^T \tag{21}$$

The equations (5), (7) and (21) give expressions for $\mathbf{B}_0$ and $\mathbf{B}_L$, and with $\mathbf{S}$ defined in equation (12a), we know all the required matrices to evaluate the element tangent-stiffness matrix $\mathbf{K}_{T_E}$ defined in equation (14).

We still have to execute the integration over the element volume V_e (Fig. 2). The substitution (19) with the resulting Jacobi determinant $\det \mathbf{J}$ leads to the transformation of integrals. For example, we have :

$$\mathbf{K}_{0_E} = \int_{Ve} \mathbf{B}_0^T(x,y,z)\; \mathbf{C}_T\; \mathbf{B}_0(x,y,z)\; dx\,dy\,dz = \int_{V_0} \mathbf{B}_0^T(\xi,\eta,\zeta)\; \mathbf{C}_T\; \mathbf{B}_0(\xi,\eta,\zeta)\; \det \mathbf{J}\; d\xi\,d\eta\,d\zeta \tag{22}$$

The numerical integration over the domaine V_0 (Fig. 1) is carried out by means of the Gaussian quadrature formulas [1]. Three-dimensional numerical integration may imply a considerable amount of computation. In the linear case it is possible to integrate just nine different dyadics $(\partial\vec{N}/\partial x)\,(\partial\vec{N}/\partial x)^T, \ldots$. and to multiply them appropriately with the constant material-matrix $\mathbf{C}$ (eq. (9)) outside the integration loop. This is not possible in the nonlinear case, where we have to integrate complicated sums of dyadics.

In order to keep the amount of computation as low as possible, we must carefully analyse the degree of the Gaussian quadrature required for a precise integration. Considering the terms $\mathbf{B}_L$ (eq. (7)) and $\mathbf{C}_D$ (eq. 9)) as second-order terms, we obtain the following approximation, neglecting all terms of order higher than two :

$$\mathbf{K}_{0_E} + \mathbf{K}_{L_E} = \int_{Ve} (\mathbf{B}_0 + \mathbf{B}_L)^T (\mathbf{C} + \mathbf{C}_D)(\mathbf{B}_0 + \mathbf{B}_L)\; dV \cong \int_{Ve} (2\,\mathbf{B}_0^T \mathbf{C}\,\mathbf{B}_L + \mathbf{B}_0^T \mathbf{C}_D \mathbf{B}_0 + \mathbf{B}_0^T \mathbf{C}\,\mathbf{B}_0)\; dV \tag{23}$$

The initial deformations $\vec{\theta}$ vary in ξ,η,ζ like the terms $\mathbf{G}$ or $\mathbf{B}_0$ (the same is true for $\mathbf{C}_D$), hence $\mathbf{B}_L$ varies with a power twice as high as $\mathbf{B}_0$ (eq. (6), (7)). The integrand in (23) thus varies in ξ,η,ζ with a power three-times as high as $\mathbf{G}$. The

same is true for the integrand in equation (13). We further have to consider the variability of $\det \mathbf{J}$ in equation (22). With a seventh order integration in ξ, η (quadratic displacement interpolation in ξ, η yields quadratic interpolation of the $\partial N_i/\partial\zeta$ in eq.(15) !) and an order of $2p + 5$ in ζ, all terms of first and of second order (at least) are integrated accurately. In the linear case (and right prism elements, i.e. constant $\det \mathbf{J}$), an integration order of four and of $2p$ respectively is sufficient.

In order to calculate free vibrations we also need the element mass matrix $\mathbf{M}_e$. Calculating the kinetic energy, we immediately find, with the angular velocity ω and the specific gravity ϱ, the following expression

$$E_{kin,E} = \frac{\omega^2}{2} \int_{Ve} \varrho\,(u^2 + v^2 + w^2)\, dV = \frac{\omega^2}{2}\, \vec{a}_e^{\,T}\, \mathbf{M}_e\, \vec{a}_e$$

With equation (18), we obtain the symmetric element mass matrix :

$$\mathbf{M}_e = \int_{Vo} \varrho(\xi,\eta,\zeta)\,(\vec{N}_1\vec{N}_1^T + \vec{N}_2\vec{N}_2^T + \vec{N}_3\vec{N}_3^T)\,\det \mathbf{J}\, d\xi\, d\eta\, d\zeta \tag{24}$$

The integration order is the same as for the tangent-stiffness matrix.
The last point to discuss in this section is the term $\Delta\mathbf{M}_e$, caused by thin electrodes of density μ (mass/area) positioned on the lower and/or upper triangle area, which contributes to the element mass matrix :

$$\Delta\mathbf{M}_e = \mu \int\int_{Go} (\vec{N}_1\vec{N}_1^T + \vec{N}_2\vec{N}_2^T + \vec{N}_3\vec{N}_3^T)_{\zeta=\pm1}\, |\vec{A}_{\zeta=\pm1}|\, d\xi\, d\eta \tag{25}$$

The integration area G_0 is the unit-triangle (Fig.1), and the vector $\vec{A}$ is the following vector cross product :

$$\vec{A} = \left(\frac{\partial x}{\partial\xi}, \frac{\partial y}{\partial\xi}, \frac{\partial z}{\partial\xi}\right)^T \times \left(\frac{\partial x}{\partial\eta}, \frac{\partial y}{\partial\eta}, \frac{\partial z}{\partial\eta}\right)^T \tag{25a}$$

where the separate terms are calculated according to equation (19).

Determination of the eigenfrequency first-order perturbation

We first have to determine, with the finite element method, the unperturbed eigenvalue λ_n and eigenvector $\vec{u}_n$,belonging to the N-dimensional eigenproblem :

$$\mathbf{K}\vec{u}_n = \lambda_n \mathbf{M}\vec{u}_n \quad , \qquad n = 1, \ldots N \tag{26}$$

with the symmetric global stiffness matrix $\mathbf{K}$ and the symmetric global mass matrix $\mathbf{M}$. The eigenvalue and the eigenvector are calculated by the subspace iteration method [2]. We obtain convergence to modes with eigenvalues adjacent to value of the negative shift β, by solving the associated problem :

$$(\mathbf{K} - \beta\mathbf{M})\,\vec{u}_n = (\lambda_n - \beta)\,\mathbf{M}\vec{u}_n \tag{26a}$$

where the Gauss method is used to factorize the indefinite matrix $(\mathbf{K} - \beta\mathbf{M})$ into $\mathbf{L}\,\mathbf{D}\,\mathbf{L}^T$. With the Sturm sequence property [3], yielding that the number of negative elements of the diagonal matrix $\mathbf{D}$ equals the number of eigenvalues smaller than β, we are able to give the mode-number n of the calculated eigenvalue. The eigenvectors fulfil the relation :

$$\vec{u}_n^{\,T}\,\mathbf{M}\,\vec{u}_m = \begin{cases} 1\,, & n = m \\ 0\,, & n \neq m \end{cases} \tag{27}$$

i.e. the calculated eigenvectors are $\mathbf{M}$ -orthonormalized.

Secondly we have to perform a finite element linear static analysis. We calculate, depending on the static perturbation forces, the nodal initial displacements and the initial strains and rotations in the Gauss-points of the elements. From this we calculate, as shown in detail in the next section, the perturbed stiffness matrix $\mathbf{K}'$ and mass matrix $\mathbf{M}'$.

Finally, we must determine the eigenfrequency of the initially deformed resonator. The principle of virtual work, equating the virtual deformation energy to the virtual work done by inertia forces, is also valid in the initially deformed state. Thus, the equation of motion for free vibrations (26) is the determining equation also in the state of initial deformation.
We calculate the resulting first order change in eigenfrequency with the Schrödinger method of perturbation [4].The N eigenvectors of the eigenproblem (26) form a complete set of base vectors. So, the following expression yields for the perturbed eigenvector $\vec{u}_n'$:

$$\vec{u}_n' = \vec{u}_n + \sum_{\substack{i=1 \\ i \neq n}}^{N} a_i\,\vec{u}_i \quad , \qquad a_i \ll 1 \tag{28}$$

We have, with $\lambda_n' = \lambda_n + \Delta\lambda_n$ and with equation (26)

$$\mathbf{K}'\,\vec{u}_n' = (\lambda_n + \Delta\lambda_n)\,\mathbf{M}'\,\vec{u}_n'$$

From equation (28) and with $\mathbf{K}' = \mathbf{K} + \Delta\mathbf{K}$, $\mathbf{M}' = \mathbf{M} + \Delta\mathbf{M}$ we find, neglecting all perturbation terms of order higher than one, the following relation

$$(\mathbf{K} + \Delta\mathbf{K})\vec{u}_n + \sum_{i \neq n} a_i \, \mathbf{K}\vec{u}_i = \lambda_n(\mathbf{M} + \Delta\mathbf{M})\vec{u}_n + \lambda_n \sum_{i \neq n} a_i \, \mathbf{M}\vec{u}_i + \Delta\lambda_n \, \mathbf{M}\vec{u}_n$$

Scalar multiplication of the above equation with $\vec{u}_n^T$ leads, with equation (27), to

$$\vec{u}_n^T \mathbf{K}' \vec{u}_n = \lambda_n \vec{u}_n^T \mathbf{M}' \vec{u}_n + \Delta\lambda_n$$

With $\lambda + \Delta\lambda \cong \omega^2 + 2\omega\Delta\omega$ we obtain :

$$\Delta\omega_n = \frac{\vec{u}_n^T(\mathbf{K}' - \lambda_n \mathbf{M}')\,\vec{u}_n}{2\omega_n} \tag{29}$$

Equation (29) gives a simple expression, which allows us to calculate the first-order change in angular velocity.

Calculation of the perturbed stiffness and mass matrix

The state of the initially deformed resonator is characterized by the perturbed matrices $\mathbf{K}'$ and $\mathbf{M}'$. Small oscillating deformations superposed on the initial static deformations are interpreted as incremental deformations. Hence the perturbed stiffness matrix is the tangent-stiffness matrix $\mathbf{K}_T'$ calculated from equation (14), and $\mathbf{M}'$ is calculated from equation (24), both with modified properties.

Considering the effects of initial deformation, we take into account only first order perturbation terms !

The linear static analysis yields initial element nodal deformations $\vec{a}_e$. We first calculate the perturbed element nodal coordinates $\vec{c}_e'$:

$$\vec{c}_e' = \vec{c}_e + \vec{a}_e \tag{30}$$

which are used in equation (19)

Linear static finite element analysis also yields initial deformations $\vec{\theta}$ at the element Gauss-points (integration points) :

$$\vec{\theta}^T = \left(\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}, \frac{\partial w}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}, \frac{\partial w}{\partial y}, \frac{\partial u}{\partial z}, \frac{\partial v}{\partial z}, \frac{\partial w}{\partial z} \right) \tag{31}$$

With equations (5a) and (7) we obtain the matrix $\mathbf{B}_L$ at the Gauss-points.

From the initial deformations we calculate the initial strain $\vec{\varepsilon}$:

$$\varepsilon_1 = \frac{\partial u}{\partial x}, \quad \varepsilon_2 = \frac{\partial v}{\partial y}, \quad \varepsilon_3 = \frac{\partial w}{\partial z}, \quad \varepsilon_4 = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \quad \varepsilon_5 = \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}, \quad \varepsilon_6 = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \tag{32}$$

From the initial strain $\vec{\varepsilon}$ we get, corresponding to equation (9), (only first order terms !) the initial stress $\vec{\sigma}$:

$$\vec{\sigma} = \mathbf{C}\,\vec{\varepsilon} \tag{33}$$

and with equation (12a) we obtain the matrix $\mathbf{S}$ at the Gauss-points.

Now we have to derive the perturbed elastic constants. Anisotropic elastic constants, depending on the cutting angle Θ for singly rotated, and (Φ, Θ) for doubly rotated crystal plates (see Fig. 3), are transformed with the orthogonal matrix $\mathbf{B}$ to the rotated coordinates :

$$\mathbf{B} = \begin{bmatrix} \cos\Psi & \sin\Psi & 0 \\ -\sin\Psi & \cos\Psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Theta + \pi/2) & \sin(\Theta + \pi/2) \\ 0 & -\sin(\Theta + \pi/2) & \cos(\Theta + \pi/2) \end{bmatrix} \begin{bmatrix} \cos\Phi & \sin\Phi & 0 \\ -\sin\Phi & \cos\Phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{34}$$

For the azimuth angle Ψ, see Figure 3.

The dependence of the elastic constants in function of the initial strain $\vec{\varepsilon}$ is expressed by third-order elastic coefficients C_{ijk}, forming a symmetric sixth-rank (6x6x6)-tensor. By definition, the elements of the symmetric fourth-rank (6x6)-tensor $\mathbf{C}_T$ in equation (9) are :

$$C_{Tij}^{o} = C_{ij}^{o} + \sum_{k=1}^{6} C_{ijk}^{o} \, \varepsilon_k^{o} \tag{35}$$

The supscript " ^{o} " means : with respect to the non-rotated coordinates . With equation (34), we calculate the initial strain $\vec{\varepsilon}^o$ from $\vec{\varepsilon}$ in equation (32) by the following second-rank tensor transformation :

$$\begin{bmatrix} \varepsilon_1^o & \tfrac{1}{2}\varepsilon_6^o & \tfrac{1}{2}\varepsilon_5^o \\ & \varepsilon_2^o & \tfrac{1}{2}\varepsilon_4^o \\ sym & & \varepsilon_3^o \end{bmatrix} = \mathbf{B}^t \begin{bmatrix} \varepsilon_1 & \tfrac{1}{2}\varepsilon_6 & \tfrac{1}{2}\varepsilon_5 \\ & \varepsilon_2 & \tfrac{1}{2}\varepsilon_4 \\ sym & & \varepsilon_3 \end{bmatrix} \mathbf{B} \tag{35a}$$

Also, we have to consider the rotations caused by large displacements. From the initial deformations in equation (31), the rotations $\vec{\varphi}$ are calculated in order to obtain the matrix $\mathbf{F}$:

$$\varphi_x = \frac{1}{2}\left(\frac{\partial v}{\partial z} - \frac{\partial w}{\partial y}\right), \quad \varphi_y = \frac{1}{2}\left(\frac{\partial w}{\partial x} - \frac{\partial u}{\partial z}\right), \quad \varphi_z = \frac{1}{2}\left(\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}\right), \quad \mathbf{F} = \begin{bmatrix} 1 & \varphi_z & -\varphi_y \\ -\varphi_z & 1 & \varphi_x \\ \varphi_y & -\varphi_x & 1 \end{bmatrix} \tag{36}$$

where the matrix $\mathbf{F}$ is orthogonal in first-order. From this we obtain the perturbed transformation matrix $\mathbf{B}'$:

$$\mathbf{B}' = \mathbf{F}\,\mathbf{B} \tag{36a}$$

calculated at the Gauss-points.

To get the perturbed rotated elastic coefficients at the Gauss-points, we have to execute the following fourth-rank tensor transformation on the tensor defined in equation (35) :

$$\mathbf{C}_T' = \mathbf{Q}'\,\mathbf{C}_T^{o}\,\mathbf{Q}'^{T} \tag{37}$$

We obtain the perturbed (6x6)-matrix $\mathbf{Q}$ [5], composed of four submatrices $\mathbf{Q}_N$, from the perturbed $\mathbf{B}$ (eq. (36a)) :

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_3 & \mathbf{Q}_4 \end{pmatrix}, \quad \begin{matrix} Q_{1ij} = B_{ij}\,B_{ij} & Q_{3ij} = B_{kj}\,B_{lj} & k = mod(i,3)+1 & m = mod(j,3)+1 \\ Q_{2ij} = B_{im}\,B_{in} & Q_{4ij} = B_{km}\,B_{ln} + B_{kn}\,B_{lm} & l = mod(i+1,3)+1 & n = mod(j+1,3)+1 \end{matrix} \tag{37a}$$

where i, j = 1,2,3.

With equations (30)-(37) we are able to calculate the perturbed element stiffness matrices $\mathbf{K}_{T_E}'$ in equation (14).

From the initial strain in equation (32), we finally obtain the perturbed specific gravity ϱ' at the Gauss-points :

$$\varrho' = \varrho\,(1-\varepsilon_1-\varepsilon_2-\varepsilon_3) \tag{38}$$

With equations (30) and (38), the perturbed element mass matrices $\mathbf{M}_e'$ of equation (24) can be calculated.

In the case of large initial displacements, the perturbation terms of equation (30) and equation (36) may produce a distinct nonlinearity in the frequency-force characteristic. Thus, to eliminate these higher-order perturbation terms, the static displacements must be limited. We multiply the initial displacements and deformations with the small factor n_o :

$$n_0 = \frac{max.di}{max.am}\;10^{-8} \tag{39}$$

with the maximum value of the resonator node-coordinates $max.di$ and with the maximum value of the nodal-displacements $max.am$, resulting from the imposed perturbation forces. The frequency change, obtained from equation (29), is then multiplied with the large factor $1/n_o$. In order to obtain zero frequency perturbation with zero perturbation forces, the degree of the Gauss-quadrature has then to be chosen, when calculating the unperturbed eigenmode, the same as in the nonlinear case! The calculations are carried out with an eight byte precision. The factor 10^{-8} in equation (39) can be varied between 10^{-5} and 10^{-11}, with changes in the resulting frequency perturbation smaller than a hundredth of a percent.

Numerical results

Anisotropic second-order elastic constants for trigonal quartz and third-order elastic constants (14 independent, 17 dependent, 25 zero), both referred to the crystallographic axis x_1,x_2,x_3 , are listed in [6]. Matrix $\mathbf{B}$ of equation (34) defines the transformation of the elastic constants from the coordinates x_1,x_2,x_3 to the coordinates x,y,z shown in Figures 3 and 4.

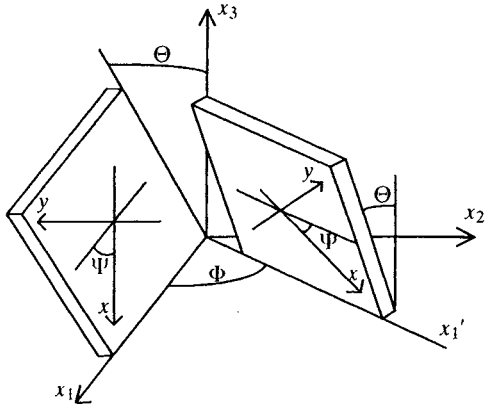


Fig. 3 : Singly and doubly rotated crystal plates

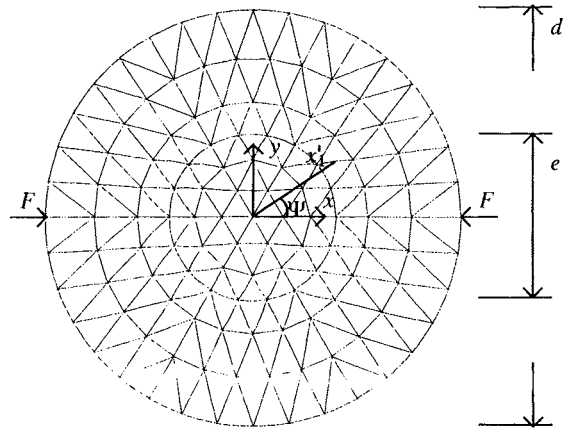


Fig. 4 : Finite element mesh. Diametrical perturbation forces F

Figure 4 shows the finite element discretisation of a quartz disk, with the total diameter d , the diameter e of the electroded area, and with the azimuth angle Ψ between the working-coordinate x -axis and the cristallographic x_1' -axis.

We first present results obtained with opposite perturbation forces F acting along the x -axis (Fig.4). The thickness of the quartz disc is $0.300\ mm$, the diameter $d = 15.0\ mm$ and the electrode diameter $e = 6.0\ mm$.

The AT cut has $\Theta = 33.90°$, an electrode mass density $\mu = 0.5\ mg/cm^2$ (both sides), and a resulting fundamental thickness-shear mode frequency of 5.4638 MHz and 16.517 MHz, respectively, for the third-overtone thickness-shear mode.

The SC cut has $\Phi = 21.90°$, $\Theta = 33.90°$ and an electrode mass density $\mu = 0.3\ mg/cm^2$. The resulting fundamental thickness-shear mode frequency is 5.9586 MHz.
In Figures 5, 6 and 7 the calculated normalized values of the frequency perturbations Δf are represented, i.e. values of the force sensitivity coefficient K_f [7], defined as :

$$K_f = \frac{\Delta f}{f} \frac{n\ d}{f\ F}$$

with the resonance frequency f , and the order n of the harmonic overtone. d is the disc diameter and F is the applied force.

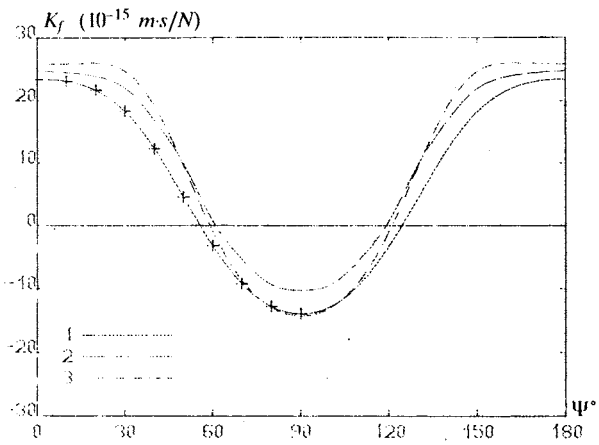


Fig. 5 : K_f vs Ψ for AT cut, fundamental mode
1 : present , 2 : experimental [7] , 3 : analytical [8]

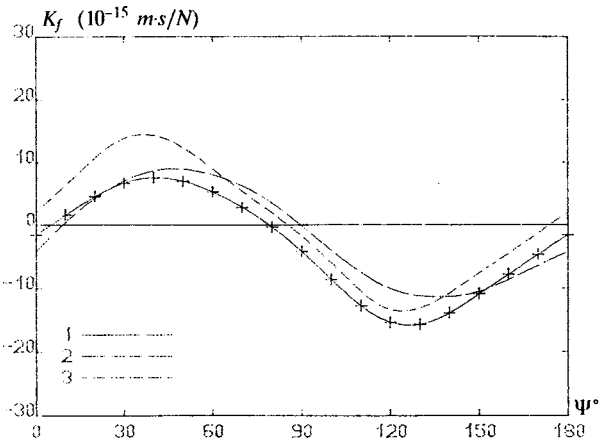


Fig. 6 : K_f vs Ψ for SC cut, fundamental mode
1 : present , 2 : experimental [9] , 3 : analytical [8]

The number of degrees of freedom for the fundamental mode (cubic interpolation in z) is 5628 , respectively, 8442 for the third harmonic overtone mode (quintic interpolation in z).
Results for the AT cut overtone mode are represented in Figure 7.

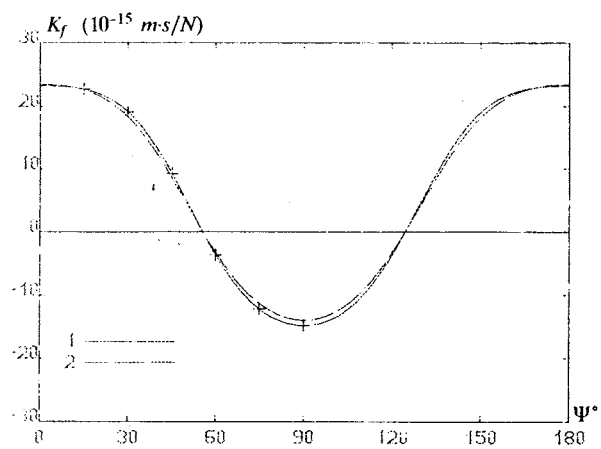


Fig. 7 : K_f vs Ψ for AT cut, third harmonic overtone
1 : overtone mode, 2 : fundamental mode

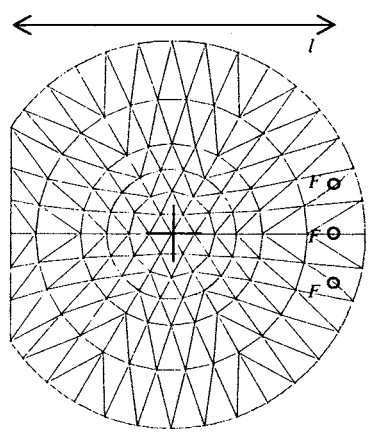


Fig. 8 : Finite element mesh. Cantilever bending.
Transversal perturbation forces F .

The effects of transverse forces have also been investigated. Figure 8 shows schematically the arrangement of cantilever bending and the finite element mesh. The quartz disc is clamped at the lefthand vertical side. At the distance l transversally directed perturbation forces F act on the disc .
The value of the three nodal-forces $F = 6.54\,mN$ (total 2 gf), and $l = 10\,mm$. The diameter of the quartz disc is $12\,mm$,and the electrode area diameter is $4\,mm$ with an electrode mass density $\mu = 0.2\ mg/cm^2$.
The AT cut has a disc thickness of $0.165mm$ and a cutting angle $\Theta = 35.25°$. The resulting frequencies are 9.9706 MHz (1258th mode) for the fundamental mode and 30.188 MHz (3649th mode) for the third harmonic overtone mode.
The SC cut has a disc thickness of $0.180mm$ and cutting angles $\Phi = 21.9°$, $\Theta = 33.9°$. The resulting frequencies are 9.9195 MHz for the fundamental mode and 30.045 MHz for the third harmonic overtone mode.
The calculated frequency perturbations for cantilever bending are represented in Figures 9-12.

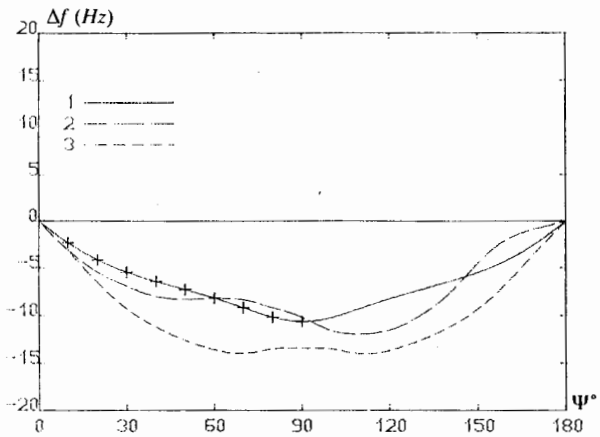


Fig. 9 : Δf vs Ψ for AT cut, fundamental mode
1 : present, 2 : experimental [10], 3 : analytical [11]

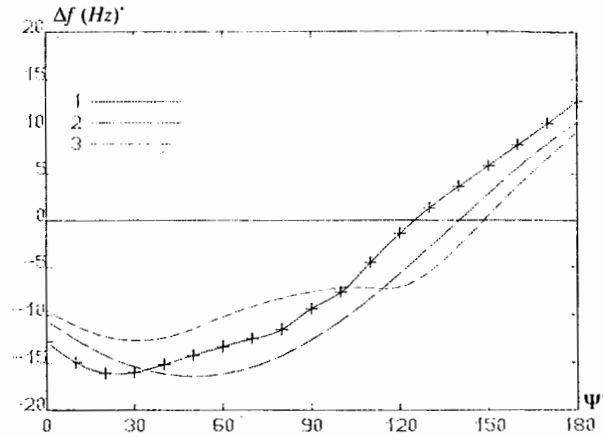


Fig. 10 : Δf vs Ψ for SC cut, fundamental mode
1 : present, 2 : experimental [10], 3 : analytical [11]

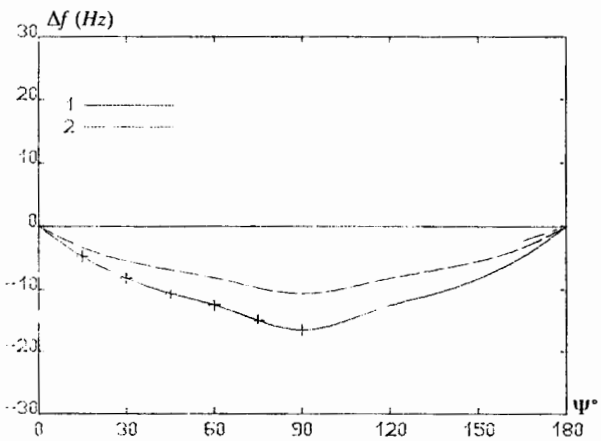


Fig. 11 : Δf vs Ψ for AT cut, third harmonic overtone
1 : overtone mode, 2 : fundamental mode

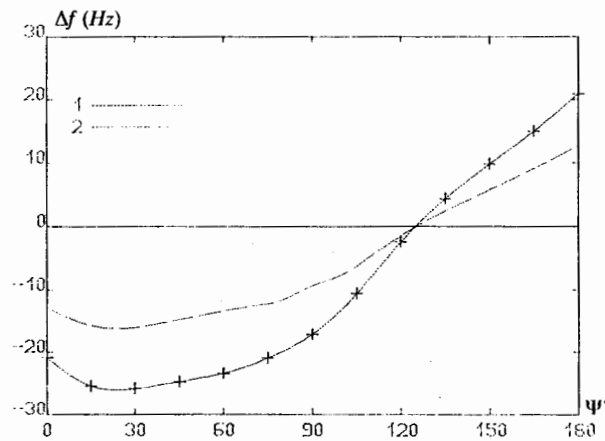


Fig. 12 : Δf vs Ψ for SC cut, third harmonic overtone
1 : overtone mode, 2 : fundamental mode

Shear-deformations for the SC cut overtone mode are represented in Figure 13. The electroded area is located in the upper-right part of the figure, and the unelectroded area in the lower-left part. We can clearly recognize the trapping effect, i.e. the fading away of the oscillations outside the electroded area. The three-axial principal tensions, characteristic of a doubly rotated cut, located in the central plane of the crystal are also represented. The AT cut has two-axial principal tensions.

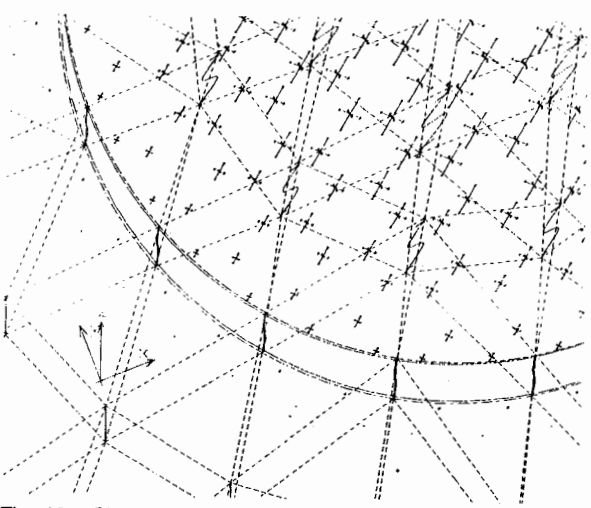


Fig. 13 : Shear-deformations and principal tensions, SC cut third harmonic overtone mode.

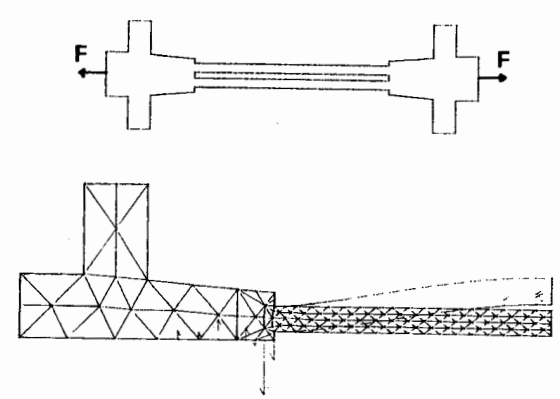


Fig. 14 : Above: double-ended tuning fork. Below: finite element mesh for a quarter of the double-ended tuning fork.

A totally different type of resonator we analysed with our method is the double-ended tuning fork force sensor, represented

in Figure 14. The static tensile forces F (Fig.14 Above) generate additional restoring forces, causing a significant increase in frequency. In Figure 14 Below is represented the finite element mesh of a quarter of the double tuning fork, with resulting nodal reaction forces on the axis of symmetry.
The arms of the $X + 1.5°$ cut double tuning fork have a length of 5.20 mm, a width of 0.245 mm and a thickness of 0.175 mm. The unperturbed frequency is 46.861 kHz and the calculated force sensitivity is 468.1 Hz/N. The experimental value [12] is 486 Hz/N. The order of interpolation in the z-direction is chosen linear (the number of degrees of freedom is 2736). About 90% of the frequency perturbation is due to initial stress.

Conclusion

The main purpose of the present study was to develop a general method to determine the first-order resonance frequency perturbation of vibrating structures, which is caused by static forces. The finite element method, with the tangent-stiffness concept, is used to analyse the frequency-force behaviour of different quartz crystal resonators. The generality of the method makes it possible to analyse the effect of arbitrarily directed perturbation forces acting on quartz crystal resonators of any geometrical shape and for any mode of vibration. We think that, with the contributions of the change in the resonator geometry, the change of the elastic constants, the nonlinearity in the strain-displacement relationship, and the initial stress, we have taken into consideration all nonlinear effects of importance.
A disadvantage of the method is the enormous amount of computation. On a VAX 8530 computer, the run time is about eleven hours to calculate, for the thickness-shear overtone mode with 8442 degrees of freedom (solution of the eigenproblem & linear static analysis & calculation of the perturbed stiffness and mass matrix), just one point in the diagrams of Figures 7, 11 or 12. For the fundamental mode the run time is about four hours.
The validity of the model is proved by the good accord of our results with those of other authors.

References

[1] O.C. Zienkiewicz, "Methode der finiten Elemente", Hanser, München Wien, 1984.
[2] H.R. Schwarz, "Methode der finiten Elemente", Teubner, Stuttgart, 1984.
[3] K.J. Bathe, "Finite element procedures in engineering analysis", Prentice -Hall, New Jersey, 1982.
[4] H. Baumgärtel, "Endlich-dimensionale analytische Störungstheorie", Akademie-Verlag, Berlin-Ost, 1972.
[5] S.G. Lekhnitskii, "Theory of elasticity of an anisotropic elastic body", Holden Day, San Francisco, 1963.
[6] P.C.Y. Lee, Y.S. Wang, X. Markenscoff, "High-frequency vibrations of crystal plates under initial stresses", J. Acoust. Soc. Amer., vol 57, pp. 95-105, 1975.
[7] J.M. Ratajski, "Force-frequency coefficient of singly rotated vibrating quartz crystals", IBM J. Res. Develop., vol. 12, pp. 92-99, 1968.
[8] P.C.Y. Lee, M.S.H. Tang, "Thickness vibrations of doubly rotated crystal plates under initial deformations", IEEE Trans. Ultrason., Ferroel., Freq. Control, vol 6, pp. 659-666, Nov. 1987.
[9] A. Ballato, E.P. EerNisse, T. Lukaszek, "The force-frequency effect in doubly rotated quartz resonators", Proc. 31th Ann. Freq. Control Symp., pp. 8-16, 1977.
[10] E.D. Fletcher, A.J. Douglas, "A comparison of the effects of bending moments on the vibrations of AT and SC (or TTC) cuts of quartz", Proc. 33th Ann. Freq. Control Symp., pp. 346-350, 1979.
[11] P.C.Y. Lee, C.S. Lam, "Effect of transverse force on the thickness-shear resonance frequencies in rectangular, doubly rotated crystal plates", Proc. 36th Ann. Freq. Control Symp., pp. 29-36, 1982.
[12] K. Trümpy, "Calcul de perturbation avec elements finis, appliqe aux resonateurs a quartz", Actes 1er Forum Europeen Temps-Freq., pp. 117-124, 1987.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-01 Number 3

DETERMINATION OF THE COMPLETE TENSOR OF NONLINEAR PIEZOELECTRICITY OF QUARTZ USING THE RESONATOR METHOD

Carl K. Hruska
Piezoelectricity Research Laboratory
Department of Mathematics, York University
4700 Keele Street, Downsview, Ontario M3J 1P3, Canada

ABSTRACT

The complete tensor of nonlinear piezoelectricity of quartz has been calculated - for the first time - exclusively using the data obtained by the resonator method. Its eight independent components were obtained by the least-squares fit to one hundred and eighty-four observations of the dc field-induced change in the resonance frequency of the extentional mode of rods and thickness modes of plates.

Simultaneously an attempt has been made to calculate the eight independent components of the electrostrictive tensor and the only independent component of the third-order permittivity tensor of quartz. Only seven values have been obtained here; three of them are linear combinations of these constants.

A comparison was made with similar tensors calculated from the transit-time experiment (Kittinger, Tichy and Friedel, J.Appl.Phys., vol.60, 1986, pp. 1465-71). The values of the third-order piezoelectric constants obtained by the two methods are in excellent agreement as the differences between the corresponding constants are comparable with their standard errors. The comparisons available for the electrostrictive constants are considerably less encouraging. No comparison for the third-order permittivity (disregarded in the transit-time experiment) is made.

The work is a classical example of an independent verification process which may also be desirable in other cases where material constants of crystals are calculated.

I. INTRODUCTION

Interactions between the dc electric field and quartz provide valuable information about its material nonlinearities. Typically, they are observed by the resonator method or by the transit-time method. The former consists in changes in the resonance frequency of quartz resonators, the latter in the variations in the transit time of acoustic pulses propagating through bulk quartz. Both changes are induced by a dc electric field acting on quartz material.

The most recent attempts by Hruska [1] and Hruska and Brendel [2] to calculate respectively the third-order piezoelectric constants and the third-order piezoelectric and electrostrictive constants from experimental data obtained by the resonator method resulted invariably in obtaining only several linear combinations of these constants which, in the latter case, were mixed with one another.

This paper overcomes the former difficulty; it isolates the third-order piezoelectric constants from the electrostrictive constants and presents the complete set of the fundamental third-order piezoelectric constants of quartz. Simultaneously, the calculation of the electrostrictive constants and of the third-order permittivity of quartz is also attempted, the latter appearing in this context for the first time.

Compared with our results are the values of the third-order piezoelectric constants and electrostrictive constants of quartz calculated from the transit-time experiment done by Kittinger, Tichy and Friedel [3].

The numerical values of all quantities (generated or referenced) in this paper are stated for right-hand quartz and the frame of reference according to the IEEE 1978 Standard [4].

II. THE RESONATOR METHOD - EXPERIMENT AND ITS INTERPRETATION

The determination of the nonlinear material constants of quartz in this paper is based on the observations of the linear coefficient L of the dependence of the resonance frequency f of quartz resonators on the dc electric field E applied to their body. This phenomenon has also been known as the polarizing effect or the electroelastic effect. The linear coefficient of the frequency-dc field dependence is defined as $L = (1/f).(df/dE)_{E=0}$.

The observations of the linear coefficient L used in this paper consist of fifty values obtained for the extentional mode of vibration of quartz rods and of one hundred and thirty-four values obtained for the three thickness vibration modes of quartz plates. They are taken from Hruska and Khogali [5], Kinigadner [6], Hruska, Merigoux and Kucera [7] and from Hruska and Brendel [8], and are all listed in Table I, parts a)-d).

The actual calculation of the nonlinear material constants in this paper is based on the linear relationship between the linear coefficient L of the frequency-dc field dependence and the nonlinear material constants which is of the following form

$$(1)\quad L_i = a_{i1}.e_{111} + a_{i2}.e_{113} + a_{i3}.e_{114} + a_{i4}.e_{122} + a_{i5}.e_{124} + a_{i6}.e_{134} + a_{i7}.e_{144} + a_{i8}.e_{315} + b_{i1}.l_{11} + b_{i2}.l_{12} + b_{i3}.l_{13} + b_{i4}.l_{14} + b_{i5}.l_{31} + b_{i6}.l_{33} + b_{i7}.l_{41} + b_{i8}.l_{44} + c_i.\varkappa_{111} + C_i \quad ,$$

where

(2) $e_{111}, e_{113}, e_{114}, e_{122}, e_{124}, e_{134}, e_{144}, e_{315}$

are the eight independent third-order piezoelectic constants, and

(3) $l_{11}, l_{12}, l_{13}, l_{14}, l_{31}, l_{33}, l_{41}, l_{44}$

are the eight independent electrostrictive constants, and

(4) $\varkappa_{111}$

is the (only) independent third-order permittivity constant.

The electrostrictive constants (3) are the components of the total electrostrictive tensor consisting of the material electrostriction and of the Maxwell vacuum stress tensor as described by Nelson [9].

All seventeen material constants (2)-(4) are the fundamental material constants of quartz related to its basic reference frame [4]. They are defined for zero strain and zero electric field in the crystal material.

In general the quantities a_{ij}, b_{ik}, c_i and C_i in Eq.(1) are functions of the second-order (linear) material constants of quartz. The index i at these quantities as well as at L_i further signifies dependence on the crystallographic orientation of the resonators and on the mode of vibration.

For the extentional mode of rods the coefficients a_{ij} and the absolute term C_i are the same as those in Eq.(8) of [1], while all coefficients b_{ik} and c_i are identically zero.

For the three thickness modes of plates the definitions of a_{ij}, b_{ik}, c_i and C_i agree with those in Eq.(12) of [2] with the following two qualifications:

(a) The work in [2] is formulated for the electroelastic constants which differ from the third-order piezoelectric constants used in this paper by their sign. As a result, the coeficients a_{ij} here and in [2] also have an opposite sign.

(b) The definition of C_i in [2] and here differ by the term $c_i.\varkappa_{111}$. This term is included in C_i in [2] but removed from it and stated separately in Eq.(1) of this paper.

The absolute term C_i in this paper as well as in [2] represents the sum of all contributions to L_i regarded as known and readily calculable. In [2] the third-order permittivity $\varkappa_{111}$ was viewed as a known constant determined by Besson and Gagnepain [10]. For this reason the term $c_i.\varkappa_{111}$ was part of C_i there. However, on closer examination, it appeared that its value was obtained for the conditions of constant (zero) stress rather than that of constant strain required here; the term $c_i.\varkappa_{111}$ has thus been removed from C_i defined in [2] and placed along the terms containing other unknown nonlinear material constants of quartz.

The general definition of the term $c_i.\varkappa_{111}$ is given by Eq.(8) in [2]. The coefficient c_i is obtained from there if the third-order permittivity tensor of quartz (with only one independent component $\varkappa_{111}$) is substituted into this equation and the third-order permittivity constant $\varkappa_{111}$ subsequently factored out.

As a whole, Eq.(1) leans heavily on the earlier results by Kittinger and Tichy [11], Tiersten and Ballato [12], Hruska and Brendel [2], Brendel [13], and Baumhauer and Tiersten [14].

III. CALCULATION OF THE NONLINEAR MATERIAL CONSTANTS AND RESULTS OBTAINED

Aiming at the determination of the nonlinear constants (2)-(4) Eq.(1) was applied to all one hundred and eighty-four observations of the linear coefficient L_i listed in Table I, parts a)-d). Due to random errors in the experimental values of L_i, this led to an overdetermined linear system

$$\begin{aligned}(5)\quad L_i - C_i = {} & a_{i1}.e_{111} + a_{i2}.e_{113} + a_{i3}.e_{114} + a_{i4}.e_{122} + a_{i5}.e_{124} + a_{i6}.e_{134} + a_{i7}.e_{144} + a_{i8}.e_{315} \\ & + b_{i1}.l_{11} + b_{i2}.l_{12} + b_{i3}.l_{13} + b_{i4}.l_{14} + b_{i5}.l_{31} + b_{i6}.l_{33} + b_{i7}.l_{41} + b_{i8}.l_{44} \\ & + c_i.\varkappa_{111} ,\end{aligned}$$

where $i = 1,2,\ldots,184$. The seventeen material constants (2)-(4) were sought by the least-squares fit to this system. Prior to it a brief analysis of the random errors of its left hand sides and of its matrix was necessary.

Given its definition, the absolute term C_i is not considered to be associated with random errors. (On the other hand, C_i may be a source of systematic errors due to the errors in the earlier determined values of the material constants of quartz which participate in its definition and which are now taken for granted. Unfortunately, this is a problem we are so far unable to deal with.) As a result, the experimental errors in the observed coefficients L_i are regarded to be the only source of random errors of the left hand sides of system (5).

Our experimental data in Table I contain forty-five subsets, each consisting of two or more values of L_i observed for resonators of equal or nearly equal orientation. Earlier estimates, sufficiently accurate for the purpose, show that within these subsets, there should be little or no variation in the theoretical value of L_i. At the same time, however, the experimental values exhibit pronounced differences.

Treating each subset as a random sample the sample standard deviation of L_i was calculated for each of the forty-five subsets. Their mean value was found to be 0.26×10^{-12} m/V. This value is typically ten times larger than the standard deviations (of the mean) of the individual values of L_i obtained from the spread of their readings. As a result the variations in L_i among nominally identical samples had to be viewed as the principal source of uncertainty in L_i and the basis for the calculation of its experimental error.

We were unable to produce this type of experimental error for about one third of the observations of L_i in system (5) as some of our resonators were specimens of orientations dissimilar from any other resonator

used for this project. For this reason we were forced to execute the least-squares procedure using the equal weighting of the left hand sides of system (5).

The matrix of linear system (5) consists of three concatenated matrices, a_{ij}, (j = 1,2,...,8), b_{ik}, (k = 1,2,3,4,...,8) and c_i. Two of its columns, b_{i2} and b_{i8}, were found to be linearly dependent on the remaining columns of the system matrix. As a result the least-squares fit did not yield the values of all seventeen sought constants but rather only of fifteen of them, three of which are linear combinations of these constants rather than their pure values.

Our results are listed together with their standard errors in the second column of Table II. The quality of the least-squares fit attained can be seen from the obtained statistical indicators [15]: r-square = 0.9964 and the analysis of variance F test value = 3,131. It is also illustrated by comparison of the experimental and predicted values of the linear coefficient L shown in Table I.

The above calculation also required the value of the density, the linear material constants and of the third-order elastic constants of quartz. They have been taken, respectively, from Sosman ([16], 2.649 x 10^3 $kg.m^{-3}$ at 20 °C), Bechmann [17] and Thurston, McSkimmin and Andreatch [18].

The experimental values L_i and the above mentioned material constants used in the calculation are a mixture of quantities referred to 20 or 25 °C or just room temperature. Consequently, the values of the nonlinear material constants calculated here are understood as valid for room temperature. Considering the existing accuracy (standard errors in Table I) of the determined material constants and their estimated temperature dependence [19] the temperature inconsistency of several degrees C is of no practical consequence. Similarly disregarded is the nonuniform thermodynamic character of the used quantities ranging from adiabatic to isothermic.

It is of considerable interest to compare the results of this work with those obtained by the transit-time metod. The latter are based on a set of twenty-four dc field-induced variations in the transit time of pulses propagating through bulk quartz [3].

The results of the two methods are easily comparable. However, using the results from [3], one is not able to state the standard errors for the two linear combinations of the electrostrictive constants obtained in this paper. Also, the iterative least-squares procedure in [3] may have inappropriately reduced the size of the standard errors in the determined material constants. For these reasons the analysis of the transit-time experiment was redone starting from the raw data and aiming at the determination of all the standard errors needed in Table II. The results of this exercise are in the last column of Table II. As the third-order permittivity $\varkappa_{111}$ is disregarded in [3], no comparable transit-time method equivalent is given in the last row of Table II.

To ensure uniformity of the data reduction procedure the recalculation was done in the same manner as all other calculations in this work. This, apart from possible numerical causes, accounts for the differences between the recalculated results and the values which would have been in their place had a direct substitution from Table IV of [3] been made instead.

IV. CONCLUSION

The complete set of eight fundamental third-order piezoelectric constants of quartz has been determined using the resonator method. The principal difficulty which has been the trademark of the resonator method - its apparent inability to produce the values of isolated third-order piezoelectric constants instead of their combinations - is thus overcome.

The typical standard errors in these constants are several percent. Their magnitude is consistent with the experimental error in the observations of the linear coefficient of the frequency-dc field dependence from which they were calculated.

An excellent agreement has been reached with the results of the transit-time experiment. The difference between the corresponding third-order piezoelectric constants averages less than is the average size of their standard errors.

The remaining seven material quantities calculated simultaneously are four pure electrostrictive constants, two combinations of the same, and one electrostrictive constant combined with the third-order permittivity of quartz. Five of these quantities have relative standard errors smaller than 30%. The remaining two values have excessively large errors rendering them totally unreliable.

Compared to the third-order piezoelectric constants the relative standard errors in the electrostrictive constants are substantially larger. This is consistent with the facts that the electrostrictive constants are a smaller contributor to the linear coefficient of the dc field-frequency dependence - typically by one to two orders of magnitude - than are the third-order piezoelectric constants.

Being obtained from two independent sources we believe that the eight third-order piezoelectric constants of quartz can be regarded now as fully verified. This is true providing that the existing values of the third-order elastic constants [18] are reliable; it is probably independent of the fact that the current data reduction method used to evaluate the transit-time experiment may require some additional attention (e.g. because of its treatment of the third-order permittivity).

No agreement similar to that obtained for the third-order piezoelectric constants of quartz has been recorded for its electrostrictive constants. The results of the transit-time experiment and those presented here agree rather poorly. Any attempt made at this point to identify the definite reasons for it or to determine the order of importance from among the known possibilities would only be risky speculation.

ACKNOWLEDGEMENTS

The author is obliged to Miss P. Hruska for her assistance with the final stages of preparation of this paper. This work was supported by a grant from the Faculty of Arts, York University and from the National Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Hruska C. K., J.Appl.Phys. 66(1989), pp.1071-1074.
[2] Hruska C. K. and Brendel R., J.Appl.Phys., in press.
[3] Kittinger E., Tichy J., and Friedel W., J.Appl.Phys. 60(1986), pp.1465-1471.
[4] 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.
[5] Hruska K. and Khogali A., IEEE Transactions on Sonics and Ultrasonics, vol. SU-18, 1971, pp.171-176.
[6] Kinigadner A., Ph.D.Thesis, University of Innsbruck, Austria, 1978.
[7] Hruska C. K., Merigoux H., Kucera M., J.Can.Ceram.Soc. 57(1988), pp.53-57.
[8] Hruska C. K. and Brendel R., J.Appl.Phys. 65(1989), pp.715-717.
[9] Nelson D. F., J.Acoust.Soc.Am. 63(1978), pp. 1738-1748.
[10] Besson R. and Gagnepain J.-J., C.R.Acad.Sci., Paris B 274(1972), pp.835-838.
[11] Kittinger E. and Tichy J., J.Acoust.Soc.Am., vol.83, 1988, pp.647-651.
[12] Tiersten H. F., and Ballato A., J.Acoust.Soc.Am., vol.73, 1983, pp.2022-2033.
[13] Brendel R., J.Appl.Phys. 54(1983), pp. 5339-5346.
[14] Baumhauer J. C., Tiersten H. F., J.Acoust.Soc.Am. 54(1973), pp.1017-1034.
[15] Mendenhall W. and Sincich T., A second course in business statistics: Regression analysis, Collier Macmillan Publishers, London, 2nd edition, 1986, pp. 150-159.
[16] Sosman R. B., "The properties of silica", Chemical Catalog Company, Inc., New York, 1927.
[17] Bechmann R., Phys.Rev. 110(1958), pp. 1060-1061.
[18] Thurston R. N., McSkimmin H. J. and Andreatch P., Jr, J.Appl.Phys. 37(1966), pp. 267-275.
[19] Hruska C. K., IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control, 35(1988), pp.637-640.

TABLE I, part a). The extentional mode of rods.

sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V		sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V	
#	ψ	φ	θ	exp	pred	exp	pred	#	ψ	φ	θ	exp	pred	exp	pred
D001	180.000	00.000	-30.00	66.15	66.55	10.62	11.46	D020	180.000	00.00	-135.00	55.42	55.53	-14.54	-14.42
D002	180.000	00.000	-30.00	66.25	66.55	10.90	11.46	D021	180.000	00.00	-135.00	55.34	55.53	-14.44	-14.42
D003	180.000	00.000	-35.00	202.2	203.6	12.83	12.84	D050*	180.000	00.00	-140.00	---	---	-15.57	-14.81
D004	180.000	00.000	-36.00	67.63	68.11	13.14	13.09	D022	180.000	-40.00	00.00	62.29	62.21	-3.17	-3.18
D005	180.000	00.000	-40.00	206.3	207.1	14.02	13.95	D023	180.000	-40.00	00.00	62.15	62.21	-3.18	-3.18
D006	180.000	00.000	-40.00	68.82	69.03	14.18	13.95	D024	180.000	-60.00	00.00	58.33	58.36	-9.27	-9.48
D044*	180.000	00.000	-40.00	---	---	14.50	13.95	D025	180.000	-60.00	00.00	58.19	58.36	-9.23	-9.48
D007	180.000	00.000	-50.00	69.92	70.00	14.96	14.70	D026	270.000	49.00	-30.83	56.29	56.23	-1.57	-1.37
D008	180.000	00.000	-50.00	69.91	70.00	14.21	14.70	D027	270.000	49.00	-32.00	55.75	55.73	-0.98	-1.03
D009	180.000	00.000	-60.00	67.95	68.01	12.03	12.82	D028	270.000	49.00	-48.50	51.39	51.48	4.14	4.00
D045*	180.000	00.000	-60.00	---	---	13.45	12.82	D029	270.000	49.00	-51.50	51.25	51.25	4.98	4.80
D046*	180.000	00.000	-75.00	---	---	6.84	6.54	D030	270.000	49.00	-58.83	51.33	51.28	6.43	6.28
D047*	180.000	00.000	-90.00	---	---	-0.04	-0.15	D031	270.000	49.00	-59.67	51.32	51.32	6.22	6.40
D010	180.000	00.000	-105.00	51.22	51.27	-5.94	-5.95	D032	270.000	49.00	-60.50	51.84	51.38	6.81	6.50
D011	180.000	00.000	-105.00	51.28	51.27	-5.90	-5.95	D033	270.000	49.00	-69.50	52.25	52.37	6.55	6.67
D012	180.000	00.000	-105.00	51.20	51.27	-5.71	-5.95	D034	270.000	49.00	-69.83	52.40	52.42	6.52	6.64
D048*	180.000	00.000	-105.00	---	---	-6.29	-5.95	D035	270.000	-35.25	-30.00	66.80	67.04	11.94	11.93
D013	180.000	00.000	-115.00	51.28	51.32	-9.52	-9.40	D036	270.000	-35.25	-40.00	69.53	69.59	11.87	11.52
D014	180.000	00.000	-115.00	51.27	51.32	-9.43	-9.40	D037	270.000	-35.25	-40.00	69.50	69.59	12.56	11.52
D049*	180.000	00.000	-120.00	---	---	-11.57	-10.98	D038	270.000	-35.25	-40.00	69.63	69.59	11.21	11.52
D015	180.000	00.000	-125.00	52.80	52.86	-12.04	-12.40	D039	270.000	-35.25	-50.00	68.21	68.27	7.74	7.96
D016	180.000	00.000	-125.00	52.76	52.86	-12.33	-12.40	D040	270.000	-35.25	-50.00	68.14	68.27	8.06	7.96
D017	180.000	00.000	-125.00	52.78	52.86	-12.25	-12.40	D041	270.000	-35.25	-60.00	63.85	63.78	3.82	3.89
D018	180.000	00.000	-130.00	53.93	54.09	-13.05	-13.58	D042	270.000	-35.25	-60.00	63.80	63.78	3.34	3.89
D019	180.000	00.000	-130.00	53.93	54.09	-13.24	-13.58	D043	270.000	-35.25	-60.00	63.81	63.78	3.70	3.89

TABLE I, part b). The slow thickness shear mode (mode C) of plates.

sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V		sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V	
#	ψ	φ	θ	exp	pred	exp	pred	#	ψ	φ	θ	exp	pred	exp	pred
R007	294.020	-0.150	00.00	2782.3	2779.2	16.76	17.08	R052	285.070	-9.240	00.00	2638.6	2626.7	17.23	16.46
R008	293.980	-0.170	00.00	2782.2	2779.4	16.38	17.08	R053	285.000	-9.190	00.00	2636.9	2625.4	15.43	16.45
R009	290.476	-0.025	00.00	2826.8	2825.1	16.83	16.71	R054	285.070	-9.220	00.00	2662.8	2657.0	17.04	16.46
R010	290.405	0.047	00.00	2831.1	2827.0	17.43	16.69	R055	285.130	-9.210	00.00	2657.5	2646.9	16.37	16.47
R011	285.280	0.150	00.00	2934.0	2925.1	16.22	15.80	R056	284.860	-9.170	00.00	2660.1	2649.4	15.78	16.44
R012	285.220	0.100	00.00	2931.4	2925.6	15.93	15.80	R057	285.080	-9.170	00.00	2682.7	2737.8	16.55	16.47
R013	280.490	0.210	00.00	3035.6	3040.7	15.10	14.42	R065	273.740	-40.430	00.00	2637.6	2627.1	-0.78	-0.91
R015	275.800	0.060	00.00	3154.3	3155.3	11.54	11.68	R066	273.620	-40.420	00.00	2639.1	2626.6	-0.76	-0.88
R016	275.680	0.260	00.00	3161.8	3162.9	11.45	11.52	R068	273.640	-40.440	00.00	2639.8	2626.9	-0.65	-0.89
R017	270.200	-0.600	00.00	3221.9	3223.5	0.53	0.62	R069	273.640	-40.440	00.00	2638.4	2626.9	-0.72	-0.89
R018	270.220	-0.590	00.00	3217.4	3223.7	0.70	0.68	R095	275.560	-41.040	00.00	2629.1	2622.3	-1.01	-1.44
R020	300.072	-10.050	00.00	2814.4	2811.0	16.11	16.15	R096	274.170	-41.100	00.00	2629.8	2620.3	-0.98	-1.09
R021	299.924	-14.688	00.00	2868.3	2873.5	14.97	15.03	R097	275.560	-41.100	00.00	2628.7	2613.1	-1.01	-1.45
R023	300.647	-20.025	00.00	2978.6	2984.9	12.93	13.37	R098	276.440	-41.090	00.00	2635.5	2623.7	-1.20	-1.67
R024	300.589	-20.350	00.00	2982.0	2990.2	13.28	13.28	R099	271.780	-40.950	00.00	2695.0	2687.0	-0.43	-0.46
R025	300.631	-24.889	00.00	3080.2	3086.3	11.94	11.82	R100	271.350	-40.970	00.00	2681.8	2673.0	-0.07	-0.35
R028	300.686	-29.939	00.00	3176.5	3179.8	8.18	8.52	R101	271.920	-40.890	00.00	2688.8	2679.4	-0.14	-0.49
R029	300.479	-35.257	00.00	3168.2	3166.8	-1.84	-2.00	R102	270.580	-41.020	00.00	2676.8	2666.5	-0.13	-0.15
R030	299.556	-35.050	00.00	3165.0	3150.5	-2.09	-1.42	R120	323.530	-26.010	00.00	5112.7	5084.5	-4.00	-3.39
R031	299.861	-39.811	00.00	3095.3	3097.1	-5.43	-5.28	R121	323.360	-27.060	00.00	5084.9	5044.3	-2.57	-2.22
R032	300.000	-39.970	00.00	3091.7	3097.2	-5.48	-5.33	R122	323.500	-29.640	00.00	4975.8	4961.4	-1.95	-1.70
R046	277.520	-9.450	00.00	2774.5	2748.9	13.51	14.09	R123	323.430	-29.150	00.00	4956.3	4951.6	-2.00	-1.74
R047	277.460	-9.450	00.00	2774.4	2767.2	14.24	14.05	R124	323.370	-29.950	00.00	4941.8	4945.3	-1.87	-1.72
R048	277.510	-9.450	00.00	2781.1	2770.8	14.64	14.08	R127	321.530	-27.000	00.00	5128.5	5104.6	-3.24	-2.66
R049	277.550	-9.490	00.00	2765.6	2743.5	13.99	14.11	R130	279.120	-28.120	00.00	3798.6	3785.7	6.93	6.04
R050	277.580	-9.470	00.00	2755.5	2743.4	14.03	14.13	R133	279.060	-29.840	00.00	3850.7	3830.0	5.14	4.41
R051	277.350	-9.590	00.00	2759.9	2749.5	14.01	13.97								

TABLE I, part c). The fast thickness shear mode (mode B) of plates.

sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V		sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V	
#	ψ	φ	θ	exp	pred	exp	pred	#	ψ	φ	θ	exp	pred	exp	pred
R007	294.020	-0.150	00.00	4201.3	4190.7	-2.11	-2.16	R032	300.000	-39.970	00.00	3506.5	3518.4	3.27	2.65
R008	293.980	-0.170	00.00	4201.2	4189.6	-2.21	-2.19	R046	277.520	-9.450	00.00	3277.9	3246.0	-10.13	-11.33
R009	290.476	-0.025	00.00	4125.3	4116.4	-4.59	-4.06	R047	277.460	-9.450	00.00	3271.7	3265.0	-10.90	-11.31
R010	290.405	0.047	00.00	4130.4	4116.6	-4.08	-4.05	R048	277.510	-9.450	00.00	3280.9	3271.3	-10.97	-11.33
R011	285.280	0.150	00.00	3994.8	3978.8	-6.78	-6.58	R049	277.550	-9.490	00.00	3264.9	3240.6	-10.89	-11.35
R012	285.220	0.100	00.00	3988.0	3975.5	-6.78	-6.63	R050	277.580	-9.470	00.00	3255.1	3241.8	-11.13	-11.35
R013	280.490	0.210	00.00	3829.3	3834.8	-8.19	-8.22	R051	277.350	-9.590	00.00	3259.8	3238.3	-11.14	-11.30
R014	280.300	0.210	00.00	3827.2	3829.0	-8.48	-8.27	R052	285.070	-9.240	00.00	3443.0	3428.3	-11.19	-10.80
R015	275.800	0.060	00.00	3693.1	3692.6	-8.21	-8.33	R053	285.000	-9.190	00.00	3438.9	3424.7	-10.91	-10.80
R016	275.680	0.260	00.00	3694.7	3695.6	-8.05	-8.22	R054	285.070	-9.220	00.00	3473.1	3468.2	-10.92	-10.79
R017	270.200	-0.600	00.00	3577.5	3579.2	-0.48	-0.51	R055	285.130	-9.210	00.00	3471.4	3457.7	-10.59	-10.77
R019	300.258	-9.750	00.00	4042.9	4051.3	-4.93	-4.85	R056	284.860	-9.170	00.00	3472.6	3450.4	-11.01	-10.83
R020	300.072	-10.050	00.00	4045.6	4037.5	-5.40	-5.10	R057	285.080	-9.170	00.00	3469.3	3575.2	-11.07	-10.77
R021	299.924	-14.688	00.00	3871.6	3873.2	-7.22	-7.31	R095	275.560	-41.040	00.00	3032.7	3038.9	0.14	0.12
R023	300.647	-20.025	00.00	3705.7	3699.8	-8.80	-8.84	R120	323.530	-26.010	00.00	5183.9	5145.8	7.41	7.57
R024	300.589	-20.350	00.00	3690.5	3687.5	-8.24	-8.95	R121	323.360	-27.060	00.00	5232.6	5182.0	5.74	6.03
R025	300.631	-24.889	00.00	3533.7	3540.7	-10.15	-9.94	R122	323.500	-29.640	00.00	5340.1	5311.5	4.12	4.57
R027	300.333	-29.708	00.00	3409.5	3418.0	-9.15	-9.08	R123	323.430	-29.150	00.00	5275.0	5258.8	4.33	4.78
R028	300.686	-29.939	00.00	3410.4	3419.1	-8.56	-8.82	R124	323.370	-29.950	00.00	5325.1	5316.6	3.98	4.51
R029	300.479	-35.257	00.00	3411.1	3428.2	-0.24	0.10	R127	321.530	-27.000	00.00	5218.1	5186.1	6.88	6.82
R030	299.556	-35.050	00.00	3413.0	3408.9	-0.03	-0.57	R130	279.120	-28.120	00.00	4297.8	4293.3	-7.33	-7.09
R031	299.861	-39.811	00.00	3505.8	3511.8	2.32	2.60	R133	279.060	-29.840	00.00	4351.0	4339.0	-5.91	-5.74

TABLE I, part d). The thickness extentional mode (mode A) of plates.

sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V		sample	(xzlwt)ψ/φ/θ			f kHz		L 10^{-12} m/V	
#	ψ	φ	θ	exp	cal	exp	pred	#	ψ	φ	θ	exp	cal	exp	pred
R007	294.020	-0.150	00.00	4809.1	4809.6	-12.97	-12.21	R031	299.861	-39.811	00.00	5515.8	5532.6	-1.88	-1.71
R008	293.980	-0.170	00.00	4809.9	4810.5	-12.53	-12.17	R032	300.000	-39.970	00.00	5519.4	5530.9	-1.47	-1.70
R009	290.476	-0.025	00.00	4845.2	4846.9	-10.88	-10.17	R047	277.460	-9.450	00.00	5014.6	5014.9	-1.77	-2.05
R010	290.405	0.047	00.00	4843.6	4845.7	-10.34	-10.17	R049	277.550	-9.490	00.00	5004.0	4976.3	-1.83	-2.07
R011	285.280	0.150	00.00	4913.6	4902.2	-7.89	-7.25	R050	277.580	-9.470	00.00	4984.2	4975.5	-2.02	-2.08
R012	285.220	0.100	00.00	4911.6	4904.6	-7.25	-7.19	R051	277.350	-9.590	00.00	4991.9	4988.5	-2.26	-2.01
R013	280.490	0.210	00.00	4935.3	4946.4	-4.94	-4.78	R052	285.070	-9.240	00.00	4877.7	4871.3	-3.89	-4.35
R014	280.300	0.210	00.00	4933.2	4947.9	-4.84	-4.69	R053	285.000	-9.190	00.00	4873.0	4866.2	-3.32	-4.34
R015	275.800	0.060	00.00	4967.4	4982.6	-2.84	-2.55	R054	285.070	-9.220	00.00	4923.1	4926.7	-3.02	-4.35
R016	275.680	0.260	00.00	4969.1	4975.6	-2.83	-2.53	R055	285.130	-9.210	00.00	4916.7	4907.8	-3.41	-4.37
R020	300.072	-10.050	00.00	4956.9	4951.9	-8.70	-9.40	R056	284.860	-9.170	00.00	4919.1	4908.6	-3.72	-4.30
R021	299.924	-14.688	00.00	5069.8	5079.9	-8.38	-7.28	R057	285.080	-9.170	00.00	4916.8	5074.3	-3.07	-4.37
R023	300.647	-20.025	00.00	5189.3	5195.8	-5.62	-5.64	R120	323.530	-26.010	00.00	6361.9	6327.5	-5.17	-4.83
R024	300.589	-20.350	00.00	5192.7	5205.1	-5.58	-5.54	R121	323.360	-27.060	00.00	6431.4	6384.3	-4.98	-4.55
R025	300.631	-24.889	00.00	5295.2	5304.2	-3.73	-4.34	R122	323.500	-29.640	00.00	6546.9	6530.4	-4.00	-3.74
R027	300.333	-29.708	00.00	5391.4	5400.8	-2.76	-3.27	R123	323.430	-29.150	00.00	6470.2	6470.8	-4.28	-3.89
R028	300.686	-29.939	00.00	5386.5	5396.0	-3.34	-3.27	R124	323.370	-29.950	00.00	6532.6	6545.3	-4.53	-3.69
R029	300.479	-35.257	00.00	5459.1	5475.7	-2.88	-2.35	R127	321.530	-27.000	00.00	6574.5	6532.0	-5.67	-5.09
R030	299.556	-35.050	00.00	5462.3	5493.9	-2.18	-2.30								

TABLE I. Experimental values of the linear coefficient of the frequency-dc field dependence L. Samples marked D and R are rods vibrating in length and plates vibrating in thickness (modes C, B, and A), respectively. The experimental values for the rods are according to Hruska and Khogali [5] (at 20°C) except for those marked * which are by Klnigadner [6]; the values for the plates are according to Hruska, Merigoux and Kucera [7] and Hruska and Brendel [8] (at 25°C). The orientation of the resonators, given using the rotational symbol (XZlwt)ψ/φ/θ, and all values of L are for right-hand quartz according to IEEE 1978 Standard [4]. The experimental and calculated frequencies are added for comparison.

The experimental values of L for rods [5] are in most cases the mean values of readings obtained for the fundamental and the third overtone. However, wherever the fundamental frequency was excitable, only its value is stated. The predicted values of L, calculated using the material constants in Table II, illustrate the least-squares fit attained in the process of determination of the nonlinear material constants.
Note: At the time of writing reference [6] was not available; the data are taken over from [11] where some of the details of the experiment e.g. temperature (presumed to be room temperature) and frequency are not mentioned.

material constant	resonator method this work	transit-time method calculated from [3]
e_{111}	-2.16 ± 0.05	-2.15 ± 0.07
e_{113}	0.43 ± 0.07	0.49 ± 0.06
e_{114}	-0.16 ± 0.04	-0.22 ± 0.05
e_{122}	1.12 ± 0.03	1.13 ± 0.06
e_{124}	-0.74 ± 0.02	-0.76 ± 0.03
e_{134}	-1.65 ± 0.03	-1.68 ± 0.03
e_{144}	-0.01 ± 0.03	-0.08 ± 0.03
e_{315}	0.78 ± 0.03	0.89 ± 0.03
l_{11}	-3.1 ± 0.9	-4.3 ± 1.0
l_{13} + 2.000 l_{44}	-9.1 ± 2.6	9.7 ± 4.2
l_{14}	-2.3 ± 0.6	-2.5 ± 1.4
l_{31} + 2.000 l_{44}	-11.2 ± 3.0	2.7 ± 3.2
l_{33}	-7.7 ± 6.9	-5.4 ± 5.2
l_{41}	-4.7 ± 0.7	-2.5 ± 2.1
κ_{111} - 2.045 .10^{-23} l_{12}	(-2.4 ± 2.5).10^{-21}	n/a

TABLE II. Calculated values of the fundamental third-order piezoelectric constants e_{ijk}, electrostrictive constants l_{ij} and of the third-order permittivity κ_{111} obtained by the resonator method. Some constants appear in combinations that could not be resolved. The first eight values are in N/(V.m), the next six are dimensionless, the last value and the numerical coefficient at l_{12} are in F/V. The errors are standard errors. Wherever possible, our results are compared with the same constants calculated from the transit-time experiment data published by Kittinger, Tichy and Friedel [3]. Given for room temperature, right-hand quartz and the frame of reference according the IEEE Standard of 1978 [4].

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-01 Number 4

ZT-CUT HF RESONATOR

J. Hermann

Centre Suisse d'Electronique et de Microtechnique
(CSEM)

Maladière 71, CH 2007 Neuchâtel, Switzerland

Abstract

The ZT-cut is useful for rectangular quartz plates vibrating by width-extension, as well as for composite structures. Such resonators are characterized by a cubic frequency-temperature curve, with a low third-order coefficient. Their practical upper frequency limit is about 4 MHz.

This paper addresses the generalization of the ZT-cut, as well as the extension of the resonator frequency limit toward the HF range. It shows that a new X-rotated cut presents some advantages over the original Y-rotated ZT-cut, particularly a smaller sensitivity to slight misorientations of the rotation axis. Higher frequencies can be reached by using a new composite resonator, realized in batch-processing by chemical etching. Experimental data relative to 8 MHz resonators are given, as well as preliminary results for 13 MHz crystals.

Introduction

In 1979, a new quartz crystal cut for contour mode resonators was presented [1]. This orientation, called the ZT-cut, is suitable for width-extensional rectangular resonators. These resonators are characterized by cubic frequency-temperature curves, with low third-order coefficients.

The usefulness of the ZT-cut has been extended to a more complex resonator, comprising two resonant plates connected by a resonant arm [2]. This structure, fabricated in batch-processing by chemical etching, includes a clamping area where both movement and stresses are evanescent. The temperature behavior remains the same as that of the rectangular resonator, but the Q factor and the aging properties are markedly better. A suitable frequency for the miniature rectangular resonator is 1 MHz, whereas the preferred frequency of the batch-processed crystal lies between 2 and 4 MHZ.

The work presented here has been directed toward two goals :

- The investigation and design of double-rotated quartz crystal cuts satisfying the same conditions as the original ZT-cut.
- The extension of the upper frequency limit of ZT-cut resonators into the HF band, which implied the development of new resonating structures.

ZT-cut Resonators

A rectangular ZT-cut resonator is shown in Figure 1.

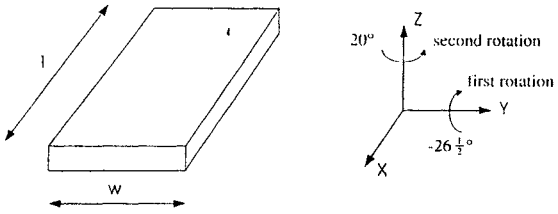


Fig. 1 ZT-cut Basic Cell

It is characterized by the following properties :

- The mode of vibration is a pure width-extension, the length-extensional and shearing movements being negligible. This property comes from the fact that two of the planar elastic constants are zero, namely : $\gamma_{12} = 0 = \gamma_{26}$.
- The first-order frequency-temperature coefficient is nearly zero, and is practically independent of the resonator width-to-length ratio because the two elastic constants are zero.
- The second-order frequency-temperature coefficient can be annulled by a suitable choice of the width-to-length ratio.

The behavior of the first three frequency-temperature coefficients as a function of the dimensional ratio is shown in Figure 2, where the vertical scale corresponds to the following units :

- first-order temperature coefficient α : $10^{-6}/K$
- second-order temperature coefficient β : $10^{-9}/K^2$
- third-order temperature coefficient γ : $10^{-12}/K^3$

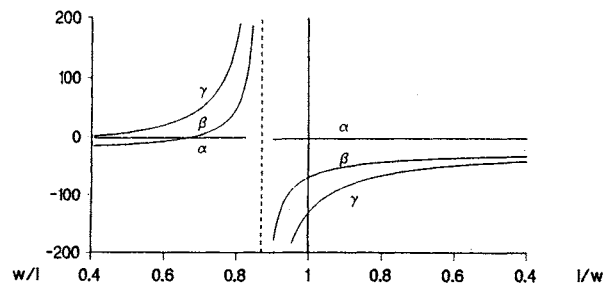


Fig. 2 ZT-cut : Frequency-temperature coefficients versus dimensional ratio

The coupling effect observed for the higher temperature coefficients is dependent upon the thermal derivative of γ_{12}, which is non-zero.

To obtain a ZT-cut resonator which can be clamped without affecting its properties, a composite resonator had to be devised [2]. It is made up of two resonant plates connected by a resonant arm. This structure, shown in Figure 3, includes a clamping area where both movement and stresses are evanescent.

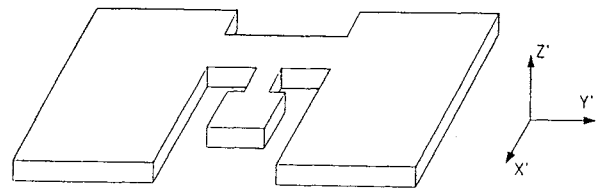


Fig. 3 ZT-cut Composite Resonator

The main properties of the 2.1 MHz composite resonators are as follows :

- Q factor : 300 to 700 $\cdot 10^3$, depending upon the relative electrode area
- Motional capacitance : up to 1.8 fF
- Static capacitance : 2.5 pF
- Frequency-temperature curve : cubic, with a third-order coefficient of 50 to 55 $\cdot 10^{-12}/K^3$
- Aging : less than 0.5 ppm for the first year at 28° C
- Static breakage limit : >15'000 g (perpendicular to the resonator plane)

Other properties also worth mentioning are [3] :

- Low acceleration sensitivity : about $3 \cdot 10^{-10}/g$
- Small dynamic temperature coefficient : 0.1 µs/K
- Good short term stability : flicker floor $8 \cdot 10^{-12}$

The preferred frequency of the structure lies between 2 and 4 MHz. The upper limit is given by the minimum acceptable size of the mounting area, as well as by impedance level considerations.

Generalized ZT-cut

The ZT-cut is relevant to width-extensional vibrations of thin plates (contour mode). Generalizing that cut requires 3 rotation angles, denoted as follows (IEEE 1978) :

(ZX twt) Φ/Θ/Ψ

To maintain the ZT-cut properties, the following quantities must be zero:

- Two of the planar elastic constants, namely γ_{12} and γ_{26}.
- The first-order frequency-temperature coefficient α.
- The second-order frequency-temperature coefficient ß. Note that this condition is satisfied for only one width-to-length ratio.

Figures 4 and 5 show the relations between Φ and Θ, and Φ and Ψ respectively, to fulfill the first of these conditions.

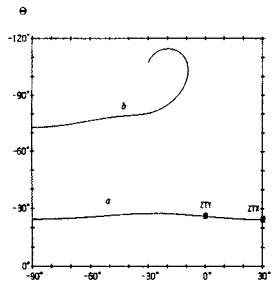


Fig. 4 Θ versus Φ for $\gamma_{12} = 0 = \gamma_{26}$

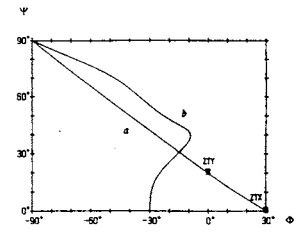


Fig. 5 Ψ versus Φ for $\gamma_{12} = 0 = \gamma_{26}$

The loci are composed of two unconnected branches *a* and *b*. Note that Θ represents the angle between the Z axis and the plate normal. As chemical etching (a prerequisite to realize complex structures) requires Θ values smaller than about 60 degrees, the *b* branch is not suitable for such cases.

Figure 6 gives the dimensional ratio w/l needed to annul the second-order temperature coefficient ß, and Figure 7 shows the first-order frequency-temperature coefficient as a function of the angle Φ.

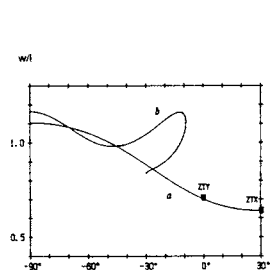


Fig. 6 Dimensional ratio versus ϕ for β = 0

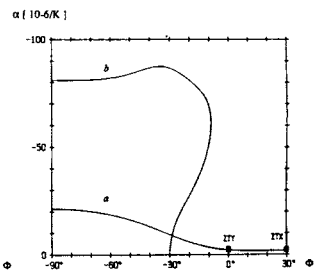


Fig. 7 Temperature coefficient α

Most parts of the locus correspond to relatively high first-order temperature coefficients, suitable for quartz thermometers. Nevertheless, a nearly zero coefficient is obtained along a portion of the *a* branch, for $0 \leq \Phi \leq 30°$. This region could be called a "generalized ZT-cut".

The fact that the first-order coefficient is not exactly zero is not a real problem : a small increase of the angle Θ is sufficient to annul α, with practically no detrimental effect on the other properties.

Two particular points on the generalized ZT locus are worth mentioning, because they correspond to single-rotated substrates. They are denoted in Figures 4 to 7 as ZTY and ZTX.

ZTY (original ZT-cut) is a Y-rotated cut with the following characteristics :

$\Phi = 0°$
$\Theta = -26.5°$
$\Psi = 20°$

ZTX is a new X-rotated cut, corresponding to :

$\Phi = 30°$
$\Theta = -24°$
$\Psi = 0°$

The new ZTX-cut [4] retains the main advantages of the original ZT-cut, namely :

- A cubic frequency-temperature curve with a low third-order coefficient (55 to $60 \cdot 10^{-12}/K^3$).
- The possibility of choosing the location of the inflection temperature by an adjustment of the dimensional ratio.
- Insensitivity of the first-order temperature coefficient to the dimensional ratio, as shown in Figure 8 (same units as in Fig. 2).

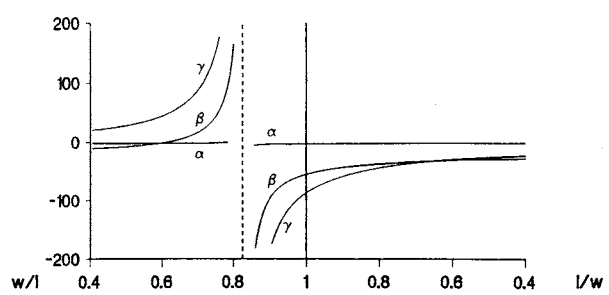


Fig. 8 ZTX-cut : Frequency-temperature coefficients versus dimensional ratio

The ZTX-cut presents some advantages over the original ZT-cut, namely :

- Due to the higher symmetry of X-rotations, the first-order temperature coefficient is much less sensitive to a misorientation of the rotation axis, or to an angular misalignment in the plane of the substrate.
- There is no in-plane rotation, so that the number of resonators on a given substrate can be increased.
- A more economical use of cultured quartz bars can be achieved.

On the other hand, the main advantage of the ZTY-cut is that the vibration mode can be excited by an electric field normal to the substrate, which gives a higher piezoelectric coupling.

The resonator shown in Figure 3 could be realized using the ZTX-cut. To obtain the same properties, slight changes of the basic resonant cell need to be made :

- Reduction of the width, to take into account the decrease of the frequency constant from 2823 kHz·mm to 2718 kHz·mm.
- Reduction of the w/l ratio, from 0.70 to 0.64.

ZT-cut HF Resonators

As there is a demand for batch-processed HF resonators with good thermal properties, the goal has been set to extend the upper frequency limit of ZT resonators. The basic resonant cell of a ZTX resonator vibrating at 8 MHz becomes quite small, precisely 0.34 mm · 0.56 mm. The structure depicted in Figure 3 is then no longer suitable :

- The overall dimensions, in particular the mounting area, would be too small for easy manipulation of the crystal.
- The reduction of the motional capacitance, compounded by the relatively small piezoelectric coupling of the ZTX-cut, would give impractically high impedance levels.

It was thus essential to find a new resonant structure which could accomodate multiple - at least ten - basic resonant cells. Among the many configurations simulated, several have been realized experimentally. The current best resonator obtained is shown in Figure 9.

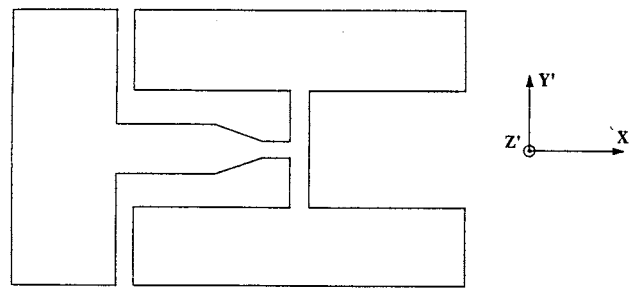


Fig. 9 8 MHz Composite Resonator

This resonator consists of the following elements :

- Two rectangular plates vibrating on the 7th length-overtone.
- A coupling arm, where the movement is evanescent, linking the resonant plates together.
- A supporting system, comprising a mounting area (shown at the left).

This structure can be enclosed in a cylindrical package, 3 mm in diameter and 8 mm in length.

The electrodes have been designed, using a method based on conformal transformations, to maximize the overall piezoelectric coupling. Due to the higher symmetry of the ZTX-cut, as compared with the ZTY-cut, only the component of the electric field along the X axis can be used for the excitation.

Numerous experiments, corroborated by computer simulations, have shown that the best results are obtained when the resonant plates vibrate with an 180 degree phase shift. The electrode configuration needed to excite this type of movement can be seen in the microphotograph of Figure 10, which shows a batch-processed crystal.

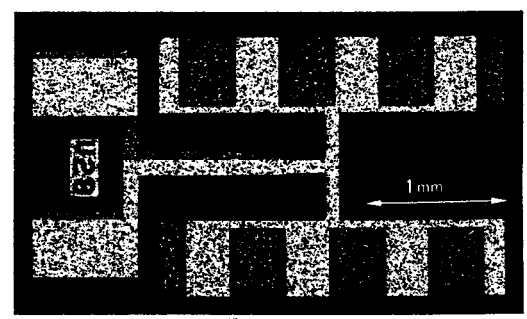


Fig. 10 Chemically-Etched 8 MHz Resonator

Experimental Results

Several slightly different versions of this resonator have been realized. Typical values for the best structures are summarized as follows :

- Frequency : 8 MHz ± 4 kHz (unadjusted)
- Q factor : 100 to $200 \cdot 10^3$
- Motional capacitance : 0.64 fF
- Static capacitance : 0.5 pF (bare resonator) + 0.2 pF (package)
- Frequency-temperature curve : cubic, with a third-order coefficient of 55 to $60 \cdot 10^{-12}/K^3$
- Aging : < 1 ppm for the first year
- Static breakage limit : >15'000 g (perpendicular to the resonator plane)

The development of this resonator has not yet been completed. Several problems need further consideration, in particular :

- The optimum thickness must be determind in order to avoid activity dips due to elastic coupling with out-of-plane modes.
- The dimensions of the coupling and suspension regions must be defined so as to minimize the energy loss due to mode conversion, which sets an upper limit on the Q factor.

Both problems are currently being examined by using a three-dimensional computer simulation program based on variational techniques.

Upper Frequency Limit

A further question is the determination of the upper frequency limit for ZT-cut resonators. This limit is set by several factors which all depend upon the minimum practical substrate thickness, usually considered to be about 80 µm :

- The influence of thickness stresses, which degrade the frequency-temperature behavior in a way which cannot be compensated for by orientation or dimensional changes.
- The deterioration of the overall geometry due to the strong anisotropy of chemical etching, characterized by the fact that all lateral faces are approximately parallel to the Z axis.
- A marked decrease of the piezoelectric coupling for too large thickness-to-width ratios.

The practical frequency limit has not yet been determined. 13 MHz resonators have been fabricated, using a somewhat different structure. The best results obtained correspond to Q factors which were much too low (10 to $20 \cdot 10^3$), but it could be verified that the frequency-temperature behavior is very similar to that of lower frequency crystals.

Future work will be concentrated on the improvement of the 8 MHz resonators, and on the quest for a structure suitable to reach higher frequencies.

Conclusions

The generalization of the ZT-cut has permitted a new X-rotated cut, characterized by a cubic frequency-temperature curve, to be found.

8 MHz resonators have been developed using that new cut. The experimental results are promising, although several problems remain to be solved.

Higher frequencies are being investigated, and first results obtained with 13 MHz crystals indicate that the cubic frequency-temperature behavior is maintained.

Acknowledgements

This work has been funded by the Swiss PTT, R+D Division in Bern. The author wishes to thank Dr.P. Kartaschoff for his unwawering support, as well as K. Hilty and F. Ruefenacht for their help.

The author is very much indebted to C.Bourgeois, who devised and ran the computer simulation programs, for many helpful discussions.

Many thanks also to F. Crevoisier for his dedication in taking care of many technological aspects of this project.

References

[1] J. Hermann and C. Bourgeois, "A New Quartz Crystal Cut for Contour Mode Resonators", Proceedings 33rd Annual Frequency Control Symposium, May-June 1979, pp. 255-262.

[2] J. Hermann, "A Novel Miniature ZT-Cut Resonator", Proceedings 39th Annual Frequency Control Symposium, May 1985, pp. 375-380.

[3] J.J. Gagnepain, "Recent Advances in Subminiature/Thin Membrane Resonators", Proceedings Ultrasonics Symposium, October 1985, pp. 291-301.

[4] US Patent Application No 07/373,470.

Applications

APP-02 Time Transfer and Navigation

Chairman: Ph. Hartl

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-02 Number 1

A review of recent advances in GPS time comparisons

B. Guinot, W. Lewandowski and C. Thomas

Bureau International des Poids et Mesures
Pavillon de Breteuil, 92312 Sèvres Cedex, France

Abstract. Using the C/A code of GPS, the inaccuracy of time comparisons, which is of the order of 20 ns at present, can be reduced to about 2 ns, even in the case of Selective Availability. This would require a number of improvements respective to the current modes of operation. These are examined. Unfortunately, the delay in the availability of precise ephemerides, of the order of a month, is a serious drawback. With the P code and advanced receivers, geodesists expect to reduce the errors of global geodesy and of the Earth rotation parameters below 10 cm. Therefore errors of time comparisons should be of under 300 ps. Such an accuracy exceeds the present needs, but it is interesting to explore the ultimate capabilities of each time comparison method and to evaluate their operational merits.

1. Introduction

After a period where only the NBS and the USNO (the acronyms are explained at the end) used the GPS time comparisons on an operational basis, a rapid development started in 1983. This is illustrated, for instance, by the number of national time services equipped with GPS time receivers which report data to the BIPM for the establishment of TAI and UTC (fig. 1). There are, of course, many other applications of GPS time comparisons in national programs.

Using the GPS time receivers available on the market and the satellites which are free of the voluntary degradation known as Selective Availability (SA), it is easy to obtain time comparisons with uncertainties of the order of 20 ns (in accuracy) over intercontinental distances. However, the resolution of the measurement of pseudo-ranges lies in the vicinity of 2 ns and it would be possible to greatly improve the accuracy by removing systematic errors at this level. Some of the actions required for this improvement will, in any case, be necessary when SA is in operation.

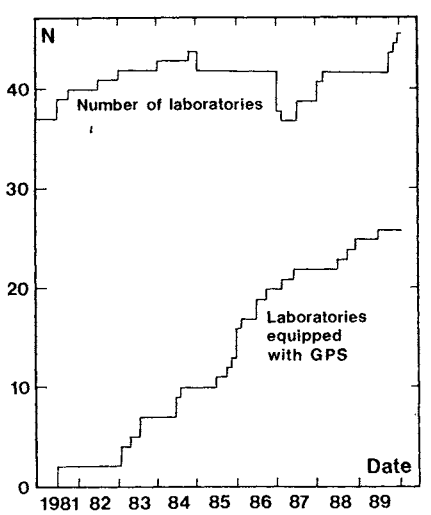


Fig. 1.- Progress of GPS time links

On the other hand, geodesists, using a new generation of receivers, expect to measure pseudo-ranges with uncertainties of 10 cm or less and to reduce ephemeris errors to less than 10 cm. These developments bring the hope that time comparisons could be achieved with uncertainties of 300 ps or less.

After showing a typical error budget for GPS time comparisons, we give an overall view of the efforts of various groups to reduce the errors. We also point out some difficulties which have been encountered.

We restrict our discussion to the GPS segment of time comparisons, i.e., we do not consider the contribution of the clocks to be compared or the time equipment in the laboratories.

2. Typical uncertainties of GPS time comparisons

The typical error budget of Table I is given for two distances, 1000 and 5000 km and for the usual tracking durations of 13 minutes, when applying the 'common view' mode (1). Two cases are considered : a single common view and a daily average of 10 common views.

This budget is established for normal operating conditions. Much larger errors may occur in case of defective receiver, lack of delay calibration, poor environment of the antenna, adoption of wrong coordinates, etc.

Table I - Typical error budget of GPS time comparisons in common view (CV), at distance d, C/A code.

Unit: 1 nanosecond

	For a single CV		For 10 CV, average over one day (1)	
d	1000 km	5000 km	1000 km	5000 km
Antenna coordinates (2)	20	20	7	7
Satellite coordinates	2	8	1	3
Ionosphere (day time, normal solar activity, elevation > 30 °)	6	15	2	5
Troposphere (elevation > 30 °)	2	2	0,7	0,7
Instrumental delay (relative)	2	2	2	2
Receiver software	2	2	2	2
Multipath propagation	5	5	2	2
Receiver noise (13 min average)	3	3	1	1
Total	22	27	8	10

(1) The noise of the clocks of the laboratories and the rise time of reference pulses bring non-negligible contributions, which are not considered here.
(2) Assuming uncertainties of the order of 3 m. In practice, errors of coordinates are often much larger and reach sometimes 30 to 40 m.

3. Improvement of GPS time comparisons

3.1. Antenna coordinates

There are two different requirements on the coordinates of antennas.
(a) Coordinates should be known accurately in a global terrestrial reference frame. The accuracy should be, for example, of order of 20 cm for 1 nanosecond accuracy in time comparisons.
(b) Satellite positions should be expressed in the same global frame as the antennas.

Concerning (a), the CCDS has recommended (2) the use of the IERS Terrestrial Reference Frame (ITRF). This frame is realized by the publication, by the IERS, of coordinates of about 100 primary sites, with uncertainties of a few centimetres (3, 4). It is usually possible to obtain antenna coordinates by relative positioning between the antenna and the IERS site with uncertainties of a few centimetres, but these measurements require the skill of geodesists.

To ensure (b), the ITRF coordinates have to be converted in the WGS 84 used by GPS by applying formulae and parameters provided by the IERS or (5). This conversion is not accurate at the level of 20 cm. But condition (b) is less stringent than (a). Consider, for example, a global shift in rotation between the satellite and receiver reference frame, which would correspond to a maximum offset of 3 meters on the ground. In the worst case the effect on time comparisons would be no more than 1 ns for stations 500 km apart. However, this type of error increases almost linearly with distance.

It has been shown that a relative positioning using only GPS time comparisons brings satisfactory results over distances of several hundreds kilometres (6). A first application of this method, with ionospheric model and broadcast ephemerides has led to homogenized coordinates in Europe, East Asia and North America (7), which are employed in the BIPM data processing. Further improvements of this technique will result from measurements of the ionospheric delays and precise ephemerides. Nevertheless accurate positioning in the ITRF is strongly advisable, in order to ensure condition (b) and as a check. It might happen, in particular, that relative positions derived from time observations be biaised by differences in the receiver softwares (see 3.5).

3.2. Satellite coordinates

Precise ephemerides are made available by the DMA ; they are received at the BIPM. Comparisons with broadcast ephemerides, for 1988, show a significant improvement in some cases (for satellites of Block I working on Rubidium clocks, during the eclipse season) (8,9). But we can expect future improvements of precise ephemerides for the needs of geodesists. A global tracking system in support to 'missions to the planet Earth' is being established (10), aiming at a fully reliable, continuously operating service. GPS satellite tracking is now one of the fundamental techniques of the IERS, together with VLBI and

satellite/Moon laser ranging. In these activities, the satellite position errors should drop below 10 cm, contributing less than 300 ps in time comparison errors for the longest terrestrial distances.

The main inconvenience of precise ephemerides provided by external bodies is the delay of availability which, at present, is of the order of one month. An alternative could be to derive the satellite positions from time observations themselves, by geometrical or dynamical methods. The geometrical approach is possible using multi-channel receivers (11). However, dynamical methods have a superiority, well known in space geodesy, in that they use an additional information through the constraint imposed by the equations of motion. The dynamical derivation of a short arc of orbit from time data has already been demonstrated (12), and is presently being studied jointly by the NIST and the University of Colorado. This approach may require additional trackings ; it also requires accurate pseudo-ranges obtained with good antenna coordinates and measured ionospheric delay.

In practice, computations with precise ephemerides require knowledge of the broadcast ephemerides used by the receiver software in order to apply differential corrections. A difficulty is the possible change of ephemeris parameters during the usual 13 minute tracking period (especially with SA) which would lead to the exchange of an enormous amount of data. A modification of receiver software is being prepared by the NIST, so that a single set of parameters be retained for the full duration of the 13-minute trackings. On the other hand, in case of current use of precise ephemerides, the BIPM will organize the collection of broadcast ephemerides for the tracking schedules (one or two volonteers per area would suffice).

Let us remark that the obtention of accurate time comparisons when the SA is implemented requires the use of corrections to the broadcast ephemerides. These corrections can be derived from precise ephemerides as said previously. But another possibility is also considered: the availability of differential corrections representing the bias introduced for SA. Such corrections could be made avalalble within a few days and would restore the quality of the broadcast ephemerides without SA.

3.3. Ionospheric refraction

Time laboratories do not have normally access to code P, however dual-frequency receivers applying correlation techniques without knowledge of code P have been developed recently. They give measurements of the ionospheric delay along the line of sight of satellites with uncertainties of about 1 ns. A first prototype was built at the BIPM by a Japanese guest visitor from the CRL (13) and operates routinely at the BIPM. After improvement at the CRL, an industrial version of this instrument has been made available commercially. An other type of 'ionospheric calibrator' is being developed at the NIST (14) and prototypes are already operating at the NIST and the OP. This new calibrator is expected to be included in the usual GPS time receivers. The improvement of time comparisons using ionospheric measurements instead of model values is evident over long distances (15). Indirectly this improvement should also be useful by increasing the accuracy of relative positioning by the method of (6). At the BIPM, measured ionospheric delays have been used since November 1989 for the link Japon - Europe. Experiments are in hand for the link NIST - OP.

3.4. Tropospheric refraction

At radio frequencies the troposphere is a non-dispersive medium, and its effect on pseudo-ranges and time comparisons cannot be estimated from two-frequency measurements as for ionospheric delay (16). Instead models are used for the estimation of the tropospheric delay. The complexity of these models and the inclusion of data from local meteorological measurements depend upon the accuracy required. It has been assumed that for the needs of GPS time transfer at the level of 1 to 2 ns and observations performed at angles of elevation above 30°, a simple global model is sufficient. However, in the practice of common view time transfer over long distances (9000 km), elevations of 20° are sometimes unavoidable. We have also observed that different types of receivers use different tropospheric models. For instance a comparison of two receivers has shown differences of 1.0 ns at 60° elevation, 1.8 ns at 30° and 3.2 ns at 20°. To achieve an accuracy of a few hundreds of picoseconds by GPS time transfer, more sophisticated troposphere models and local meteorological measurements will be necessary.

3.5. Instrumental delays, receiver software

Several experiments on the relative calibration of receiver delays have been performed by transportation of a GPS receiver, used as a transfer standard. The resolution is of the order of 1 ns for 1 - 2 days of simultaneous tracking. However, a few of the receivers have been checked. Some received a single visit and very few received two or more visits. Our experience concerning the long-term stability of receiver delays is very limited and drifts or steps of several tens of nanoseconds could occur without being noticed.

In these relative calibrations, as well as in current GPS time comparisons, the receiver software, the adopted systems of reference and constants should be identical. Unfortunately, differences have been found, as noted in 3.4 and shown in (17). An important advance would be to agree a standard data processing. It should also be possible to extract from receivers sufficient information to check the conformity of formulae and constants with the adopted standards.

But, if we really wish to reach ultimate accuracy, we shall have to organize trackings so that strictly simultaneous pseudo-ranges, in satellite time, are measured, and send these measurements to a centre where they would be processed in an homogeneous manner. The differences between receivers and receiver software are undoubtedly an obstacle to highly accurate work.

3.6. Multipath propagation

Multipath propagation arises from reflections at objects located around and under a GPS antenna. Resulting instantaneous errors can be as large as several tens of nanoseconds (18, 19). Fortunately, these errors are partially averaged over 13 minute tracks. Still, special care should be taken in the installation of a GPS antenna. An ideal installation would consist in total isolation of the antenna from its environment.

One approach of this ideal situation would be to install the antenna on the top of a high tower (20). Another option is to locate the antenna directly on the ground in an open flat field with no obstacles within a radius of several tens of meters. In practice, good locations are difficult to find. In any event GPS antennas should be located in open areas and equipped with protection planes to eliminate reflections from below.

3.7. Receiver noise

With the averaging performed in time services, receiver noise does not seem to be a serious limitation when aiming at an accuracy of 1 to 2 ns, as long as there are no systematic effects such as temperature dependence. For better performance there exist new advanced GPS receivers which have demonstrated an accuracy of 10 cm on smoothed P-code pseudo-ranges over averages of 2 minutes and better than 5 cm on averages of 30 minutes (21).

4. Improved GPS time comparisons

Table II gives a revised error budget for GPS time links, using receivers which are presently in operation, subject to the following suppositions :

- error of antenna coordinates of 10 cm (the figures for a single CV corresponds to the worst direction in both sites),
- error of satellite coordinates of 5 m (in the worst direction for a single CV),
- measured ionospheric delay with existing ionospheric calibrators,
- modelled tropospheric delay (as in Table I),
- measured relative instrumental delays by receiver transportation,
- identical, and correct, receiver softwares,
- good shielding of the antennas against multipath propagation,
- receiver noise as in Table I.

Among these suppositions, the most difficult to realize is the compatibility of receiver software. However the largest improvements arise from the adoption of accurate coordinates for the antennas and measurement of ionospheric delay : these improvements can now be easily achieved. The estimates of errors in Table II are conservative, nevertheless we observe that the contribution of GPS to the uncertainties of time comparisons at 5000 km distance, on daily averages of 10 common views, can be of the order of 2 ns, using C/A code only.

Table II - Possible error budget, with optimum operation, in common view (CV), at distance d, C/A code.

Unit: 1 nanosecond

	For a single CV		For 10 CV, average over one day	
d	1000 km	5000 km	1000 km	5000 km
Antenna coordinates (± 10 cm)	0,6	0,6	0,2	0,2
Satellite coordinates	1,0	4,0	0,3	1,3
Ionosphere (measures)	1,0	1,0	0,3	0,3
Troposphere (model, elevation > 30 °)	2,0	2,0	0,7	0,7
Instrumental delay	1,0	1,0	1,0	1,0
Receiver software	0,0	0,0	0,0	0,0
Multipath propagation	1,0	1,0	0,3	0,3
Receiver noise (13 min average, phase measurements)	3,0	3,0	1,0	1,0
Total	4,1	5,7	1,7	2,1

5. Conclusions

For time comparisons, we still use the GPS in a rather primitive manner. Even so, it provides an accuracy of about 20 ns on daily averages.

An increase of accuracy by a factor 10 could be achieved with existing receivers complemented by ionospheric calibrators. This would require an effort in operation, mostly for discrete actions such as better positioning of the antennas, measurement of instrumental and cable delays, checking of receiver software. Routine work would be increased by the use of precise ephemerides, but the BIPM is willing to undertake this part of the work. Anyway, the use of precise ephemerides could be obligatory when the Selective Availability is implemented. Even in that case, however, the accuracy of 2 ns seems possible, with worldwide coverage.

Useful steps forward would be the use of a new generation of receivers and cooperation with geodesists. An accuracy of 300 ps would then be within reach.

The main inconvenience brought by the up-grading of accuracy is a delay, of order one month, in the availability of time comparisons, required for to the preparation and dissemination of the precise ephemerides.

We are faced with a new situation in time metrology in which the potential accuracy of time comparisons exceeds the level needed to match clock performance. It has thus become necessary to evaluate the accuracy needed to carry out routine tasks and asses the cost of achieving it. We might have to optimize the network of time comparisons by a combination of techniques.

However, we must be prepared to use clocks with better accuracy and stability, so that we must explore the ultimate limits of the various methods of time comparison.

References

(1) Allan, D. W., Weiss, M. 'Accurate time and frequency transfer during common view of a GPS satellite', Proc. 34th Ann. Freq. Control Symp, 1980, pp. 334-346.
(2) Comité Consultatif pour la Définition de la Seconde, Rapport de la 11e session, 1989 (in press).
(3) Boucher, C., Altamimi, Z. 'The initial IERS Terrestrial Reference Frame', IERS Technical Note 1, 1989.
(4) IERS, Annual Report for 1988, 1989, Observatoire de Paris.
(5) CCIR Handbook on Satellite Time and Frequency Dissemination, in preparation.
(6) Guinot, B., Lewandowski, W. 'Improvement of the GPS time comparisons by simultaneous relative positioning of the receiver antennas', Bull. géodésique, 63, 1989, pp. 371-386.
(7) Lewandowski, W., Douglas, R. J., Klepczynski, W. J., Strange, W., Suter, J., Weiss, M. A.'Positioning of GPS antennas in Time-keeping laboratories of North America', Proc. 43rd Freq. Control Symp.,1989, pp. 218-224.
(8) Remondi, B. W., Hofmann-Wellenhof, B. 'GPS broadcast orbits versus precise orbits', Proc. of the Meeting of Int. Ass. of Geodesy, August 1989 (in press).
(9) Lewandowski, W., Weiss, M. A. 'The use of precise ephemerides for GPS time transfer', Proc. 21st PTTI meeting, 1989 (in press).
(10) Neilhan, R.E., Melbourne, W.G., 'GPS Global tracking systems in support of missions to the planet Earth', CSTG Bulletin, 11, 1989, pp. 129-140.
(11) Fliegel, H.F., Beard, R.L., Landis, G.P., Stebbins, F.B. 'An alternative common view method for time transfer with GPS', Proc. Second Intern. Meeting Inst. of Navigation, 25-29 Sept. 1989, Colorado Springs, USA.
(12) Capitaine, N., Gambis, D. 'A possibility of clock synchronization between two stations by the way of an artificial satellite', Ann. Géophys., 37, fasc. 1, 1981, pp. 139-142.
(13) Imae, M., Miranian, M., Lewandowski, W. and Thomas, C. 'A dual frequency codeless GPS receiver measuring ionospheric effects and its application to time comparison between Europe and USA', Proc. 3rd European Forum on Time and Frequency, 1989, pp. 89-93.
(14) Davis, D., Weiss, M.A., Vidmar, N. 'A codeless ionospheric calibrator for time transfer applications', Proc. Second Intern. Meeting Inst. of Navigation, 25-29 Sept. 1989, Colorado Springs, USA.
(15) Imae, M., Miki, C., Thomas, C. 'Improvement of time comparisons results by using GPS dual frequency codeless receivers measuring ionospheric delay', presented 21st PTTI meeting, Redondo Beach, USA, 28-30 November 1989.
(16) Dodson, A.H., Hill, J.C. 'The effects of atmospheric refraction on GPS measurements', Proc. Seminar on the GPS, 12-14 April 1989, Nottingham, U.K.
(17) Kirchner D., Bessler, H., Fassl, S. 'Experience with two collocated C/A-code GPS receivers of different type', Proc. 3rd European Time and Freq. Forum, 1989, pp. 94-103.
(18) Wells, D. et al. 'Guide to GPS positioning', 1986, Canadian GPS Associates, Fredericton, NB, Canada (see 9.11 and 9.12).
(19) Bishop, G.J., Klobuchar, J.A., 'Multipath effects on the determination of absolute ionospheric time delay from GPS signals', Radio-Science, 20, 1985, pp. 388-396.
(20) Nard, G., Rabian, J. Gounon, R., 'Utilisation des signaux du GPS en mode différentiel instantané pour les applications temps-fréquence de haute précision', Proc. 1st European Time and Frequency Forum, 1987, pp. 237-248.
(21) Srinivasan, J.M., Duncan, C., Purcell, G., Young, L., DiNardo, S., Hushbeck, E., Munson, T., Meehan, T., Yunck, T. 'Measurement of Aircraft Position, Velocity and Orientation Using Rogue GPS Receivers', Oct. 1988, A Preliminary Report, NASA CDP Fall Meeting, Munich, FRG.

List of Acronyms and abbreviations

BIPM	Bureau International des Poids et Mesures
CCDS	Comité Consultatif pour la Définition de la Seconde
CRL	Communications Research Laboratory
CV	Common view
DMA	Defense Mapping Agency
GPS	Global Positioning System
IERS	International Earth Rotation Service
ITRF	IERS Terrestrial Reference Frame
NIST	National Institute of Standards and Technology
NBS	National Bureau of Standards (now NIST)
OP	Observatoire de Paris
SA	Selective Availability of GPS
TAI	International Atomic Time
USNO	US Naval Observatory
UTC	Coordinated Universal Time

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-02 Number 2

GEOSTATIONARY NAVIGATION SATELLITES AS A MEANS OF TIME TRANSFER

S. Starker, H. Nau, J. Willner
German Aerospace Research Establishment (DLR), Oberpfaffenhofen,
Federal Republic of Germany

G. V. Kinal
International Maritime Satellite Organisation, London, United Kingdom

Introduction

As is generally known, GPS-NAVSTAR and GLONASS are national satellite navigation systems currently under construction and dedicated primarily for national military use. Civil applications will be allowed under several constraints and at the own risk of civil users [1].

Because of this, general civil use of these attractive systems will cause several problems in the future, and large user organizations may hesitate to rely on these systems in their routine navigation work. In this situation, national and international user groups have identified the additional requirements of civil users [2]. Diverse technical proposals were considered with respect to these requirements leading to the result that the complementary use of geosynchronous satellites for navigation and communication appeared as the most attractive solution.

Such a complementary system to GPS-NAVSTAR using geostationary satellites was outlined in a first step by DLR [3], INMARSAT [4, 5] and CNES [6]. The concept has been verified by some preliminary measurements. In the following this system shall be presented with special attention to the aspects of time transfer.

Main problems in the general use of GPS-NAVSTAR

The constraints for civil applications of the GPS-NAVSTAR pertain to the availability, integrity and the accuracy of the system.

The system availability, defined as the percentage of time, at which a minimum of 18 satellites will be operational and will meet the specified error budget, ist 98% [1]. There will be regions and times of reduced PDOP-values (≥ 6), as shown in Fig. 1 and 2.
With respect to integrity, a warning-message for each satellite to indicate satellite internal failures and notify users within 6 sec is required particularly for aeronautical users. In the absence of such an integrity warning channel, some failures detectable only by the control segment could take from 15 min to several hours to rectify [1].

Special features for denying the system to enemy forces are the 'Selected Availability' (SA) and the 'Anti-Spoofing' (A-S). Both are system capabilities, which may be exercised in unannounced test periods at any time, potentially causing degradation or loss of capability for unauthorized users. A reduction of the C/A-code accuracy is provided according to national security interests: $2\,\sigma_{hor} \approx 100$ m; $2\,\sigma_{vert} \approx 156$ m; $2\,\sigma_t \approx 90$ ns. The Selective Availability is accomplished by manipulating navigation message orbit data and satellite clock frequency. The degradation for the frequency L_1 is about 40 ns → 300 ns. By using the common-view technique, an essential part of this effect can be eliminated in time transfers.

Thus, the main problems will concern the civil navigation, while a relatively small reduction of time comparison accuracy to the order of 50 to 100 ns may be expected in the future. Nevertheless an additional undisturbed time signal, transmitted by a civil navigation complement may be of interest for a precise and safe time transfer.

The concept of a complementary geosynchronous navigation satellite

An improvement of integrity requires the improvement of the system monitoring and the acceleration of the evaluation and distribution of monitoring results. If this distribution is performed at the navigation frequency L_1 in a GPS-like format, the user receiver-costs can be minimized. And if this signal can be synchronized to the GPS-time an additional undisturbed navigation signal is available, which may be useful to improve the system coverage, the PDOP-values and the time transfer accuracy.

Considerations like these led to the concept of an additional civil complement to GPS-NAVSTAR, as visualized in Fig. 3: The GPS-Space-Segment is increased by one or two geosynchroneous navigation satellites, which emit GPScompatible signals and which are controlled by a civil ground segment. Because communication satellites may be used for this purpose and because their communication channels will have some importance for the complementary system, these kind of satellite may be denoted shortly as NAVCOM satellites. The ground control segment consists of several monitor-stations at well known positions, a data processing center and a feeder station, which provides the NAVCOM with a GPS-like navigation signal, carrying the necessary integrity data to the users.

If some of the monitor-stations possess a transmitting link to NAVCOM, several advantages can be reached: The monitoring results can be transferred via satellite without delay to the Data Processing Center; the orbit determination of NAVCOM can be performed by two-way ranging measurements and a two-way time transfer can be established between these monitor stations and the feeder station with its master time scale.

As soon as a synchronization of the satellite signal with the ground master time scale is demanded, a two-way radio link has to be established between NAVCOM and feeder station. In this case it will be necessary to produce a replica of the GPS time-scale at the feeder station as well as at the monitorstations by smoothing the received time information of many GPS satellites. At this point the question arises, if this replica of GPS-time should be used directly for synchronizing the civil complement and its navigation signal, or if another independent and more precise time scale should be used for this purpose.

Improved possibilities of time transfer with Geostationary Navigation Satellites

The method, to receive GPS time signals and to generate a smoothed time scale as an approximation to the GPS system-time was applied in the so-called GPS disciplined clocks. With such a clock first experiments have been performed, which are described later. The advantage of this method is that each monitoring station and each feeder station can generate its own time scale using GPS-satellites (or GLONASS-satellites) and no additional network synchronization is necessary. The precision of these GPS-dependent time scales will be limited as well as that of the civil navigation signal. These limitations may be of low importance for civil navigation applications, but for users interested in highly accurate time and frequency signals this solution would imply that the time signals from the NAVCOM-satellite would have only few advantages compared with a direct reception of GPS-signals: the continuous availability, the smoothing of SA-effects and the possibility of an improved S/N ratio and the suppression of multipath effects by using fixed directional antennas.

An essential further improvement of time transfer can be reached only, if an independent high precision clock would be used as a master time scale at the feeder station, and when the difference between this time scale and the replica of GPs-time would be transmitted together with the monitoring data. The same procedure could be used to distribute the deviations of GLONASS-time or UTC from this primary time scale. By this way an essential improvement to the existing GPS-capability in time transfer could be reached by NAVCOM satellites.

For the comparison of the master-time with UTC a two-way link to the primary standards of time-institutions would be desirable. This could be realized with a minimal expense, if some monitor stations would be placed at time institutes.

This short outline may be sufficient to show that a civil navigation complement to GPS implies the possibility for a worthwhile improvement of continental time distribution and that therefore the time community should be involved in the preparation of these systems.

First measurements with an INMARSAT-DLR Testbed

INMARSAT examined the possibilities for fitting its future generations of communication satellites also for navigation. Their geostationary 'OVERLAY' system concept to complement GPS and GLONASS is shown in Fig. 4. The navigation signal is generated in a coastal earth station and transmitted at C-band to the communication satellite.

In the navigation payload the signal is received and retransmitted in C-band to the earth station for control purpose and may be used for range measurement. It is also transponsed in L-band and any user may receive it as an additional to GPS or GLONASS.
INMARSAT started in spring last year by transmitting a GPS like signal from the Goonhilly earth station to the MARECS-B satellite. The signal is generated by a GPS-simulator. A rubidium Clock is disciplined by a GPS-time-receiver to GPS time. Only one data frame is transmitted.

This signal has been received since this time at DLR Oberpfaffenhofen with a modified GPS receiver. The first oscillator frequency had to be changed by an amount of 33,42 MHz according to the different carrier frequencies of GPS and MARECS-B, and an antenna of high gain had to be used (Fig. 5).
After regeneration of the time mark (1 pps) in the receiver, time difference measurements were performed between the receiver signal and the time mark (1 pps) from DLR cesium atomic clock running on UT. The time difference between this clock and GPS time is measured by a GPS time-receiver. Fig. 6 shows the result of one week measurements on different satellites dayly. The evaluation gives the possibility to determine the slope and the time difference between this atomic clock and GPS-time for a certain time. Fig. 7 shows the time differences measured over 24 hours, which describe the motion of the satellite. The residual measurement noise is shown on Figure 8. The rms value is 5 ns in a time interval of about 15 minutes.

The time difference measured depend mainly on 3 influences: · the time deviation between the participating clocks · the satellite position · the technical- and propagation delays. All three influences vary with time. They should be separated and determined as accurately as possible during the next experimental steps.

Conclusion and Outlook

The advantages of a geostationary overlay to GPS, and in a similar way to GLONASS, appear so convincing that it will be only a question of time, until additional geostationary navigation satellites will exist. INMARSAT, as the most appropriate international organization for operating such an overlay, has formulated agreements with a number of other organizations regarding cooperation in the development of a global civil satellite navigation system. These organizations include the European Space Agency, national space research groups in several European states, as for instance DLR in FRG, CNES in France and MORSVIAZSPUTNIK in USSR. In the specifications for the INMARSAT-3 generation of satellites, which will be available 1994 onwards, an

on-board navigation package is envisaged as an option. A decision on this option will be necessary by the end of 1990.

ESA's NAVSAT concept, proposed some years ago as a neutral international civil alternative to GPS, gradually evolved to include both inclined orbit and geostationary satellite elements, and studies have been undertaken to modify the NAVSAT system concept to make the signal format compatible with GPS, and to permit integrity warning services covering GPS as well as GLONASS.

Regarding additional activities of user organizations it looks very probable that a pilot project of this kind will come to existence in Europe in the near future. The authors believe that the european community for time and frequency should participate in these activities and use the resulting facilities for an improved time and frequency transfer in Europe.

In a further step the possible advantages should be investigated for the case that geosynchroneous navigation satellites could be used simultaneously as favored platforms for two-way time transfer transponders.

References

[1] Technical Characteristics of the NAVSTAR-GPS. September 89.

[2] Recommendations of the Special Commitee on Future Air Navigation Systems (FANS).

[3] Starker, S.
A GPS-NAVSTAR-Supplemlent as a Possibility for Autonomous Civil Satellite Navigation.
35th Congress of the International Astronautical Federation, Oct 1984. Lausanne, IAF-84-426.

[4] Kinal, G.V.
International Satellite Radionavigation and Radiolocation: Choosing the Options.
Preliminary Proceedings of the First International Technical Meeting
ION GPS-88, U.S. Institute of Navigation, Colorado Springs, 21-23 Sept., 1988

[5] Kinal, G.V.
Combining the Use of Geostationary and Inclined Orbit Satellites for Integrated Communications and Navigation Applications.
IEE 4th International Conference on Satellite Systems for Mobile Communications and Navigation, London, 17-19 October, 1988.

[6] Eydaleine, G.
Navigation by satellite - the next step for civil aviation.
ICAO Bulletin, March 1988.

[7] Rosetti, C.
A Civil Complement to GPS and GLONASS.
The Royal Institute of Navigation, Satellite Navigation Conference, London, 17-19 October 1989.

[8] Diederich, P.
The Development of Civil Satellite Navigation in Europe.
Preliminary Proceedings of the First International Technical Meeting ION GPS-88, U.S. Institute of Navigation, Colorado Springs, 21-23 September, 1988.

[9] Kinal, G., Starker, S.
Time Synchronization/Distribution Applications of Navigation Signals Repeated by Geostationary Satellites.
Radio Technical Commission for Aeronautics, 1989 Annual Assembly Meeting, and Technical Symposium, December 4-6, 1989, Washington, D.C.

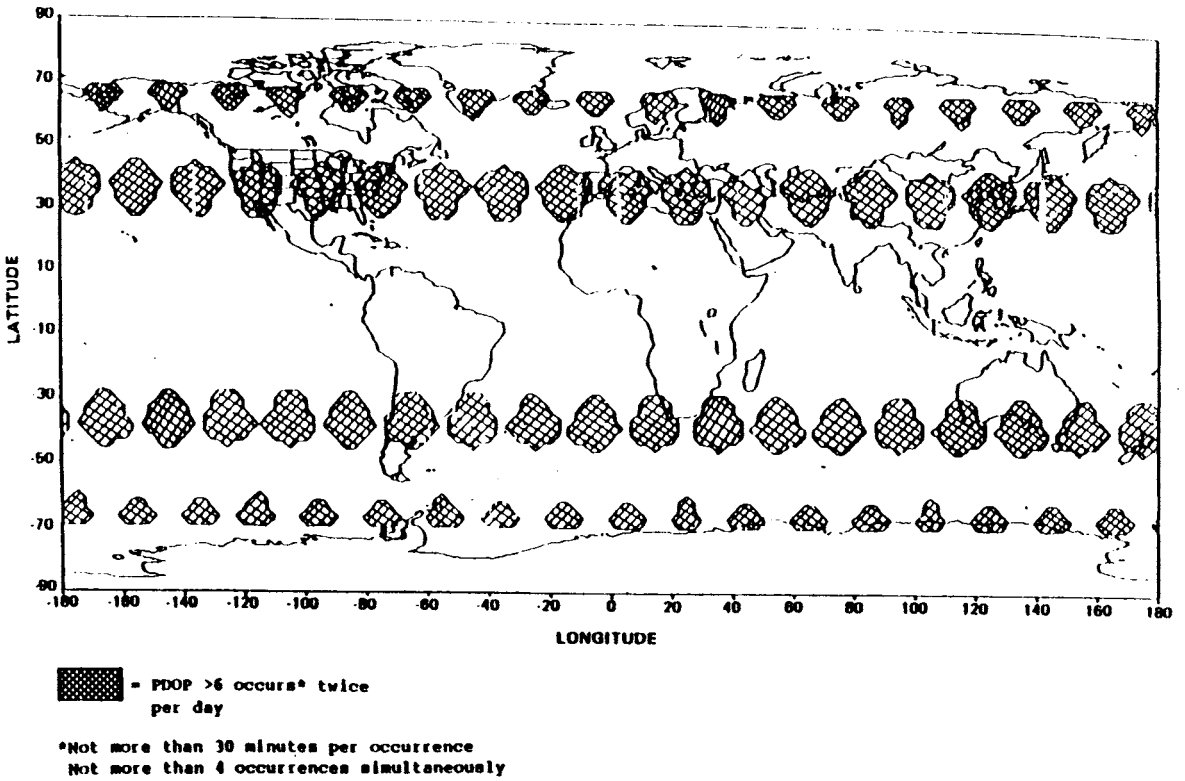


Figure 1. Daily Composite Operational Coverage (PDOP ≤ 6), 5° Elevation Mask Angle, 18 Satellites

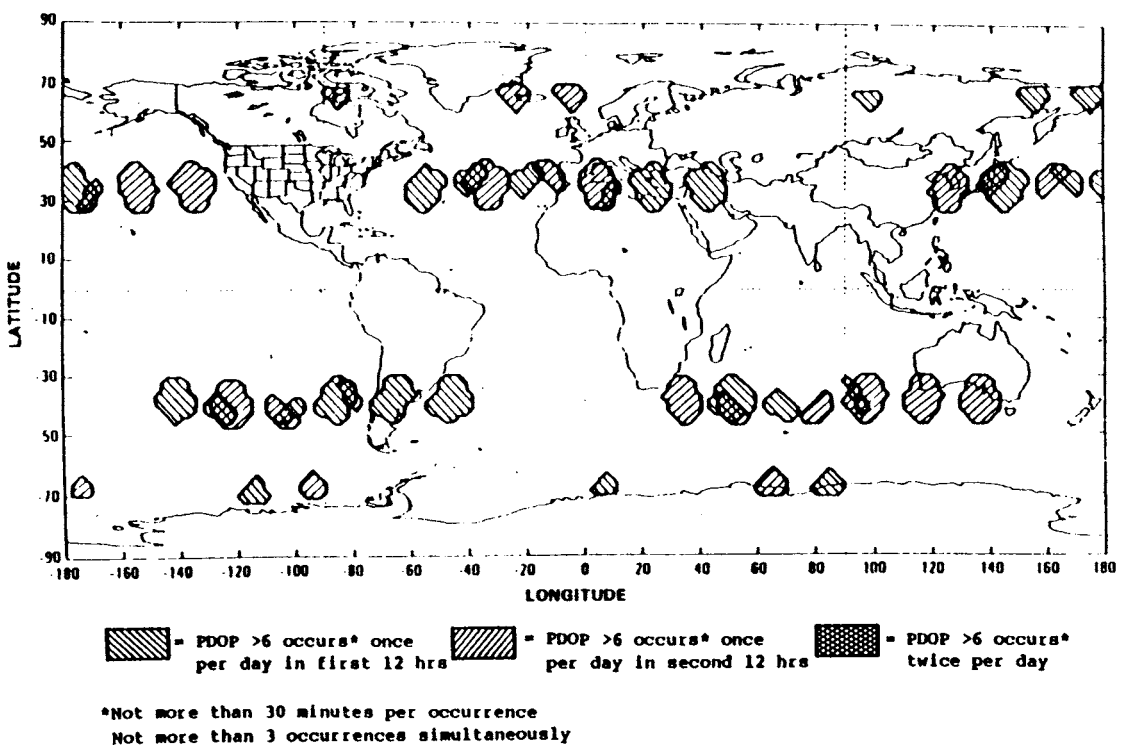
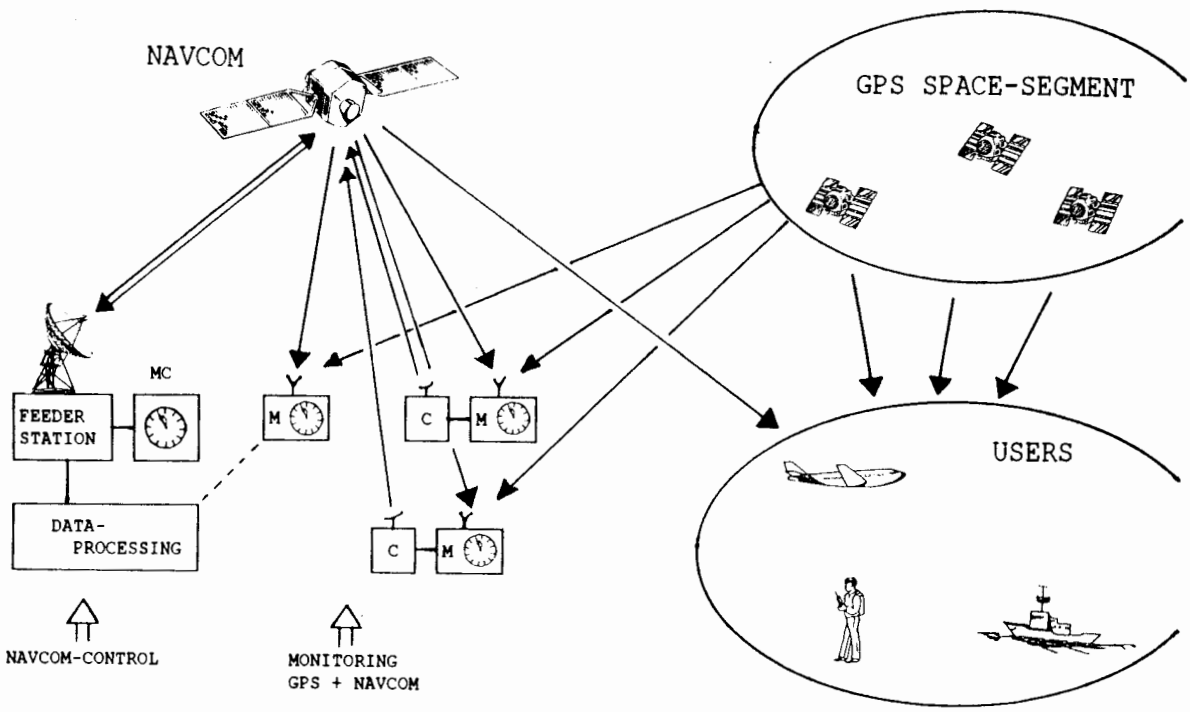


Figure 2. Daily Composite Operational Coverage (PDOP ≤ 6), 5° Elevation Mask Angle, 18 Satellites Plus 3 Active Spares



MC = MASTER CLOCK
M = MONITOR STATION
C = COMMUNICATION STATION

Fig. 3 PRINCIPLE OF A CIVIL COMPLEMENT TO GPS-NAVSTAR USING GEOSYNCHRONOUS SATELLITES

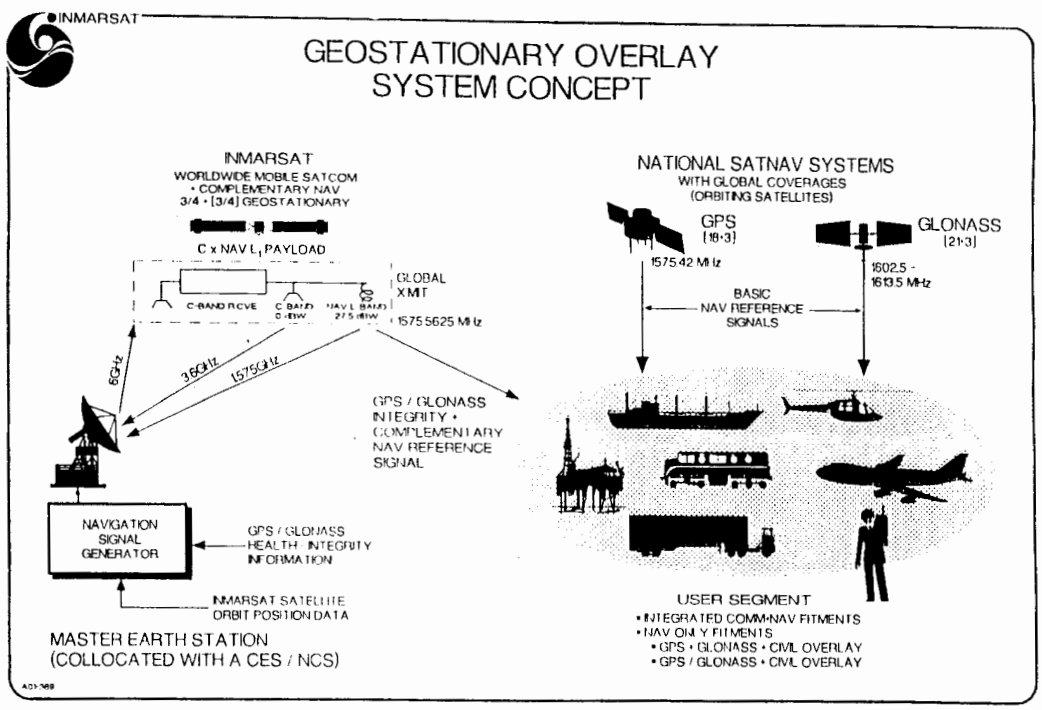


Fig. 4 The Complementary Overlay System Concept of INMARSAT

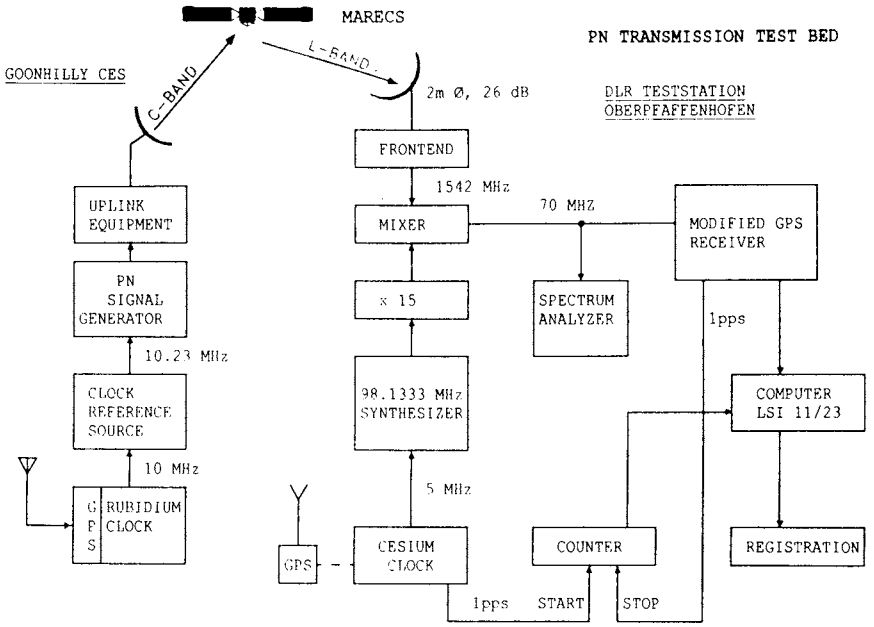


Fig. 5 PN Transmission Test Bed

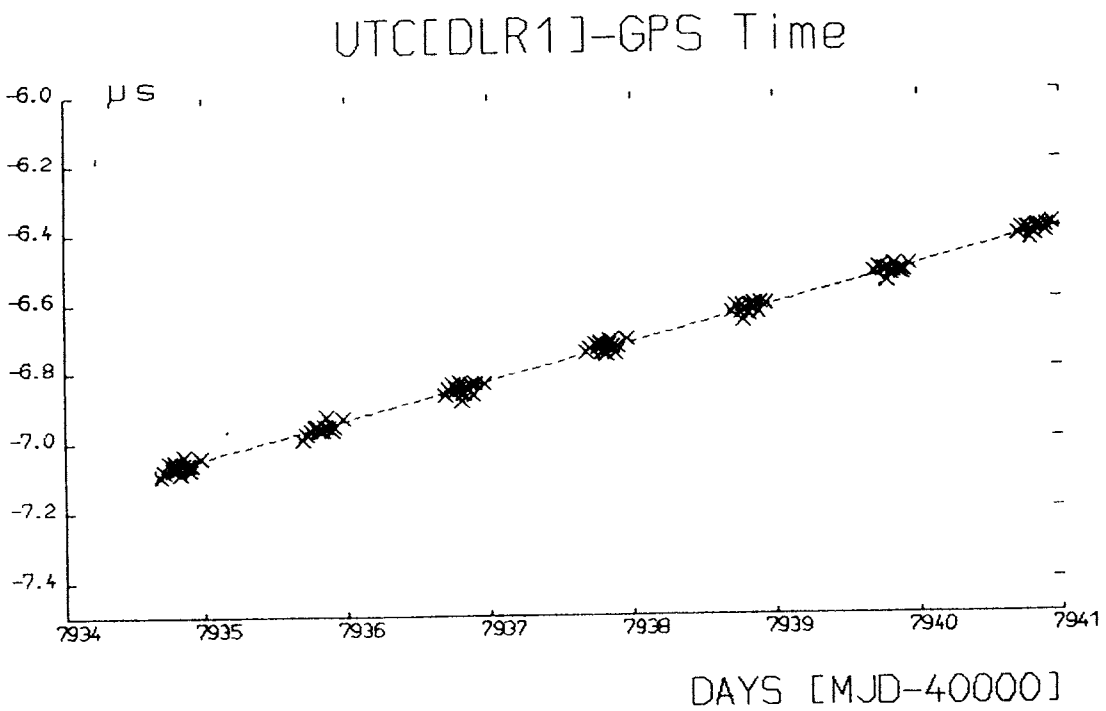


Fig. 6 Time Comparisons with GPS

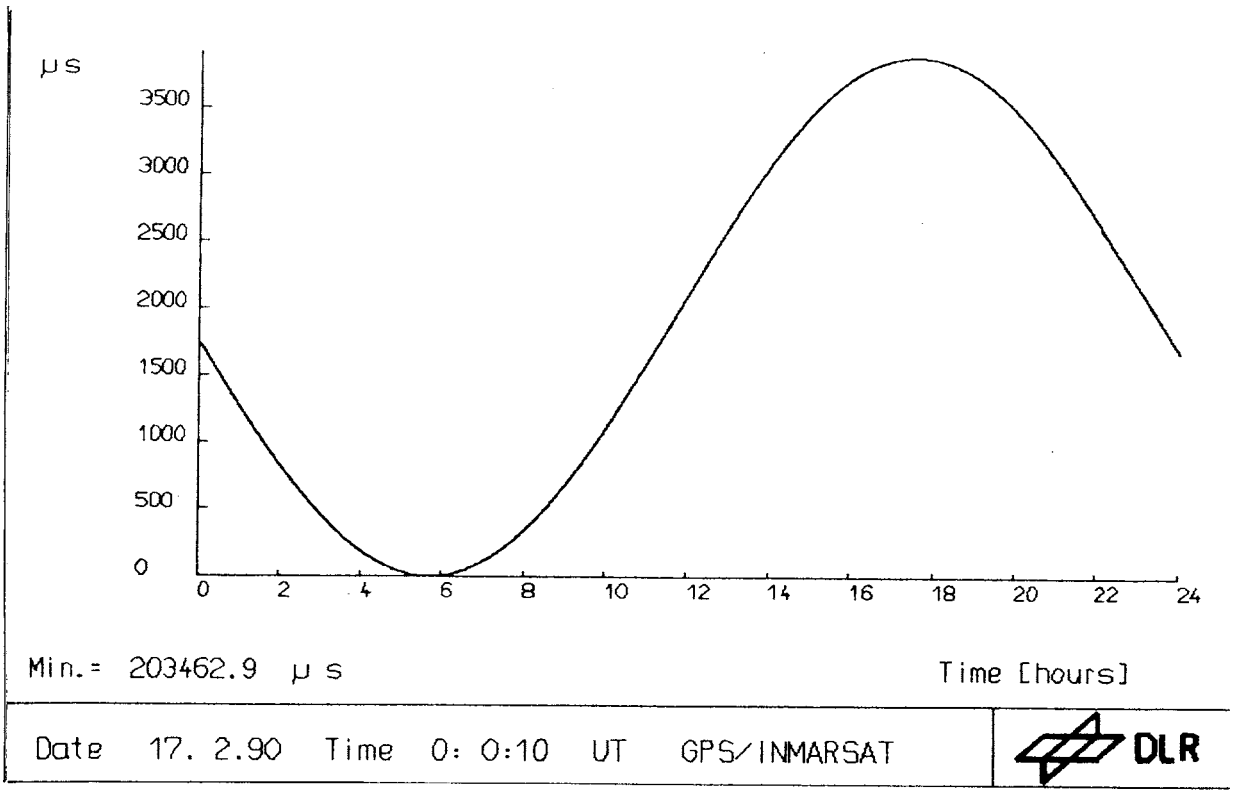


Fig. 7 Time Intervals Measured

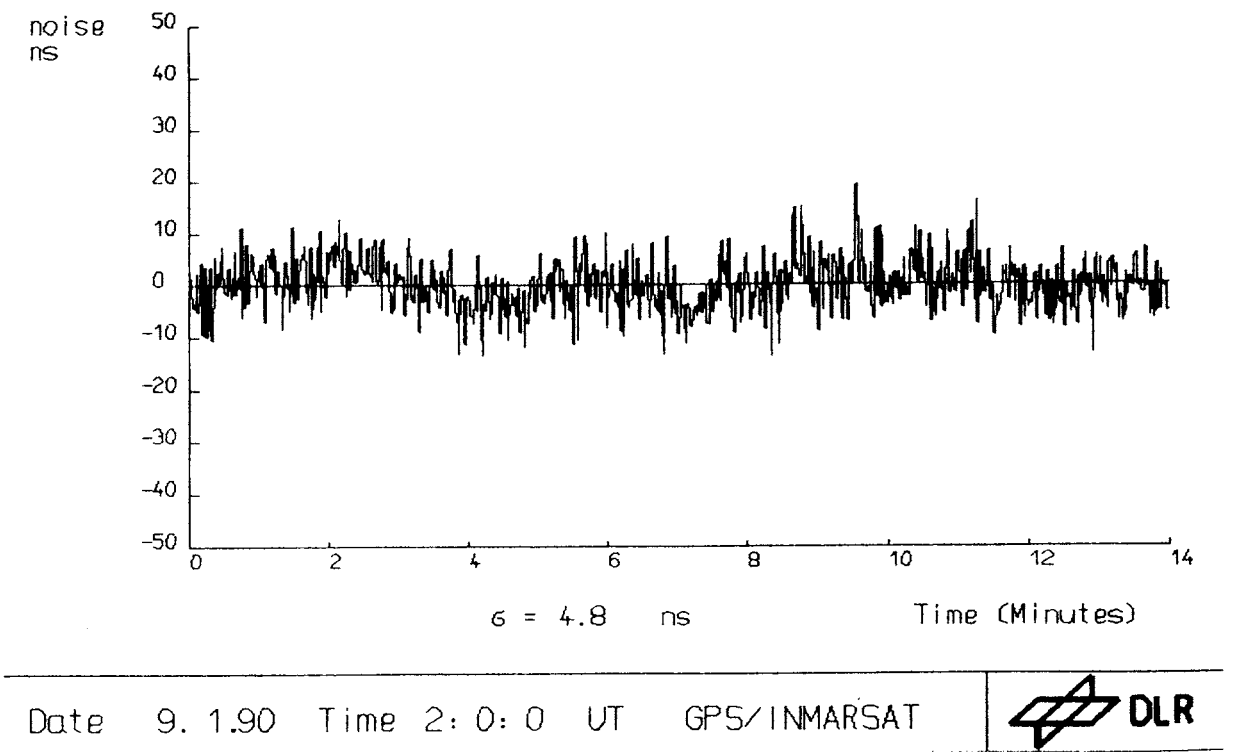


Fig. 8 Noise of Time Intervalls Measured

Session APP-02 Number 3

GPS / GLONASS -Receiver

Karl-Heinz Thiel
University Stuttgart
Institute of Navigation
D-7000 Stuttgart 1
F.R.G.

Introduction

GPS and GLONASS are the two similar satellite systems,where GPS already has proven its excellent performance. The GPS is said to have become the navigation system of the nineties. But the anounced Selective Availability (SA) will reduce the accuracy for civil positioning to about 100m SEP and for time synchronization to more than 100 nsec.

At the same time when USA installed its GPS the USSR built up their own satellite navigation system, GLONASS. In the last two years this system has become more and more attractive, due to the fact, that along with GLASNOST more information about GLONASS was offered to the West. A first bilateral agreement was signed on May 31,1988 by USA and USSR on cooperation in Transportation Science and Technology, which provided for a joint effort to develop civil navigation systems that could use signals from both GPS and GLONASS. The main forum for these exchanges is the Future Air Navigation System (FANS) committee initialized by the ICAO (International Civil Aviation Organization) after the Korean Airlines disaster. Thus the basis for a combined GPS/GLONASS receiver was given.

To show how this can be done, first a short overview is given for both systems, then differences and similarities are discussed, also the possible integration in one receiver. Finally, some results from our codeless receiving equipment are demonstrated.

Demands for integration of GPS and GLONASS

GPS itself is a military system that is funded and controlled by the Department of Defence. Although the system was designed to serve both civil and military user, there will be a degradation by Selective Availability (SA) due to reasons of national security. Beside this degradation there is another deficiency (shortage) with GPS concerning the temporary gaps in the coverage. For a short time when four satellites are needed for accurate 3D positioning, the forth satellite is out of view or the PDOP will increase above 6 (Fig. 1). This will also occur with a 21 satellite configuration with three active spares. Such a system constellation only provides a service with a probability of reliability of about 72% for 24 satellites.

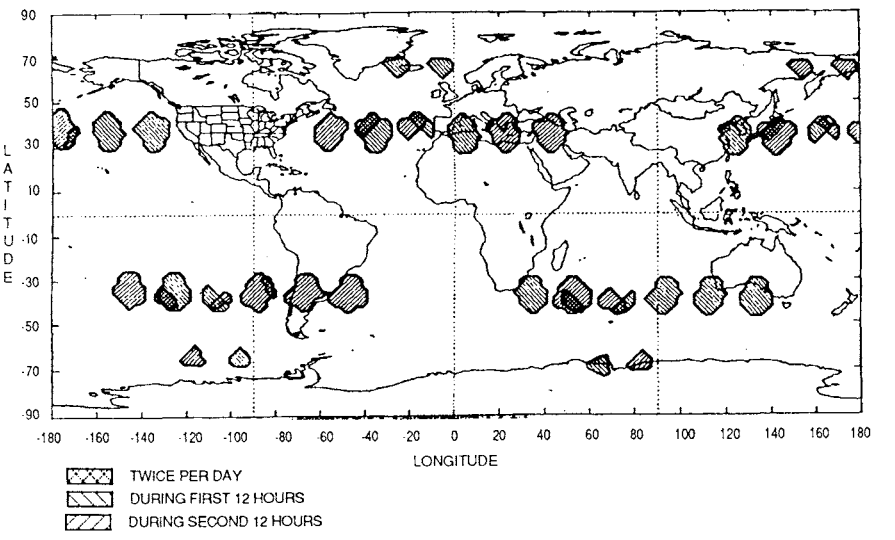


Figure 1: Temporary Gaps in coverage of the US Global Positioning System

However, designing a receiver to use both GPS and GLONASS there always will be enough satellites within view, thus also allowing for Receiver Autonomous Integrity Monitoring (RAIM). Against this background, the funding of the Lincoln Laboratory by the Federal Aviation Administration (FAA), to analyze a navsat receiver that can use both GPS and GLONASS signals, may be explained. In 1992 flight tests with this combined receiver are planned.

The Soviets seem to aim for a different policy now, though GLONASS is a also a military program. They have agreed to deliver the technical documentation for the GLONASS in February 1990 and they have stated that they will not degrade the system and perhaps make P-code access available to general user [see GPS-Bulletin]. This provides the possibility for combining both systems for cival usage and designing integrated receivers with better performance and higher system independence. The seriousness of the Soviets in the integration of both the systems for civil usage is demonstrated in their activities for looking for commercial joint ventures, as already realized with ASHTECH INC., when signing a memorandum of understanding to develop a combined GPS/GLONASS receiver with the ISDE (Institute of Space Device Engineering / USSR) (Source: GPS WORLD, March/April 1990).

System characteristics

In combining both systems, first the system characteristics have to be analyzed, to look for similarities and differences. In Table 1, the system characteristics are given. The most severe problem results from the inaccuracy of time synchronization of GLONASS to UTC and thus to GPS. The time accuracy achievable with GLONASS is specified to 1 msec, but the time synchronization is only given to Moscow time.

Also the geodetic reference systems are different and there must be a determination of the relative orientation elements between both systems. A possible way is the usage of the satellite systems by theirselves. Placing combined receivers at all Monitor Stations of both systems, these differences in time and geodetic reference systems can be solved. There are also differences in the Almanach data, transmitted by the satellites. The orbit position of a satellite is given by GPS with Kepler Elements, while with GLONASS the position is defined in the Greenwich rectangular geocentric coordinate system with speed and acceleration vector components. Even if the GLONASS satellite position calculation is easier, there is a need for different software algorithm.

Another severe difference is with distinguishing between different satellites. GPS uses code multiplex (CDMA) while GLONASS uses frequency multiplex (FDMA). Designing a multiple channels parallel receiver the GLONASS receiver will need a more complex RF part while the signal processing part will be almost the same.

SYSTEM CHARACTERISTICS

	GPS	GLONASS
ORBITAL ELEMENTS		
Number of Satellites in operational constellation	24	24
Orbital planes	6	3
Satellites per plane	4	8
Period	11h 56.9min	11h 15min
Ground track repitition	2	17
Drift per period	180°	169.4°
Altitude, km	20.200	19.100
Inclination	63°	64.8°
SIGNAL STRUCTURE		
Frequency bands		
L1, MHz	1565.2-1585.7	1602.5625-1615.5
L2, MHz	1217.4-1237.8	1240-1260
Trans. civil frequency L1, MHz	1575.42	Fj = F1 + (j-1)ΔF F1=1602.5625 ΔF=0.5625 j = 1,2,...24
Satellite separation	CODE MULTIPLEX CDMA	FREQUENCY MULTIPLEX FDMA
Bit rate (PN-Code)		
SPS (C/A), Mbit/s	1.023	0.511
PPS (P-Code), Mbit/s	10.23	5.11
Code repitition rate	1 msec	1 msec
Max. spectral power dens. IN L1, dBW/Hz	- 41	- 39 to - 44
Polarisation	right hand circular	right hand circular
Data rate, bit/s	50	50
Time synchronisation	? (40 ns)	1 msec

Table 1: GLONASS system characteristics

Launch date	Kosmos No.	GLONASS	Inclination degrees	Orbital Plane	Status
12 Oct 1982	1413	1	64,7	B	not op.
	1414	2	64,7	B	n/a
	1415	3	64,7	B	not op.
10 Aug 1983	1490	4	64,8	B	OK
	1491	5	64,7	B	n/a
	1492	6	64,8	B	not op.
20 Dec 1983	1519	7	65.3	A	OK
	1520	8	65.3	A	OK
	1521	9	65.3	A	not op.
19 May 1984	1554	10	65.5	A	OK
	1555	11	65.5	A	OK
	1556	12	65.5	A	not op.
4 Sep 1984	1593	13	64.8	B	not op.
	1594	14	64.7	B	not op.
	1595	15	64.7	B	OK
17 May 1985	1650	16	64.8	B	OK
	1651	17	64.8	B	OK
	1652	18	64.8	B	not op.
24 Dec 1985	1710	19	65.2	A	OK
	1711	20	65.2	A	OK
	1712	21	65.2	A	not op.
16 Sep 1986	1778	22	64.8	B	OK
	1779	23	64.8	B	OK
	1780	24	64.8	B	OK
24 Apr 1987	1838	25	64.9	A	Launch fail.
	1839	26	64.9	A	Launch fail.
	1840	27	64.9	A	Launch fail.
16 Sep 1987	1883	28	64.9	A	OK
	1884	29	65.0	A	OK
	1885	30	64.9	A	OK
17 Feb 1988	1917	31	64.8	B	Launch fail.
	1918	32	64.8	B	Launch fail.
	1919	33	64.8	B	Launch fail.
21 May 1988	1946	34	64.9	B	OK ?
	1947	35	64.9	B	OK ?
	1948	36	64.9	B	OK ?
16 Sep 1988	1970	37	64.8	B	OK ?
	1971	38	64.8	B	OK ?
	1972	39	64.8	B	OK ?

Source: SPACE markets, Vol.5, No. 3, July/August 1989
(Leeds University; TRW Space Log; Teledyne Brown Eng.; NASA)

Received satellites: No. 6,9,10,12,18,23,24

Table 2: GLONASS launch record

Combined receiver design

Regarding possible potential users of a combined receiver, on will find most demands from the civil aviation, due to high integrity demands. Other groups, like land transportation (trucks) and shipping can even work with the decreased accuracy of GPS alone. With combined receivers there is certainly one special group of high complex and accurate recevers, devoted for monitoring and determining the differences between both systems. Such receivers will be designed as multi-channel parallel receivers with precise clocks and coherent derived internal frequencies, to provide the accuracy demanded.

The more general solution, to be used with civil aviation, is shown in a block diagram in figure 2. A combined front-end with one antenna and one low noise amplifier (LNA) will receive all satellite signals of both systems. Separate RF-parts will distinguish between the signals and convert them to digital signals. Due to almost identical signal structures the signal processing units are almost the same. Depending on the knowledge of the differences in time and reference systems, the pseudo range and phase measurement can be used in one algorithm to solve for position and receiver time offset to UTC. While more than 4 satellites are in view, receiver autonomous integrity monitoring can be done, allowing for rejection of worse satellite signals.

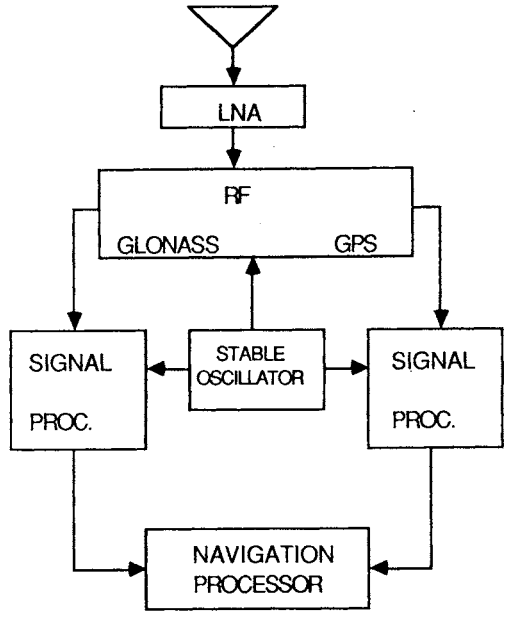


Figure 2: Combined GPS/GLONASS Receiver

Receiver Development at the INS

The INS is working on two different receiver types, a codeless one and a one (or more) channel PN-code digital receiver with phase measurement.

Codeless Receiver

This receiver makes use of some commercial instruments and is mainly devoted for monitoring of the satellites. A PC is controlling the complete measurement setup. This allows for continuous doppler measurement with one satellite or switching from one satellite to the next, thus producing an alert table of all 24 satellites for a complete GLONASS cycle of about 8 days. This measurement setup is already operational at INS and the gained results are given later.

PN-Code Receiver

This digital receiver is designed, such that, it can be used for GLONASS as well as with GPS. In a first step we will work with one channel in sequential mode. To realize high accurate measurements and position results, an atomic reference frequency source is used, deriving all other frequencies coherently. Figure 3 shows the block diagram of the receiver. The modular and general design allows for easy changing the the PN-Codes and code clock.

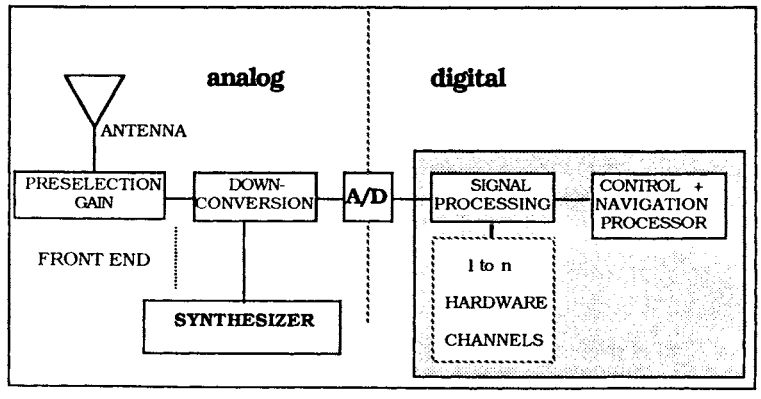


Figure 3 : Block Diagram of the PN-Code Receiver at INS

First Measurement Results

With the codeless receiver we have monitored the GLONASS satellites for several days, searching for all operating satellites. While with this receiver no data can be decoded, the status of the satellites can not be controlled. This means all the satellites that we were able to detect are given in the launch table (Table 2). Compared to the great number of launched satellites there is an enormous deficiency, which brings up the question on integrity and reliability of the GLONASS satellites itself and the complete system. In Fig. 4 and 5 the measured Doppler frequencies of the different satellites are plotted versus time. Regarding both figures the greater difference in vissibility and doppler shift within 24 hours due to the satellite configuration can be recognised, compared to GPS. After 8 days we will find almost the same configuration.

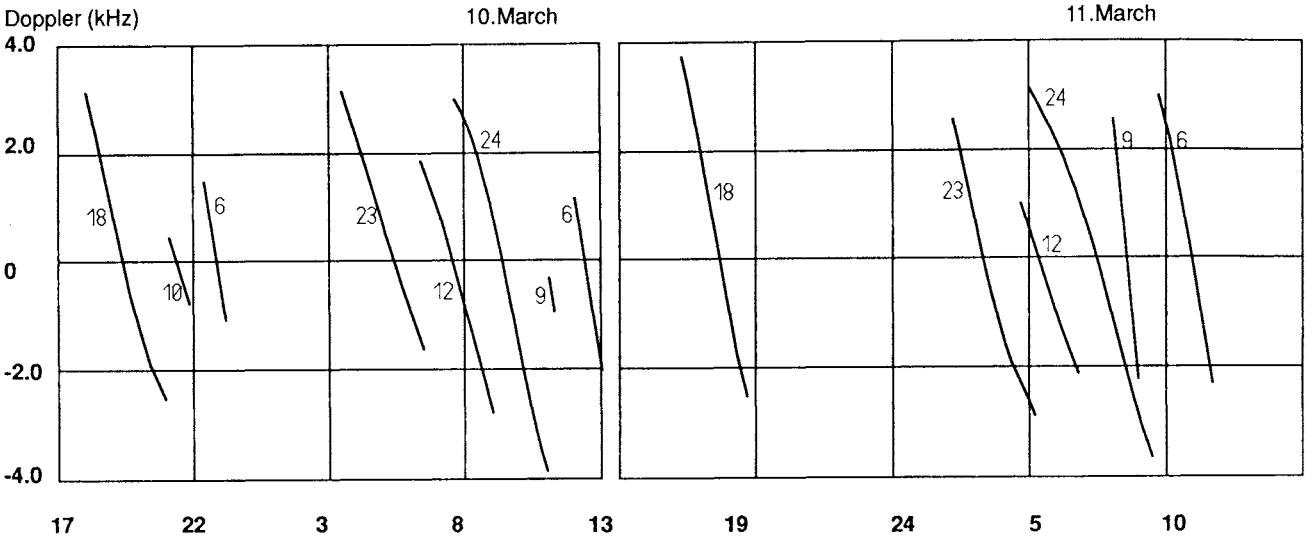


Figure 4 / 5: Doppler shift of GLONASS satellites

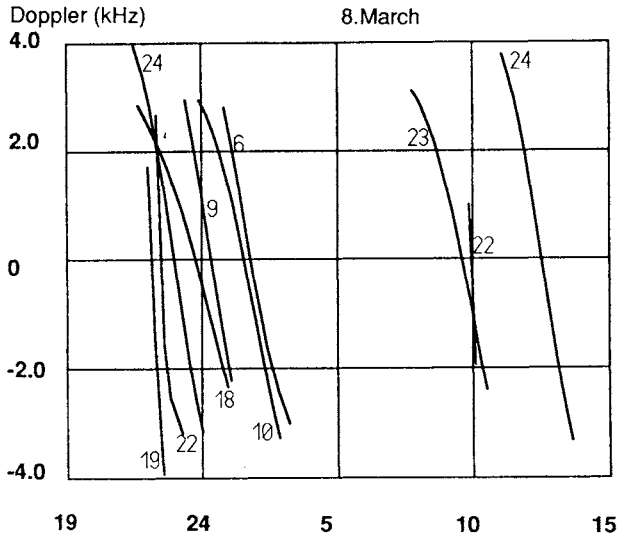


Figure 7: Doppler shift of GLONASS satellites with overlapping of satellites nr. 9 and nr. 10

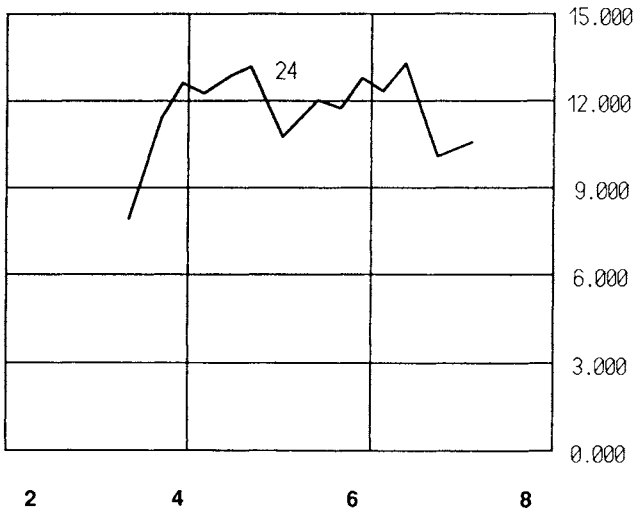


Figure 8. Signal strength of GONASS satellite

Figure 6 shows the received signal to noise ratio of the regained carrier.

In Fig.7 we found an astonishing effect. We received the two satellites, number 9 and 10 with a frequency offset of only one frequency step (0.5625 MHz) simultaneously. Formerly supposing that the frequency spacing of all satellites in view would be at least two steps (2x0.5625 MHz) to prevent an overlapping of the satellites signal spectra. Fig. 8 shows the spectra of the satellites and the amount of overlapping can be realized. This situation will cause an error in the pseudo-range measurement when the phases of both codes correspond to within one PN-code chip. The error produced with this effect is almost the same as the multipath error, often discussed with GPS. The amplitude of this error depends on the relation of the signal strength of both the received signals. If the signal of the selected satellite is weak compared to

the interfering satellite signal a false lock resulting in correlation loss may occur. Also the phase measurement can be spoiled due to a diminished code correlation. We will analyse this effect in more detail, as soon as we have finished our PN-Code receiver. But we suppose that with the operational system this effect will be removed, due to a proper satellite

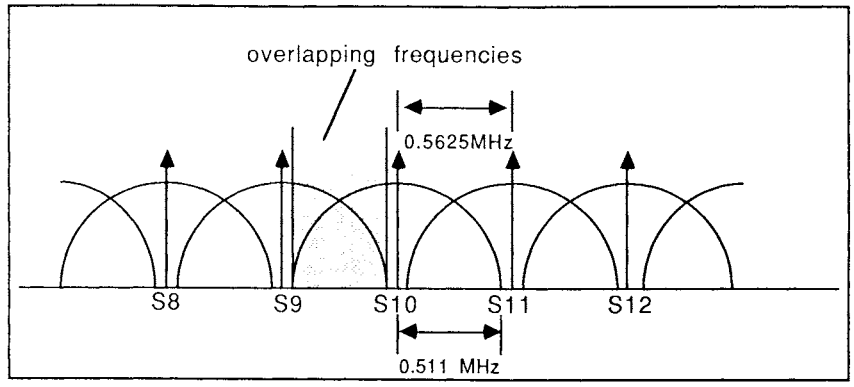


Figure 8: Overlapping spectras of GLONASS satellites (first mainlobe)

frequency assignement.

Conclusions

Combining both navigation systems in one receiver demands for some coordination between the conducting countries, especialy in establishing a common time and geodetic reference system. Also information on changes, possibilities, trends and failures must be available, thus allowing for an usage like with only a single system. The more complex receiver will be more expensive and thus reducing the number of possible users, only demands for reliability and integrity as with the civil aviation can promote the development of a combined GPS/GLONASS

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-02 Number 4

THE USE OF IONOSPHERIC DATA IN GPS TIME TRANSFER

M. Weiss, T. Weissert
National Institute of Standards and Technology

C. Thomas
Bureau International des Poids et Mesures

M. Imae
Communications Research Laboratory

K. Davies
National Oceanic and Atmospheric Administration

ABSTRACT

With solar activity near maximum, the largest error in the use of the Global Positioning System (GPS) for common-view time transfer can be the correction for ionospheric delay. This paper describes the use of data from two different codeless ionospheric measurement systems to correct common-view time transfer measurements. We first describe the design of the systems, then look at the data. We compare the measured ionospheric values with those from the ionospheric model parameters broadcast from the GPS satellites, computing time transfer values among three time standards laboratories: Observatoire de Paris, Paris, France; National Institute of Standards and Technology, Boulder, Colorado, USA; and Communications Research Laboratory, Tokyo, Japan. Combining these time transfers we can obtain closure around the world. Since this closure should add to zero, we obtain a clear measure of the quality of the overall time transfer system. We make another comparison between these results and ionospheric measurements using a Faraday rotation system.

INTRODUCTION

For most of the last decade, solar activity has been low and the contribution of the ionospheric delay model error to the error budget for GPS common view time transfer has been tolerable [1]. However, since the model accuracy is about 50%, an increase in the overall delay increases the error in using the model. Now that we are approaching a maximum in the approximately 11 year sunspot cycle, the L1 ionospheric delays are approaching 100 ns for the low angles used for intercontinental time transfer. We therefore need to be able to measure the ionospheric delay rather than use the parameters from the transmitted model. Cost and security dictate that we consider only a codeless type receiver, as opposed to a unit that fully decodes the precise code (P-code).

Two such calibrators with somewhat different designs are used in this study. M. Imae built a calibrator with the support of the Bureau International des Poids et Mesures (BIPM) while employed at the Communications Research Laboratory (CRL). The first prototype of these BIPM type calibrators is still operating at the BIPM. An improved commercially available version is located at CRL, Kashima Space Research Center (SRC). Data from both of these were used in this paper. A different calibrator was built at the National Institute of Standards and Technology (NIST), and this design is now being deployed at various locations around the world. For this study, data from the NIST type of calibrator were available primarily from that at NIST, with a small amount of data available from an identical calibrator at the Observatoire de Paris (OP).

For this paper, ionospheric measurements were applied to GPS data from MJD 47886 to 47913, much of the month of January, 1990. For data from Paris, we used the calibrator at the BIPM to correct GPS data from OP. In Japan the ionospheric calibrator was at Kashima SRC and the GPS receiver was at the CRL laboratory at Koganei about 100 km away. In Boulder, both the NIST ionospheric calibrator and a GPS receiver were located at NIST. The GPS data were taken according to the BIPM schedule at each location, thus insuring simultaneous measurements pairwise between locations.

Calibrator Design: The BIPM Calibrator

The BIPM calibrator conceived and built by Michito Imae was the first codeless ionospheric calibrator. It measures the ionospheric delay by measuring the relative delay of the phase of the received P-code clock between the two L-band GPS frequencies, L1 and L2. These frequencies are 1.575 GHz and 1.228 GHz respectively. Since the delay of the P-code through the ionosphere is inversely proportional to the square of the carrier frequency, the total phase delay of the code on either frequency is proportional to the differential phase change between them. We illustrate this below. Let:

dphs1 = the change in phase of the code at L1,
dphs2 = the change in phase of the code at L2.

Since the L1 frequency is 154 times the P-code clock frequency, N = 10,23 MHz, and L2 is 120 times the same frequency, we may write

$$dphs1 = k/(154*N)^2, \quad \text{and} \quad dphs2 = k/(120*N)^2,$$

where k is proportional to the total electron content along the signal path.

If the code starts out in phase between L1 and L2, then we measure (dphs2 - dphs1), the differential phase change. The total delay on L1 then is:

$$dphs1 = [dphs2 - dphs1]/[(154/120)^2 - 1].$$

Thus, the total delay is proportional to the received phase difference. As suggested by MacDoran prior to 1982, the differential phase between the L1 code and the L2 code can be measured without knowing the code itself, since it is the same on both frequencies [2,3,4].

The BIPM calibrator has been described in detail elsewhere [5,6], so we simply outline its construction here. The systems at the BIPM and at CRL both receive the L1 and L2 signals through a directional antenna which provides some gain, cuts back multipath effects, and helps eliminate interference between satellites. The L1 and L2 signals are amplified and multiplied down on separate channels, then are cross-correlated. Two cross correlations are done between the L1 and L2 channels: one direct, and one with a fixed delay in the L1 channel. From these two outputs, the differential delay between the phases of the two codes can be determined.

The system tracks a single satellite for 4 min, taking 1 min to point the antenna and 3 min for measurement. The noise level after 4 minutes of averaging is at the nanosecond level. The system computes total ionospheric delay of the L1 code, as well as the vertical delay. This data is stored in a microcomputer which operates the system. The calibrator tracks based on an entered schedule. The international common view schedule divides the day into 16 min tracks. The common practice is to schedule four, 4 min tracks by the ionospheric calibrator on the satellite being tracked during that 16 min according to the BIPM schedule.

The calibrator is a stand-alone unit. It needs a 5 MHz source with at least the stability of a good quartz oscillator, and an almanac of the GPS satellites no more than about 3 weeks old.

Calibrator Design: The NIST Calibrator

Like the BIPM calibrator, the NIST calibrator also was described in detail elsewhere [6], so we summarize its design here. It is also a codeless ionospheric calibrator that recovers the P-code clock on L1 and L2 and uses the phase difference to compute the L1 ionospheric delay. Major features of the hardware include dual-volute omni-directional antennas on a choke ring ground plane, very low-noise front end, and alternate L1--L2 phase sampling through a common IF channel. S/N of the recovered P-code clocks is typically positive by several decibels in a 100 Hz bandwidth. Signals are processed as 8 bit data, with all satellites in view individually (and simultaneously) tracked in real time to recover the ionospheric delay values. All processing is with an internal 8 bit CMOS microprocessor.

This system differs from the BIPM calibrator both in that it uses an omni-directional antenna, receiving signals from all satellites in view, and it samples the L1 and L2 frequencies sequentially. Instead of correlating the two signals, the NIST calibrator uses a delay-and-multiply operation to recover the P-code clocks for all satellites in view on one of the L-band frequencies, then samples this signal at 250 Hz with 8 bit samples. The software measures the phase of the P-code on that frequency for all satellites as follows. Satellites are separated by their different Doppler shifts and the rates of change of these Doppler shifts. The phases of the P-code clocks are recovered by multiplying the sampled signal by sine($-\omega p$) and cosine($-\omega p$) and summing the sine and cosine components for 7.5 s. A four-quadrant arc tangent of the quotient of the sine sum over the cosine sum yields the clock phase. Measurements on L1 and L2 are carried out sequentially; that is, L1 is measured for 7.5 s followed by a 7.5 s measurement of L2. This is analogous to the tau-dither loop in a GPS receiver and provides similar advantages and disadvantages. The major disadvantage is a degradation of the signal-to-noise by 3 dB. Advantages include greater simplicity and stability; since the same narrow-band IF is used for both L1 and L2, the potential for errors due to unequal phase shifts through separate channels is minimized.

To separate satellites and determine their code phases the NIST calibrator needs to know the Doppler shift on each satellite to better than 10 mHz. This cannot be done using the broadcast ephemeris from the satellite. Instead, the NIST calibrator uses a second-order frequency-locked loop. Phase difference from one L1 measurement to the next is used to compute frequency error in (ωp) and close the tracking loop.

When development is complete, the normal mode of operation of the ionospheric calibrator will be in parallel with a GPS coarse-acquisition (C/A) code time transfer receiver, with the measured ionospheric data automatically available to replace the modelled ionospheric data transmitted by the satellite. In this way, the ionospheric calibrator operation will be nearly transparent to the time-transfer user.

The calibrator can also be used as a stand-alone unit for ionospheric measurements, with values of L1 delay for all satellites in view available from one of the two serial ports as often as every 15 s. Also, these data are stored after smoothing for 15 min with a linear fit and are available via modem for up to two weeks, depending on the number of satellites in the system. In this mode, the calibrator must be set on time within ± 1 min and requires a copy of the GPS almanac, updated every 1 to 4 weeks. A stable source of 5 MHz power is required, preferably from a cesium or rubidium standard; however, a crystal oscillator with a drift rate of parts in 10^{-9} per day is satisfactory.

The stability of this ionospheric calibrator is 1-4 ns for 15 saverages with the ultimate limitation to accuracy being multipath interference [8]. The antennas are equipped with choke ring ground planes. The L1 radiation patterns of three ground planes were measured on the antenna test range. The ground planes provided at least 7 dB additional attenuation at 0° elevation, compared to the bare helix. Until now, it has

not been determined whether noise figure or multi-path interference reduction is predominant in the measured noise reduction; however, since both a good noise figure and reduced multipath effects are desirable, the calibrators are equipped with choke-ring ground planes. Multipath interference was studied by comparing the measurements between two different calibrators. It was found that by adding microwave absorbent material inside of the inner choke ring, variations between the two systems were reduced to peak-to-peak values of a few nanoseconds.

Comparison Between the BIPM and the NIST Calibrators

In figures 1 and 2 we see three days of ionospheric delay data converted to the vertical delay, taken in Paris by the two calibrators: the NIST calibrator at OP in figure 1 and the BIPM calibrator in figure 2. The conversion assumes that the ionosphere is a spherical shell with uniform electron density. A correction is applied for azimuth and for the curvature of the earth. The NIST data plot vertical delays for many satellites tracked simultaneously at varying elevations and azimuths, thus having different points of intersection with the ionosphere for the vertical delay computation. Since the satellite elevation can vary from 15° to 90°, using a vertical height of 300 km for the ionosphere we see that the ionosphere could be pierced anywhere in a circle of diameter 2000 km. Thus we expect the true vertical delay to have a significant variation among different satellites measured simultaneously, especially near sunrise and sunset. With this in mind, the agreement between the two units seems very good. Some of the variation between measurements is also due to differences in the hardware of the satellites. Our measurement of the ionospheric delay depends on the phase coherence of the code between L1 and L2 at transmission. In fact there are some small unknown delays between these phases which differ among the satellites.

Comparison with Faraday Rotation

Figure 3 shows a comparison between the vertical ionospheric delay from the NIST unit and the corresponding data from the Faraday polarimeter at Boulder. This system records the 136.38 MHz signal from the GOES-3 geosynchronous satellite located near 0° N, 130.1° W. The ray path transits the ionosphere near 36.4° N,109.4° W; the coordinates of Boulder are 40° N, 105.3° W. The plot shows data from the month of January, 1990 averaged daily to give the average diurnal effect. 0 h UT is 5 pm local time. Most of the GPS satellites were up at night during this month. At night the GPS values are higher than the Faraday values. This is to be expected since the Faraday technique is sensitive to electrons up to a height of only about 2000 km, whereas the GPS measurements are affected by electrons up to the height of the satellites [9]. The difference is partly due to the protonospheric content [10], and partly due to the unknown phase delay between the L1 and L2 codes among the various satellites. The discrepancies during the day are probably due to the different geographical locations between the Faraday and the GPS raypaths, as well as the relatively small GPS data samples taken during these times.

GPS Time Transfer with Ionospheric Calibration

Three GPS common-view time transfers were computed using the two types of GPS ionospheric calibrator: UTC(OP) - UTC(NIST),UTC(CRL) - UTC(OP), and UTC(NIST) - UTC(CRL). The GPS data were taken according to the BIPM schedule at each location, thus insuring that there be simultaneous measurements pair-wise between locations. Strict common-view data was used: each measurement needed to be present at each leg of a pair of stations. Each branch was corrected either using the modelled or the measured ionospheric values.

We compared the deviation in time transfer values over each day between those using the GPS modelled ionosphere and the measured. This deviation is due primarily to either satellite ephemeris error, errors in the ionospheric correction applied, or multipath disturbances [1]. Other possible sources of deviation include tropospheric modelling errors, incorrect antenna coordinates, and clock dispersion between measurements. However, we think these latter errors do not contribute significantly here. For the comparison between the modelled and measured ionospheric values we insisted that both forms of common-view data be available before using a point in computations.

We have simply determined if the deviations among daily tracks decrease with the use of measured ionospheric values. We could later study more carefully the sources of these deviations.
Time transfer using a particular satellite and time of day gives a time series of one point per sidereal day. It is possible to study these time series and determine if there is a consistent statistically significant bias, and if the variations around this bias are white.

We present the average daily standard deviations in Table 1 below. The results were surprisingly disappointing. The only time transfer for which there was a decrease in the average daily standard deviation was UTC(NIST) - UTC(CRL) (figures 4 and 5). Looking at the tracks themselves we see that many were at night for each pair of locations. At night the ionospheric delay is much smaller hence variations are smaller and the model is perhaps better. On the other hand, we used the results from these time transfers to compute a closure around the world. For the closure we found a significant improvement using the measured ionospheric values.

Table 1: Average Standard Deviations

From Common-View Time Transfer

Labs	Using Modelled Values	Using Measured Values
UTC(OP) - UTC(NIST)	8,3 ns	10,6 ns
UTC(OP) - UTC(CRL)	7,1 ns	10,1 ns
UTC(NIST) - UTC(CRL)	11,5 ns	8,6 ns

For the path OP - NIST there were seven tracks per day. Of these, six were day tracks at OP and night tracks at NIST, while one was a night track at OP and a day track at NIST. It is surprising that there is not some improvement using the BIPM calibrator since so many tracks were during the day at OP. In a previous report an improvement was found for transfer from OP to the United States Naval Observatory in Washington, DC, USA [6]. One possible explanation is that the BIPM calibrator does not work as well for low elevation angles, and three of the daytime tracks were below 31° at Paris. It is also true that ephemeris error becomes more pronounced with larger baselines. Similarly for the path OP - CRL, all the tracks were at night for OP and five were during the day at CRL. But of these five, three had elevations below 30° at CRL. A previous result showed an improvement for this same path [6], but it was done at a different time when the satellites were in different positions. In addition, in the previous result there were more data. For the path NIST - CRL all the tracks were at night at both locations except one during the day at CRL.

In figure 6 we show the closure around the world using the time transfer values discussed above. The average values and their standard deviations are shown in table 2 below.

Table 2: Closure Around The World

Ionospheric Estimate	Average	Standard Deviation
Modelled	12,7 ns	5,4 ns
Measured	- 4,9 ns	3,4 ns

There is a clear improvement with the use of the measured ionosphere. We must conclude that even though the daily deviations were not lessened for individual time transfers in this experiment, the deviation did drop for the closure, and there was a bias in the use of the GPS ionospheric model.

CONCLUSIONS

We have seen two different ionospheric calibration systems, both giving similar measurements of the ionospheric delay from GPS satellites. We have shown here that the NIST calibrator agrees with Faraday rotation measurements. The same has been shown previously for the BIPM calibrator. We applied these measurements to GPS common-view time transfer, and found that, in this experiment, the standard deviation across the different tracks each day does not generally decrease, though it does in one of the three cases. We do find, however, that combining the three time transfers using the ionospheric measurements to obtain a closure yields an improvement. We conclude that if the baselines are long and many tracks are taken at night, use of the GPS ionospheric model can produce time transfer results which are as self consistent as results which use measured ionospheric values. But the measured values still improve the accuracy of the time transfer. We have also shown that the ionospheric calibration systems can make many measurements of ionospheric delay, and these should be quite useful for studying the ionosphere itself.

ACKNOWLEDGEMENTS

The authors wish to thank Pierre Uhrich of the Laboratoire Primaire du Temps et des Fréquences at OP for kindly providing some of the data used in this study.

REFERENCES

1. M. A. Weiss, D. W. Allan, "An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data," IEEE Trans. IM., Vol. I&M-36, No.2, pp 572-579 (June, 1987).

2. P. MacDoran, D.J. Spitznesser, L.A. Buennagel, "SERIES: Satellite Emission Range Inferred Earth Surveying," Proceedings of the Third International Symposium on Satellite Doppler Positioning. New Mexico State University (February 1982) 1143-1164.

3. P. MacDoran, D.J. Spitznesser, U S Patent 4,797,677 Jan 10, 1989.

4. R. Bruce Crow et al., "SERIES-X Final Engineering Report" JPL D-1476 Jet Propulsion Laboratory, Pasadena California (August 1984).

5. C. Thomas, M. Imae, W. Lewandowski, and C. Miki , "A Dual Frequency GPS Receiver Measuring Ionospheric Effects without Code Demodulation and its Application to Time Comparisons,"Proc. of the Twentieth Annual Precise Time and Time Interval(PTTI) Applications and Planning Meeting, Nov. 29-Dec. 1,1988, pp. 77-85.

6. M. Imae, C. Miki, C. Thomas, "Improvement of Time Comparison Results by Using GPS Dual Frequency Codeless Receivers Measuring Ionospheric Delay," Proc. of the Twenty-first Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Nov. 28 - 30, 1989.

7. D. Davis, M. Weiss, 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), Sep. 25 - 29, 1989, Colorado Springs, Colorado.

8. G. J. Bishop and J. A. Klobuchar "Multipath effects on the determination of absolute ionospheric time delay from GPS signals" Radio Science, v. 20, no. 3, pp. 388-396, (May-June 1985)

9. J.E.Titheridge, "Determination of Ionospheric Electron Content from the Faraday Rotation of Geostationary Satellite Signals," Planet. Space Sci., vol.20, p.353.

10. K. Davies, Ionospheric Radio, chapter 8, P. Peregrinus, 1990,London.

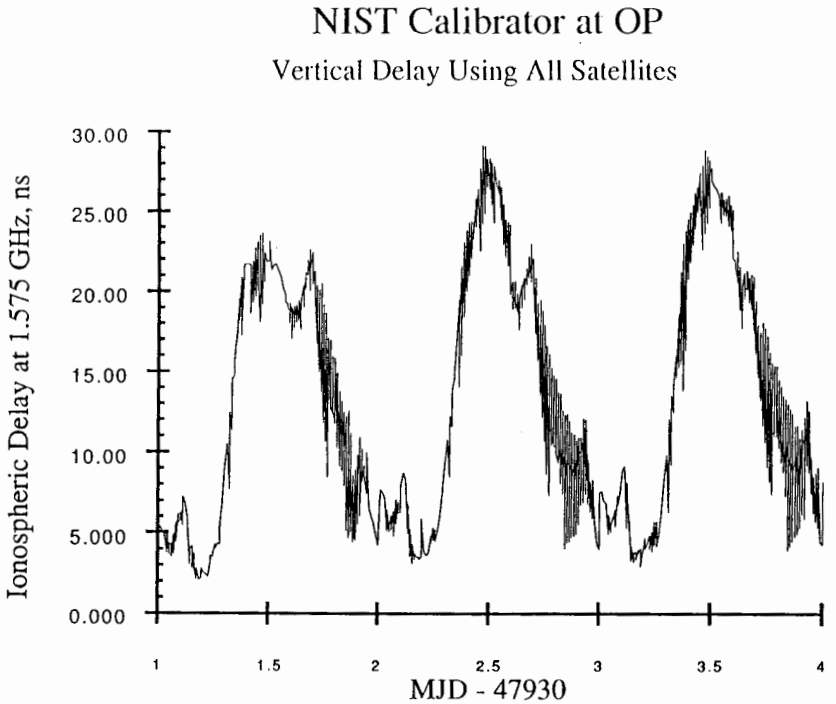


Figure 1: Data from the NIST ionospheric calibrator at OP converted for the vertical delay. Since the calibrator tracks all satellites in view there can be significantly different points of intersection with the ionosphere at one time. Hence there is a spread in the vertical delay at a given time. In addition, there is a variable bias between satellites and perhaps measurement noise.

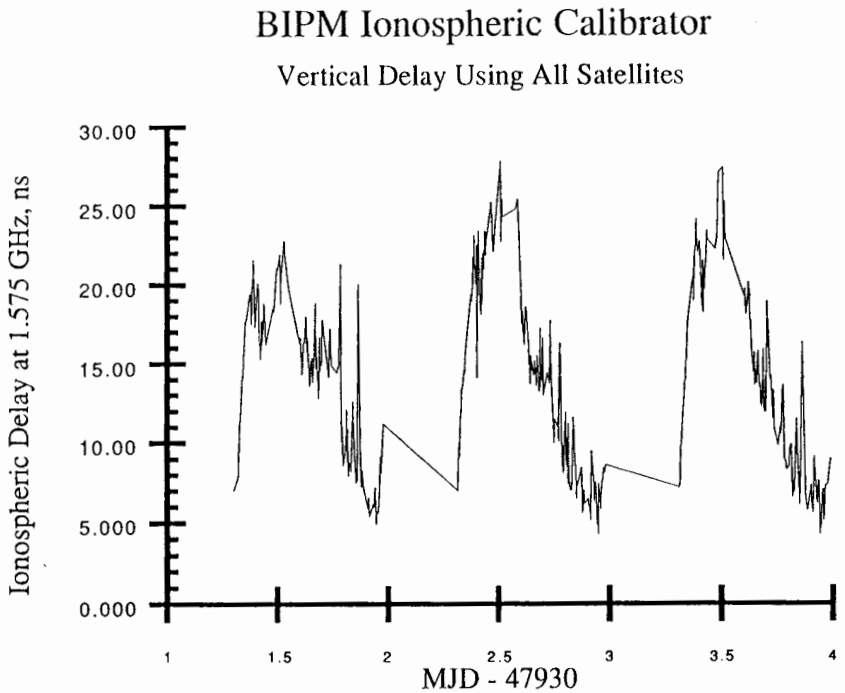


Figure 2: Data from the BIPM ionospheric calibrator. Fluctuations here can be due to the same effects as for the data in figure 1.

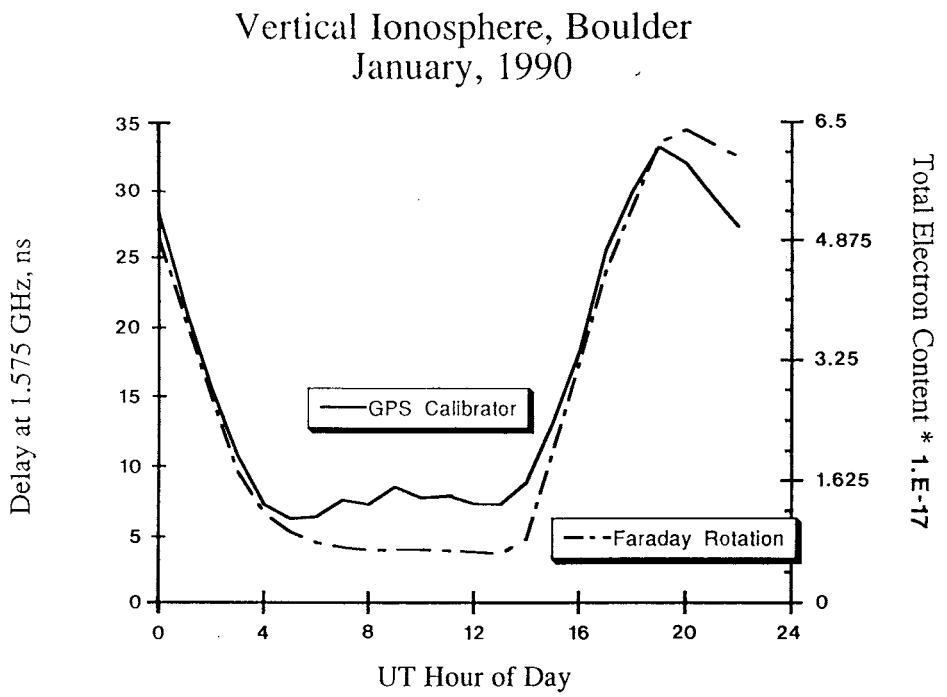


Figure 3: A comparison of measurements of the ionosphere using two different techniques: the NIST/GPS calibrator and Faraday rotation from a GOES satellite. Faraday rotation measures the ionosphere only up to 2000 km. Signals from GPS satellites pierce the ionosphere in different places than that from the GOES satellite. Also, there are biases in data from GPS satellites.

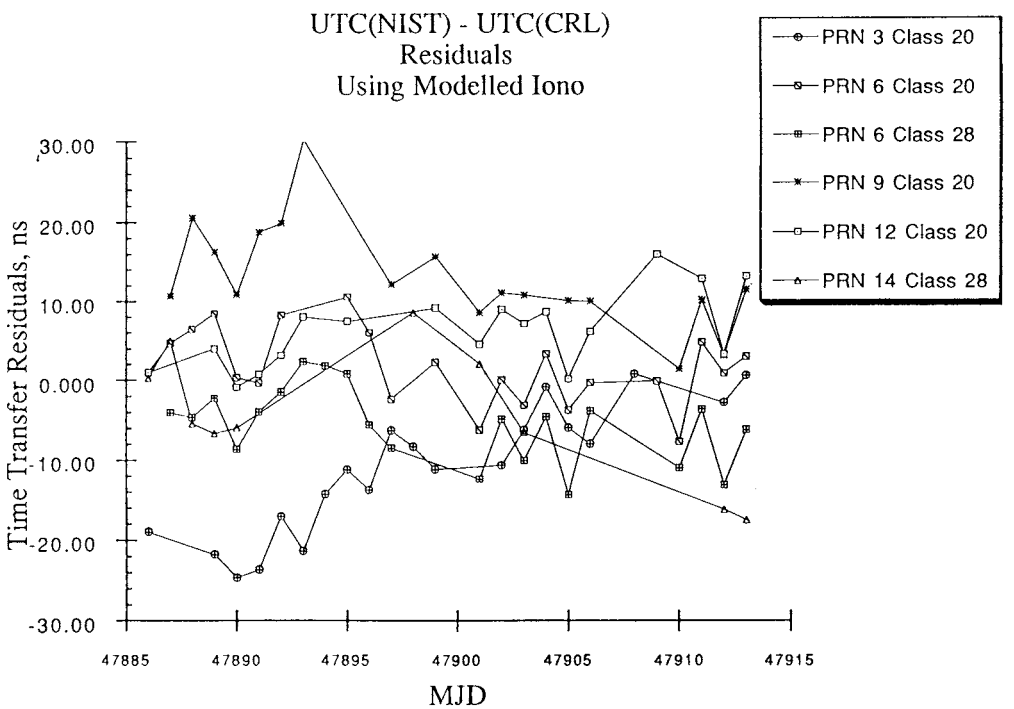


Figure 4: Time transfer residuals from NIST, Boulder to CRL, Tokyo with the data corrected for the ionosphere using the broadcast GPS model parameters.

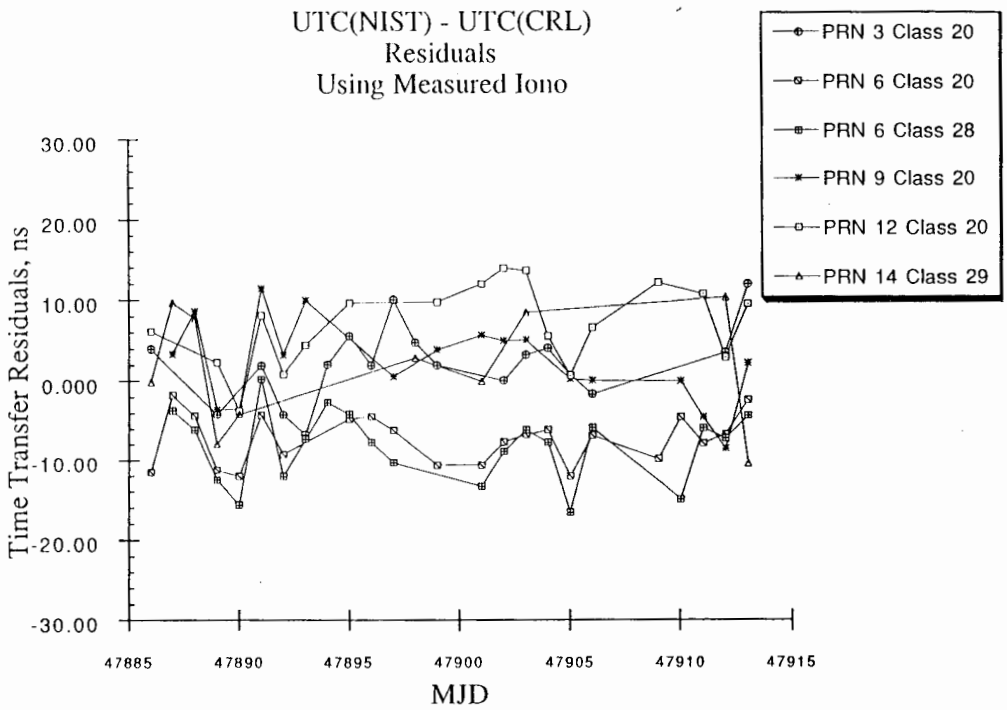


Figure 5: Time transfer residuals from NIST to CRL with the data corrected for the ionosphere using measured values from the NIST type calibrator in Boulder and the BIPM type calibrator in Tokyo.

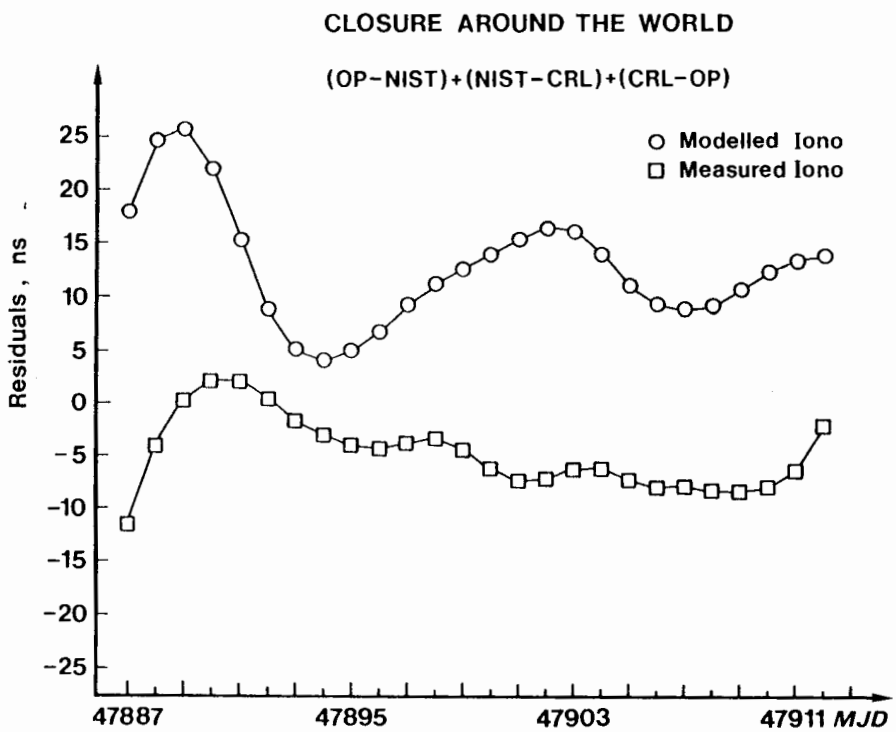


Figure 6: Closure around the world using the time transfer links: UTC(OP)-UTC(NIST) + UTC(CRL)-UTC(OP) + UTC(NIST)-UTC(CRL), computed with measured and modelled ionospheric delays. There is an improvement in the closure using the measured values, a decrease both in the bias and the standard deviation.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-02 Number 5

TOWARDS SUBNANOSECOND GPS TIME TRANSFER USING GEODETIC PROCESSING TECHNIQUES

T. Schildknecht, G. Beutler, W. Gurtner, M. Rothacher

Astronomical Institute, University of Berne,
Sidlerstrasse 5, CH - 3012 Berne, Switzerland

Abstract. Today GPS time transfer is mainly realized using C/A-Code receivers with the common view method. The method is limited by the satellite broadcast information, the uncertainities of the receiver antenna position and the wavelength of the C/A-code signal. This situation can be improved considerably using phase measurements from geodetic GPS receivers and appropriate processing techniques.

1. INTRODUCTION

GPS time intercomparisons using the so-called "common view" method are in principle performed in the following way: The time of arrival of the same C/A-code signal (from one satellite) is observed at two stations with respect to the local reference clock. In order to obtain the difference between the two reference clocks the differential signal travel time between the satellite and the stations has to be calculated. There the satellite position, the coordinates of the receiving antennae (all in the same geodetic reference frame!), and atmospheric delays for both signal paths are potential error sources.

Satellite positions and a set of parameters for a global ionosphere model are obtained from the satellite broadcast messages. As shown in investigations concerning GPS orbit determination the actual broadcast orbits are accurate to 10 to 30 meters (Remondi, 1989). This is a severe limitation especially on long baselines (USA-Europe). In addition the US Departement of Defense (DoD) will degrade this information in the near future (selective availability with an orbit accuracy of ~100 m).

A position error of the receiving antennae enters with roughly 1 ns per 30 cm. It should be pointed out that these positions have to be known in the same (global) reference system as the satellite orbits (e.g. WGS-84). This is usually not the case unless the station coordinates were determined using high precision geodetic methods (GPS phase measurements, Satellite Laser Ranging, VLBI).

A further major contribution to the error budget comes from the atmospheric delays. Differences in the ionospheric delays of the two paths trough the atmosphere start to be important on baselines of 100 kilometers and longer. Considering the high spatial and temporal variability of the ionospheric electron content measured during recent geodetic GPS campaigns it is evident that the use of the general model of the broadcast messages does not much improve the results.

Finally we have to consider the noise of the used C/A-code signal. 10 to 30 meters is a good value for the best currently available receivers. Compared to the phase measurement noise of several millimeters for geodetic receivers the C/A-code has a poor intrinsic resolution.

2. USING RESULTS FROM GEODEDIC APPROACHES

We may profit in many respects from investigations in the field of high precision geodetic GPS applications. Most of the above mentioned critical points are also research topics in geodesy.

In most cases the geodesist is interested in differential quantities only. In a local network of GPS stations for example the primary goal will be the determination of the relative geometry (today on the centimeter level), whereas orientation of this network in a global reference system is of minor importance. This leads to differential processing techniques with the consequence that many errors will be scaled down considerably. Let us consider the principal error sources in this context without going too much into details.

Geodetic results of GPS measurements depend to a large extent on the quality of the satellite orbits used. It has been shown that the error db in the determined components of a baseline of length b between two receivers R1 and R2 is correlated (in a first order

approximation) with the satellite position error dr in the following way (Bauersima, 1983):

$$\frac{db}{b} = \frac{dr}{r} \qquad (1)$$

Where
- b is the baseline length
- r is the average distance between satellites and receivers
- db is the error in one of the components of the baseline b
- dr is the orbit error

In our case, since we are not determining baselines (differences of station coordinates), but differences of station clock readings, the baseline error db has to be interpreted as a synchronization error dt of station 2 with respect to station 1 where

$$dt = \frac{db}{\text{speed of light}} \qquad (2)$$

Table 1 lists the maximum allowed orbit errors for different baseline length, assuming that we want to constrain the introduced synchronization errors to 1 nanosecond.

Table 1

Tolerable orbit error for 1 ns time transfer

Baseline length [km]	Orbit error [m]
100	60
500	12
1000	6
5000	1

Assuming an accuracy of 10 m for the current broadcast orbits, this is sufficient for 1 ns time transfers over distances up to 600 km only. In the case of deteriorated accuracy these orbits will be degraded by a factor of 10 (~ 100 m), thus the according maximum distance is reduced to 60 km ! These considerations clearly ask for better orbit sources. Fortunately modern GPS phase processing techniques are able to determine satellite orbits in the 1 to 2 m range (Dong and Bock, 1989). To provide precise a posteriori orbit information for the civil user community on a regular basis, civil tracking networks will be established in the near future. Such networks are of growing importance especially during periods of deteriorated accuracy. Although precise orbits will be provided only a posteriori, this does not impose to many disadvantages to time transfer applications.

Time transfer experiments often suffer from uncertainties in the receiver antenna positions. To estimate the induced errors in the time difference dt we use again formula (2) where db now means the error in the receiver antenna location. The resulting 30 cm for a 1 ns time transfer are usually not attainable with conventional surveying techniques (except over very short distances). Therefore it is absolutely necessary to measure the network of involved stations directly with geodetic GPS receivers or to tie it at least (with GPS) to nearby geodetic reference stations like LASER or VLBI sites. This will guarantee that the influence of station coordinate uncertainties on the results are below 1 ns in any case.

For our application ionospheric effects become critical on rather short distances (100km). Consequently high precision time transfer asks for direct measurements of the ionospheric delays on each station, which implies the use of dual band receivers or at least instruments capable of measuring some P-code differences between the two carriers.

For GPS time transfer the C/A-code signal is almost exclusively used as observable. We already mentioned its modest accuracy, which is due to its large wavelength of ~300 m (also referred to as C/A-code chip-length). Another possibility is to use the so-called P-code of the GPS signal. Thanks to its 30 m chip-length modern receivers are able to measure this code with an observation noise of 1 to 5 m . Unfortunately P-code is under control of the DoD. Current policy however is to confine the unavailability of the P-code to short periods of time (except in the case of a military conflict). The most accurate observables of the GPS signal are the reconstructed carrier phases. A single observation noise of a few millimeters is achieved by most receivers. But carrier phases have a distinct disadvantage with respect to time transfer: their lack of time tags. However, as we will see in section 4, there are possibilities to combine the advantages of both, P-code and phase measurements, in order to achieve an optimal benefit for time transfer applications. There will be even an interesting application using carrier phases alone.

3. RECEIVER REQUIREMENTS

Our final goal is the intercomparison of the output signals of two (or more) station clocks. The task of the GPS receivers is to measure these reference signals with respect to signals coming from the satellites (see Figure 1). The actual observation of the time differences between the satellite and the station clock signals may also be performed outside the receivers (see Figure 2).

In both cases we have the following fundamental requirements:

- The receiver internal software clock related to the code measurements (pseudorange measurements) has to be synchronized with the internal hardware clock at the required accuracy level (< 1ns).
- The receiver internal software clock must be identical for code and phase measurements.
- The receiver internal hardware clock must either be synchronizable with the incoming reference signal on the required accuracy level or the produced output signal must be derived directly from the hardware clock.
- The oscillator used as receiver time base (receiver internal or external) must be of good short term quality (phase noise) so that carrier phase cycle slips may be repaired in the GPS data processing. In other words the influence of the oscillator on consecutive observations should not exceed 2 centimeters (or 60 ps).

The realization of these requirements is to a great extent receiver design dependent and can therefore only be met in collaboration with the manufacturer. They may in fact be the limiting factor in our approach.

Receiver requirements are much less stringent if we consider the case where we are not interested in the absolute time differences between station clocks but in their differential behavior relative to a starting epoch, in other words if we allow for an arbitrary constant in our time transfer measurements. Consequently we need, as we will show in section 4, only the phase observations and from the hardware point of view we need the following:

- The receiver internal oscillator must be either replacable by the time base of the station clock or its signal should be available as an output for phase comparision with

FIGURE 1

TIME TRANSFER USING RECEIVERS WITH REFERENCE INPUTS

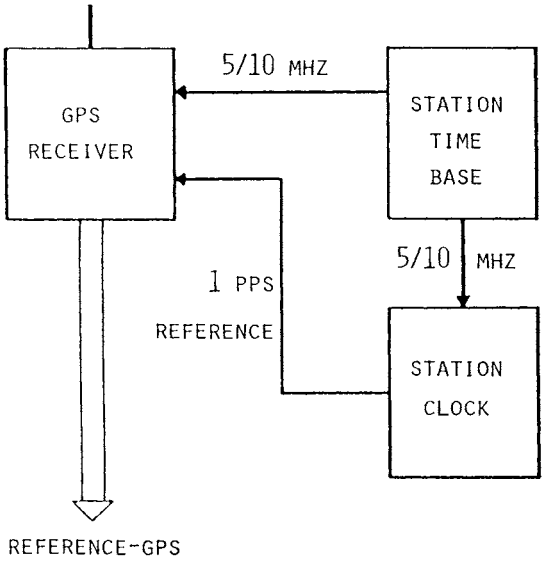
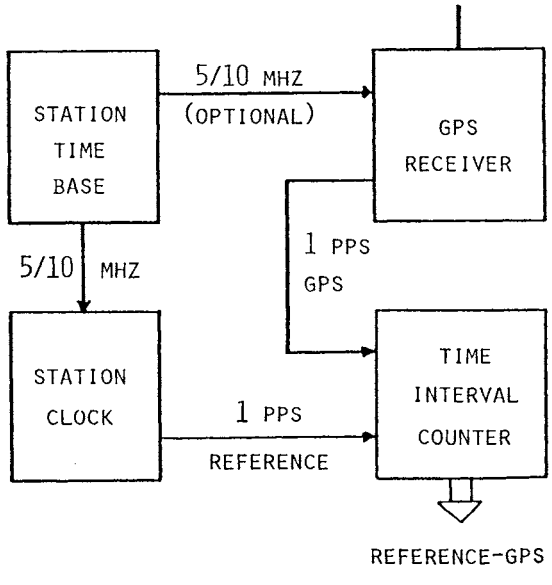


FIGURE 2

TIME TRANSFER USING RECEIVERS WITH GPS TIME OUTPUTS



the station clock oscillator.

Apart from these special requirements related to time transfer applications, we should ask for the following features:

- Dual band phase measurements
- Dual band P-code measurements
- 'Simultaneous' multiple satellite tracking
- In the case where the P-code is not available, the difference between the Y-codes on both carriers should be measured (for ionospheric correction purposes).

There is a further important issue which is often overseen: the mutual calibration of all involved receivers. Especially if different receiver types are involved internal delays may show up which cancel between instruments of the same type.

4. OPTIMUM TIME TRANSFER OVER REGIONAL AND CONTINENTAL DISTANCES

We discuss this topic using the GPS data set prepared by IAG SSG 1.104 (Special Study Group 1.104 of the International Association of Geodesy). In section 4.1 we briefly describe the data set, in section 4.2 we summarize the results obtained from an orbit determination step with GPS data from all over the continent (phase and code was processed), and we discuss typical results from regional/local GPS- network estimations (using the previously determined orbits). The purpose of this section is to demonstrate that

- a civil orbit determination service using GPS observations stemming from known VLBI- and LASER- sites (=fiducial point concept for orbit determination) does determine the relative geometry between GPS satellites and terrestrial sites in the region of the fiducial net with an accuracy of a few centimeters.
- using this concept it is no problem at all to include GPS- time stations into the fiducial network defined by the LASER and VLBI-sites.
- the entire geometry (orbits and station coordinates) may be assumed to be known for time transfer purposes on the 1 cm level if geodetic methods are used.

In section 4.3 we outline the optimum procedure for time transfer starting from the observation equations for code and phase observations.

4.1 The IAG SSG 1.104 Standard Data Set

The campaign and the M.I.T.-processing are described in (Dong and Bock, 1989). Here we just summarize the important aspects of this campaign for our subsequent discussion: A total of 19 station were equipped with GPS receivers from 1987 January 3 to 7. All five sessions lasted for approximately seven hours, a total of six satellites was observed. Dong and Bock made the distinction between a local net (9 sites), a regional net (5 sites), and a continental net (5 sites). The continental and regional sites are given in Figure 3, the regional and local sites in Figure 4 (reproduced from (Dong and Bock 1989)).

4.2 Processing the IAG SSG 1.104 Data Set in Berne

The following strategy was used to process the data:

(1) Orbit determination/improvement step:

The observations of the stations Platteville, Ovro, Algonquin, Westford, Churchill, and Austin were used to generate the best possible 5 day arcs for all satellites. The coordinates of Platteville, Ovro, Algonquin, and Westford are known from VLBI, they were kept fixed in the orbit determination. The coordinates from Austin and Churchill were estimated together with the orbit parameters (see (Beutler et al., 1990)).

(2) Analysis of the Regional and Local Net

The orbits generated in step (1) were used and left unchanged. The initial phase ambiguities were resolved using the Melbourne/ Wuebbena approach ((Melbourne, 1985), Wuebbena, 1985)) for the wide-lane, using the ionosphere-free linear combination for the remaining L1-ambiguities. All but 5 ambiguities could be resolved in the local and regional net, no attempt was made to resolve ambiguities in the continental net.

This bisection of processing seems appropriate for a region with an established regional civilian tracking/orbit determination service. Task (1) would then be accomplished by the

Figure 3

Continental and Regional Sites of IAG SSG 1.104 Data Set

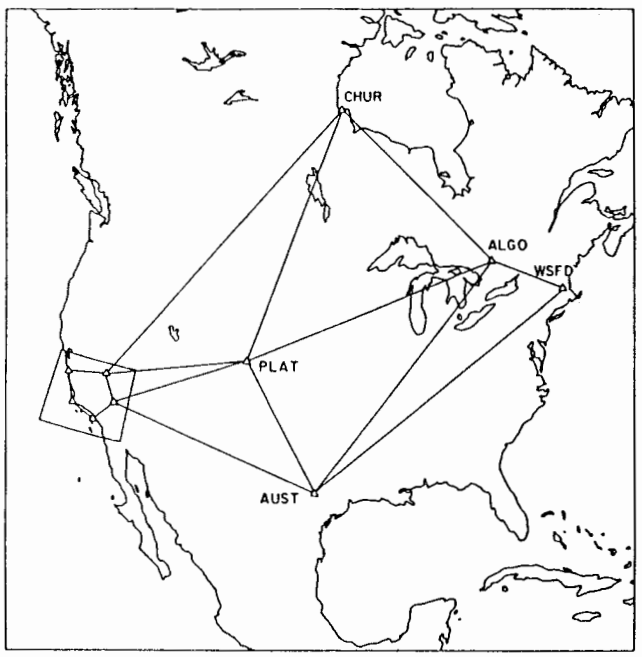
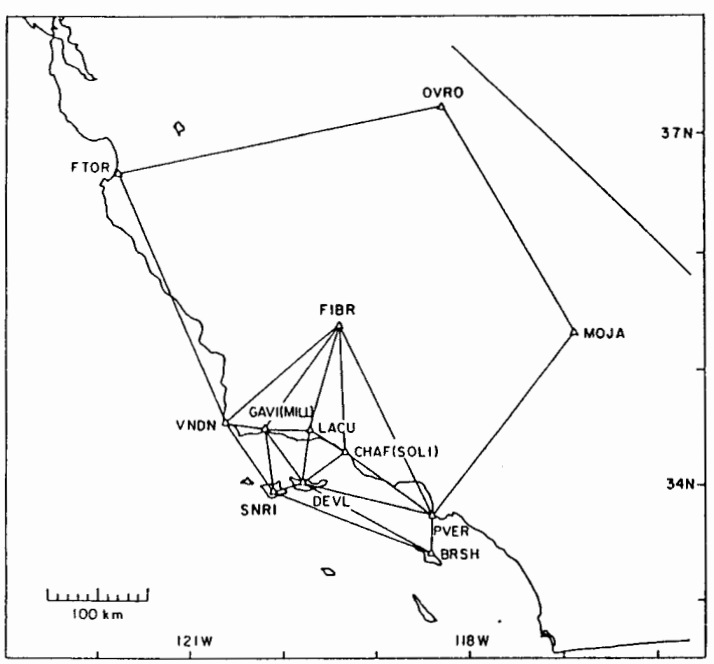


Figure 4

Regional and Local Sites of IAG SSG 1.104 Data Set



orbit determination center, only task (2) would be left to the (usually less experienced) consortium organizing the observation campaign. Unnecessary to say that such a splitting up of the work is time/money/energy - saving, if more than one local/regional campaign are organized simultaneously. It would be ideal for time transfer purposes.

It is not a trivial task to assess the quality of an estimated GPS orbit. Formal errors computed by the orbit determination program usually are much too optimistic, only the ratios between formal errors of different orbit parameters are reliable. It has become standard to judge the quality of orbits by using them to process (independent) GPS data stemming from the region of the fiducial network. If baseline repeatability is good, the orbits are said to be good. A valuable indicator is also the estimated rms of the single observation: If this value is close to the expected a priori value we may conclude that no systematic (orbit) errors remained in the analysis.

Sometimes it is even possible to process data from baselines which are accurately known from non-GPS sources (e.g. terrestrial, LASER, or VLBI measurements). In the IAG SSG 1.104 data set the Ovro - Mojave baseline (see Figure 4, length about 246 km) determined by VLBI was included to give an independent accuracy check: Baseline-length, latitude- longitude- differences were recoverd by our analysis roughly within the centimeter, the height difference between the two stations differed from the VLBI result by about 2 cm. This is a clue that the relative geometry of the entire network really is of the order of a few centimeters over distances up to, let us say, thousand kilometers.

Repeatability studies are important for the validation of results. The data set considered here is exemplary in the sense that five times essentially the same network was measured. This means that we can use session solutions to find a measure for the baseline (coordinates and lengths) repeatability. We followed the procedure outlined by (Dong and Bock, 1989) by defining latitude-, longitude-, height- and length- repeatability of a particular baseline as the rms error of the mean value of all (at most five) determinations of the considered quantity. As opposed to Dong and Bock we did not weight the individual solutions. The so defined repeatabilities may now be plotted as a function of baseline-length.

In Figures 5 and 6 we plot length- and height- repeatability as a function of baseline length. We can see the well known fact that GPS (like VLBI) measures baseline-lengths much more accurately than station heights.

From Figures 5 and 6 we draw the conclusion that it is no problem to include every site of interest for time transfer into a regional fiducial net with an accuracy of the order of one centimetre in all three coordinates as long as this site is not farther than, let

Figure 5

Repeatability of Baseline-length (seven hour sessions)

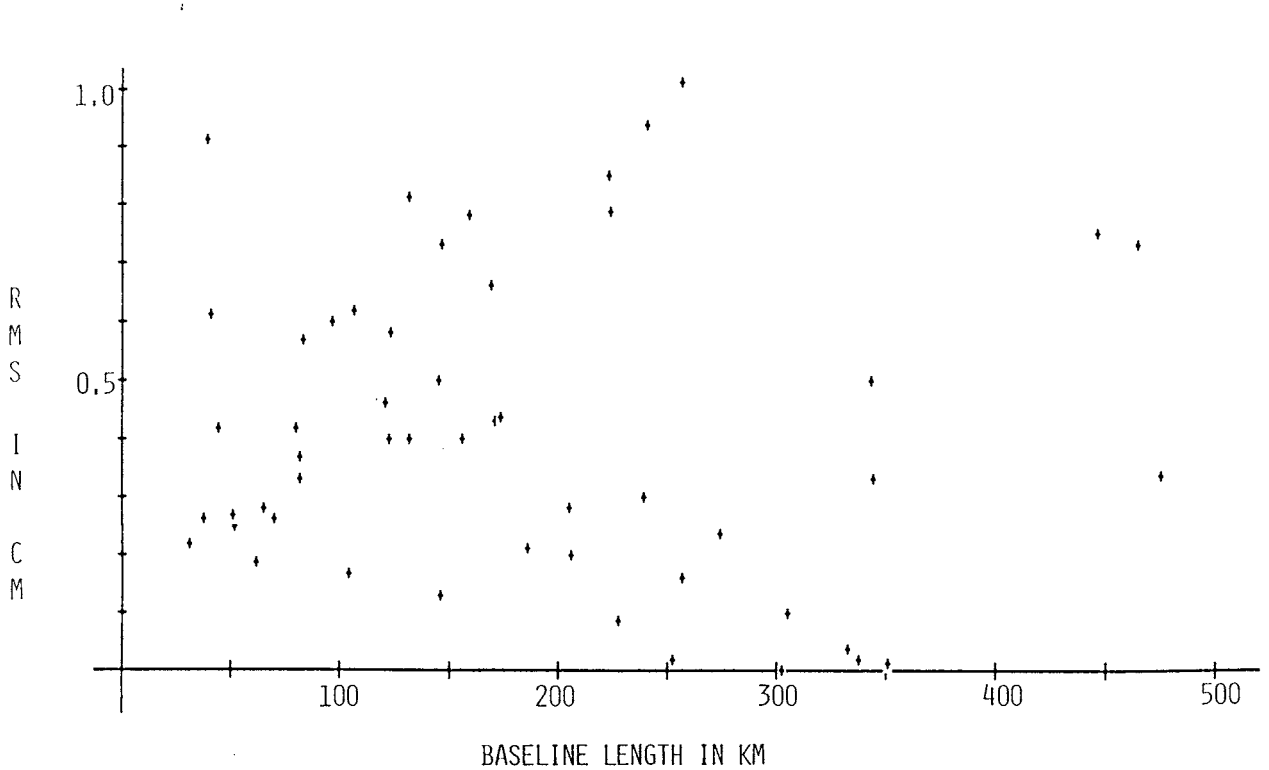
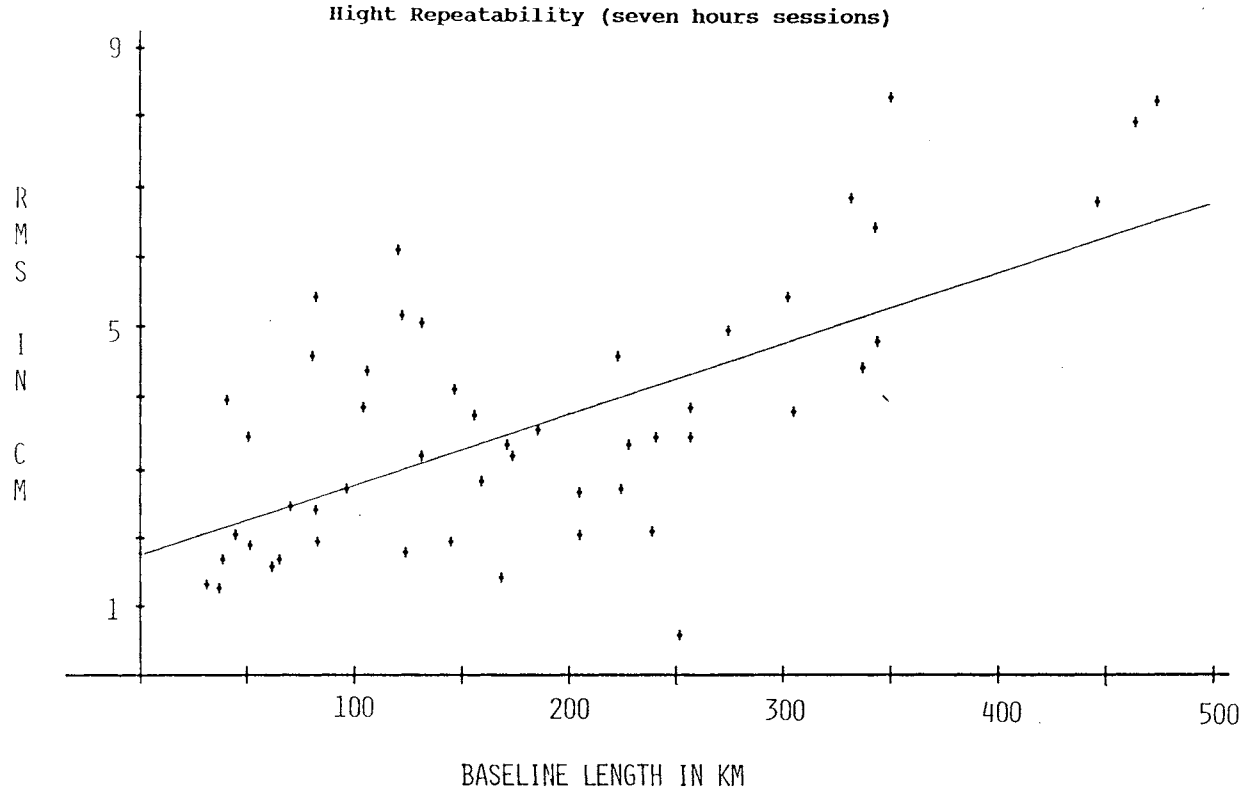


Figure 6



us say, 200 km away from the nearest reference site.

We now have to answer the question whether Figure 5 or 6 is relevant for time transfer problems. In principle it is clear that Figure 6 has to be used, since height is important to form the observable for our purpose. On the other hand we should not forget that a rotation of the entire network (including satellites and stations) from session to session is of no interest (the clock difference is invariant under (translation and) rotation). No attempt was made to eliminate possible rotations from Figure 6. Therefore we should answer the question how much of the effects seen in Figure 6 may be explained by such rotations. We cannot give a general answer at present, but we can give some hints indicating that probably most of the effect is due to these uninteresting rotations of the entire network: Table 2 contains the residuals of a pure translation between two daily solutions (January 5 and 6), Table 3 contains the same quantities for the same solutions for a transformation including translation and rotation. Clearly in this case the height residuals are greatly reduced. A similar behaviour is observed for all session solutions. We may therefore be quite optimistic and state that it is possible to model the geometry between stations and satellites on the centimeter-level in all three coordinates provided the methods of geodesy are used.

Table 2

Residuals of a 3 parameter HELMERT transformation between L_3
Solutions of days 5 and 6 of IAG SSG 1.104 data set
(3 Parameters = Translations)

NUM	NAME	RESIDUALS IN METERS		
		LAT	LONG	HEIGHT
3	BRUSH, CATALINAI	-0.0022	0.0024	0.0349
4	BUTTONWILLOW	0.0042	0.0068	-0.0120
7	DEVILS PEAK, SCI	0.0013	0.0038	0.0144
8	FORTORD NCMN1981	-0.0020	-0.0068	-0.0802
10	LA CUMBRE	-0.0016	-0.0074	-0.0049
13	OVRO 7114	0.0055	0.0085	-0.0203
14	PALOSVERDES 7680	0.0019	-0.0073	0.0546
16	SOLEDAD SRI	-0.0017	-0.0029	0.0020
17	SOLIMAR	-0.0001	0.0048	0.0357
18	VANDENBURG 7223	-0.0052	-0.0020	-0.0241

NUMBER OF PARAMETERS : 3
NUMBER OF COORDINATES : 30
RMS OF TRANSFORMATION : 0.0225

Table 3

Residuals of a 6 parameter HELMERT transformation between L_3
Solutions of days 5 and 6 of IAG SSG 1.104 data set
(6 parameters = translations and rotations)

NUM	NAME	RESIDUALS IN METERS		
		LAT	LONG	HEIGHT
3	BRUSH, CATALINAI	-0.0037	0.0048	-0.0162
4	BUTTONWILLOW	0.0037	0.0055	-0.0045
7	DEVILS PEAK, SCI	0.0012	0.0046	0.0048
8	FORTORD NCMN1981	0.0017	-0.0088	0.0012
10	LA CUMBRE	-0.0019	-0.0076	-0.0074
13	OVRO 7114	0.0045	0.0054	-0.0012
14	PALOSVERDES 7680	0.0004	-0.0056	0.0098
16	SOLEDAD SRI	-0.0012	-0.0018	-0.0021
17	SOLIMAR	-0.0007	0.0053	0.0215
18	VANDENBURG 7223	-0.0040	-0.0018	-0.0060

NUMBER OF PARAMETERS : 6
NUMBER OF COORDINATES : 30
RMS OF TRANSFORMATION : 0.0075

TRANSLATION IN X : -0.006 +- 0.002 M
TRANSLATION IN Y : -0.009 +- 0.002 M
TRANSLATION IN Z : 0.020 +- 0.002 M
ROTATION AROUND X-AXIS: 0 0 0.05 +- 0.01 "
ROTATION AROUND Y-AXIS: - 0 0 0.03 +- 0.00 "
ROTATION AROUND Z-AXIS: - 0 0 0.00 +- 0.00 "

4.3 Outline of Correct Time Transfer Algorithms

The single difference observation equations for code and phase and for carriers ℓ=1,2 may be written (somewhat simplified) as:

$$\begin{aligned}
&\text{Code:} \quad \Delta d_k + (c+\dot{d}_k)\Delta t_k + (\Delta\delta_{\ell k})_{ion} + (\Delta\delta_k)_{trop} - c'_{\ell k} = v_{c\ell k} \\
&\text{Phase:} \quad \Delta d_k + (c+\dot{d}_k)\Delta t_k + (\Delta\delta_{\ell k})_{ion} + (\Delta\delta_k)_{trop} + N_\ell \lambda_\ell - p'_{\ell k} = v_{p\ell k}
\end{aligned} \qquad (3)$$

$$\ell = 1,2;\ k = 1,2,\ldots,n^s_{obs}$$

Where: Δd_k: Slant range difference between satellite S and the two stations R_1, R_2 involved.

Δt_k: Synchronization errors between the two receiver clocks at observation time t_k.

c: Speed of light.

$\dot{d}_k$: Radial velocity of S with respect to R_1.

$(\Delta\delta_{\ell k})_{ion}$: (Differential) ionospheric range correction. It may be shown that:

$$(\Delta\delta_{2k})_{ion} = \frac{f_1^2}{f_2^2}(\Delta\delta_{1k})_{ion} \quad , \quad k=1,2,\ldots \qquad (4)$$

where f_i, i=1,2 are the carrier frequencies.

$(\Delta\delta_k)_{trop}$: (Differential) tropospheric range correction.

N_ℓ: Initial ambiguity parameter (in general unknown).

λ_ℓ: Wavelength of carrier ℓ.

$c'_{\ell k}, p'_{\ell k}$: Code- respectively phase-observations.

$v_{c\ell k}, v_{p\ell k}$: Residuals.

n^s_{obs}: Number of observations for satellite S.

There is one set of equations (3) for each satellite S observed. For geodetic work usually the so called double difference equation (the difference between two equations of type (3) for two different satellites S_1, S_2) is analyzed. The reason for this resides in the fact that the influence of the clock synchronization terms Δt_k is greatly reduced ($\sim 3\cdot 10^5$). For our purpose this would not make sense since our main interest is the series of clock terms Δt_k, k=1,2,... . We may assume however that in our case the geometry and tropospheric refraction is "perfectly" known. We may therefore rewrite eqns. (3) as:

$$\begin{aligned}
&\text{Clock:} \quad (c+\dot{d}_k)\Delta t_k + (\Delta\delta_{\ell k})_{ion} - \tilde{c}'_{\ell k} = v_{c\ell k} \\
&\text{Phase:} \quad (c+\dot{d}_k)\Delta t_k + (\Delta\delta_{\ell k})_{ion} + N_\ell \cdot \lambda_\ell - \tilde{p}'_{\ell k} = v_{p\ell k}
\end{aligned} \qquad (5)$$

$$\ell = 1,2\ ;\ k = 1,2,\ldots,n^s_{obs}$$

where: $\tilde{c}'_{\ell k} = c'_{\ell k} - \Delta d_k - (\Delta\delta_k)_{trop}$

$\tilde{p}'_{\ell k} = p'_{\ell k} - \Delta d_k - (\Delta\delta_k)_{trop}$

Taking into account eqns. (4) we have for satellite S

$$n_{p/s} = 2\cdot n^s_{obs} + 1 \qquad (6)$$

unknowns and

$$n_{\ell qn} = 4\cdot n^s_{obs} \qquad (7)$$

observation equations of types (5). Simultaneous observations of additional satellites will only introduce

$$n_{p/s} = n^s_{obs} + 1 \qquad (8)$$

new unknowns (the ambiguity parameter and the ionosphere parameters). Therefore we have enough information to solve for all the unknowns in eqns. (5). If we process all eqns. of type (5) for all simultaneously observed satellites in one adjustment run we obtain an over-determined linear system of equations which - of course - will be solved by a least squares algorithm. When forming the normal equation system one has to take into account that phase and code equations have a different observation noise (a few millimeters for the phase, a few meter for the code). Since the ratio $\sigma_{code}:\sigma_{phase}$ is known for each re-

ceiver type, proper weighting in the least squares solution is no problem. This procedure is in principle the only correct one from the mathematical point of view.

For demonstration purposes we will analyze the structure of eqns. (5) in more detail. As a result of our adjustment process we do not only obtain the unknown parameter values but also the full variance-covariance matrix associated with the parameters. This allows us to compute the rms errors of the parameters and of every linear combination. In the order of magnitude we obtain

$$\sigma(\Delta t_k) \approx 0.5 - 0.1 \text{ nsec} \tag{9}$$

and

$$\sigma(\Delta t_k - \Delta t_\ell) \approx 0.05 - 0.01 \text{ nsec} \qquad k \neq \ell \tag{10}$$

The reason that the "relative" synchronization error (with respect to an arbitrary epoch t_ℓ) is much better determined than the "absolute" synchronization error becomes obvious if we subtract the observation equation (5) **for the phase** for k=1 from the phase equations for k=2,3,... (see eqns. $(5)_2$): The ambiguity term is eliminated, only the clock terms $(\Delta t_k - \Delta t_\ell)$ and the ionosphere terms remain in the equations for carriers ℓ=1 and ℓ=2. It is now possible to form a special linear combination (the so called ionosphere free lienar combination) of the equations for ℓ=1 and ℓ=2) for each k=2,3... . What remains is an observation equation for each k=2,3,... containing only one unknown, namely $\Delta t_k - \Delta t_\ell$, which may be determined essentially with the accuracy of the phase observation. This means that even a dual-band C/A code receiver will allow us to compute the differences $\Delta t_k - \Delta t_\ell$ with the high accuracy indicated by eqn. (10). We would like to point out that it is not necessary to go through this somewhat tedious procedure of forming linear combinations of the original observation equations (5) - we only did that for illustration pruposes. In practice the observation equations will be formed starting from eqns. (5), we will solve for all the unknowns which figure there, and we will compute linear combinations of the unknowns and their mean errors using the well-known formulae of the adjustment algorithms.
We would like to ask one more question: Why do we expect (see eqn. (9)) that the absolute Δt_k are accurate on the sub-nanosecond level although the individual code measurements at time t_k are much less accurate: The answer is "code-smoothing": Using the phase equations to replace the Δt_k for k=2,3,... by $\Delta t_\ell - (\Delta t_k - \Delta t_\ell)$ in the code equations $(5)_1$ leads ultimately (after forming the ionosphere-free linear combinations of the code equations) to n^s_{obs} equations for the one and only unknown Δt_ℓ. Since this may be done for observation periods of hours and for all satellites observed simultaneously the high accuracy is explained. Again: We will actually not form these linear combinations: The information is in the variance-covariance matrix of the adjustment. It should be pointed out that this sub-nanosecond level time transfer is possible even with rather cheap quartz oscillators: No assumptions concerning the smoothness of clock was made. If good oscillators are involved (e.g. masers) it is possible to model the Δt_k by low degree polynomials, which would lead to even better results.

Let us conclude this section by two illustrative examples stemming from the IAG SSG 1.104 standard data set. In Figure 7 we plotted the ionosphere-free linear combination of phase single difference residuals $\tilde{p}'_{\ell k}$, ℓ=1,2 for k=1,2,... for the baseline La Cumbre - Chaffee (41 km) for the observation session of January 3 (for illustration purposes a fifth degree polynomial was subtracted from the residuals). Although usually 3-4 satellites were observed simultaneously we cannot see any satellite dependent differences in the residuals - which is exactly what we expect from eqns. (5) for phase observations (the measurement errors are of the order of $\leq$ 1 centimeter). The clock behaviour is as expected for the quartz oscillators of the TI-4100. Figure 7 also illustrates very nicely that the clock difference between any two epochs may be recovered very accurately. Figure 8 contains the same information as Figure 7 but using code instead of phase observations. Obviously the pattern is the same as in Figure 7 but the noise is for orders of magnitude larger. It should be very clear now that we can use the curve of Figure 7 to remove the relative clock error (relative to a certain epoch t_ℓ) in Figure 8, and that then all measurements of Figure 8 can be used to solve for one unknown Δt_ℓ.

Figure 7

Phase Single Difference Residuals for the Baseline La Cumbre - Chaffee (41 km)

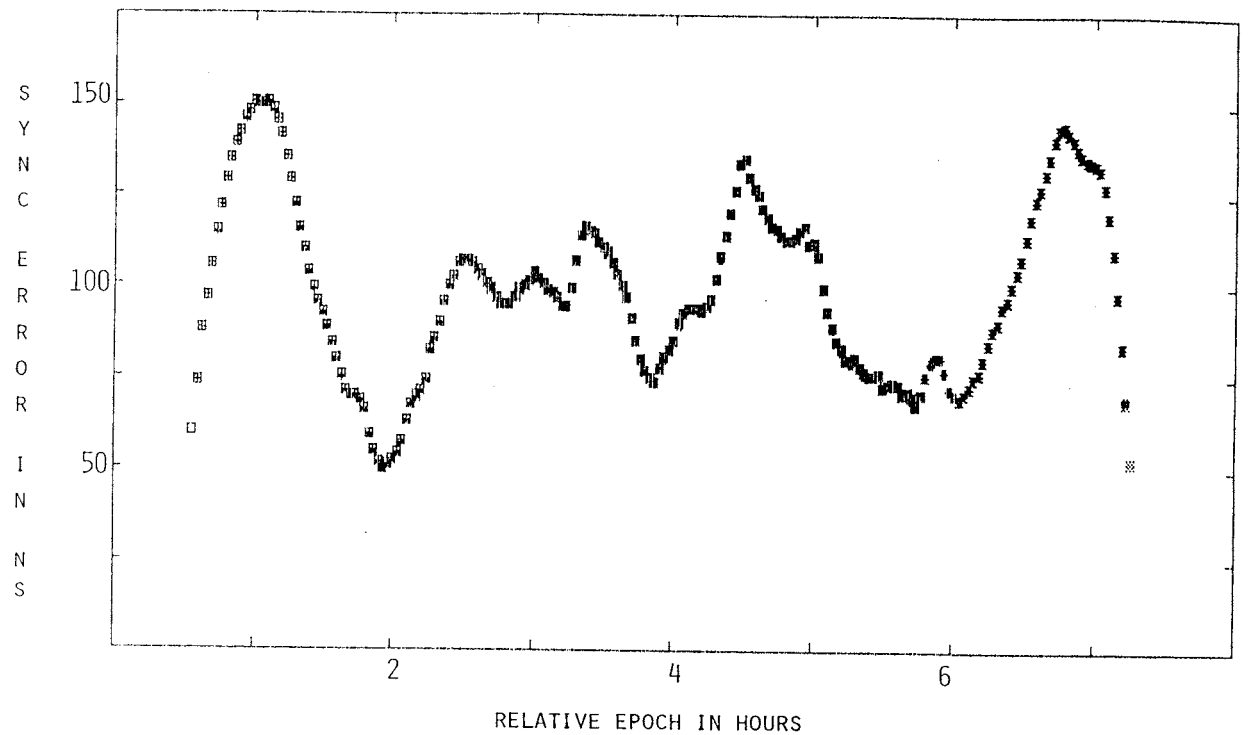
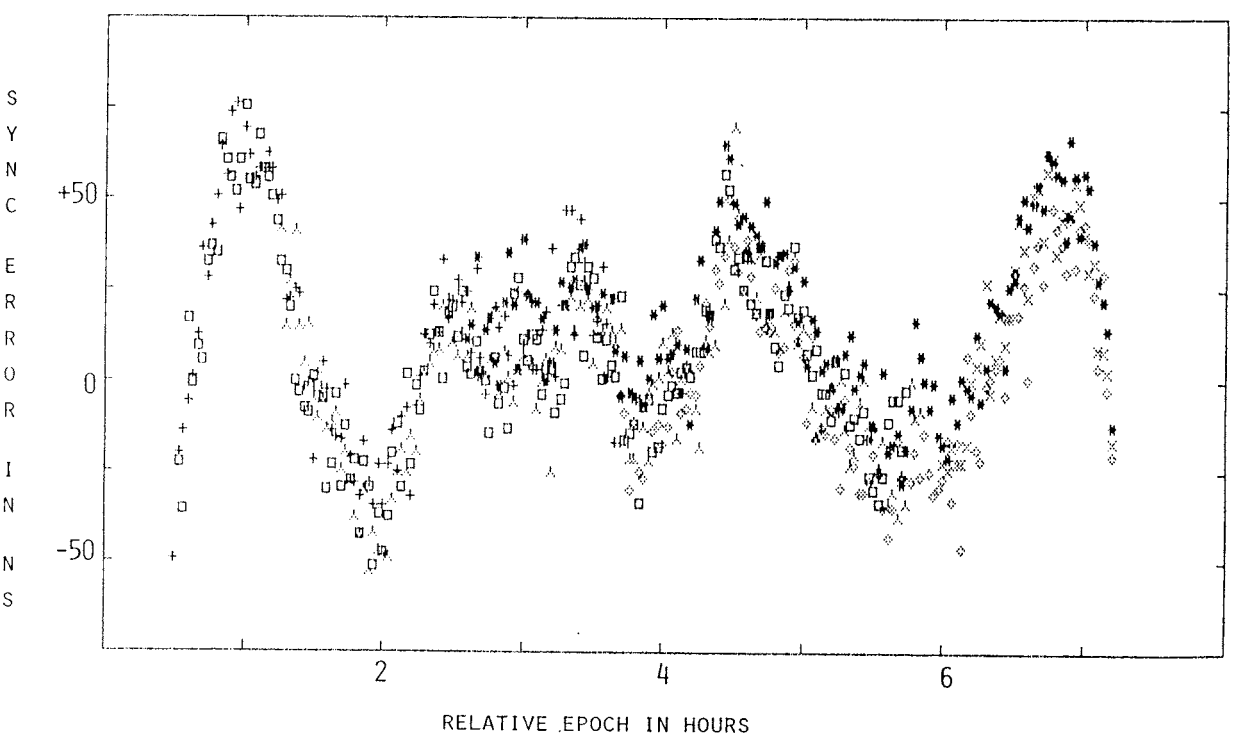


Figure 8

Code Single Difference Residuals for the Baseline La Cumbre - Chaffee (41 km)



5. CONCLUSIONS

The use of geodetic approaches in time transfer applications makes it possible to overcome the basic limitations of the traditional C/A-code common view method:

- The relative geometry between stations and satellites can be modelled on the centimeter level in all three coordinates in networks of continental size.
- Phase observations allow to reduce the intrinsic noise of the code observations by a factor of 100.
- With correct algorithms all simultaneously observed satellites may be processed together.

Using dual band code and phase observations stemming from geodetic receivers and provided the mentioned receiver requirements are met it as feasible to establish 1ns time transfer links on continental baselines.

REFERENCES

Bauersima, I. (1983). "Radiointerferometrische Satellitenbeobachtungen, (GPS II). "Mitteilungen der Satellitenbeobachtungsstation Zimmerwald, Nr. 10, Astronomisches Institut der Universität Bern, Schweiz.

Beutler, G., S. Fankhauser, W. Gurtner, M. Rothacher, U. Wild (1990). "Considerations concerning GPS software development: Structure, models, algorithms. "Proceedings of 1989 Luxembourg GPS workshop, in press.

Dong, D., Y. Bock (1989). "Global Positioning System Network Analysis With Phase Ambiguity Resolution Applied to Crustal Deformation Studies in California. "Journal of Geophysical Research, Vol. 94, No. B4, pp. 3949-3966.

Melbourne, W. G. (1985). "The Case for Ranging in GPS Based Geodetic Systems." First International Symposium on Preceise Positioning with the Global Positioning System, pp. 373-386, Rockville, MD, U.S.A.

Remondi B. W. (1989). "GPS Broadcast Orbits Versus Precise Orbits: A Comparision Study." CSTG GPS Bulletin, Vol. 2, No. 6.

Wuebbena, G. (1985). "Software Developments for Geodetic Positioning with GPS Using TI-4100 Code and Carrier Measurements.", First International Symposium on Precise Positioning with the Global Positioning System, pp. 403-412, Rockville, MD, U.S.A.

Sensors and Actuators

SA-01

Chairman: B. Neubig

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session SA-01 Number 1

CHEMICAL ETCHING OF MONOLITHIC QUARTZ VIBRATING BEAM SENSORS

F. Deyzac, D. Janiaud, S. Muller

Office National d'Etudes et de Recherches Aérospatiales,
BP 72, 92322 Châtillon Cedex, FRANCE.

SUMMARY

Differential vibrating beam sensors take advantage of the frequency variations of flexure vibrating beams, under pressure or tensile axial strengths.

In the present state of the art, such sensors are obtained by assembling the component units of the active part (in the case of accelerometers: vibrating beams, proof-mass, articulation of proof-mass, fixed part).

The purpose of this study, performed at ONERA under DRET contract, is to suppress the assemblage by proposing a monolithic quartz active part.

The expected benefit consists in starting a new generation of vibrating beam sensors that should allow a manufacturing cost reduction while providing performances depending only on crystal qualities and not on assemblage unstabilities.

A first monolithic differential accelerometer geometry is proposed, featuring reduced dimensions (over-all volume of the accelerometer-case is 1 cm^3).

INTRODUCTION

For several years, now, many research activities are devoted to the miniaturization and to the mass production of the sensors.

But it seems that no real work has been performed to these ends for the Vibrating Beam Accelerometer (V.B.A.).

In its classical configuration (see Figure 1), the V.B.A. has a complex geometry which requires specific tools and procedures that are incompatible with a low cost production.

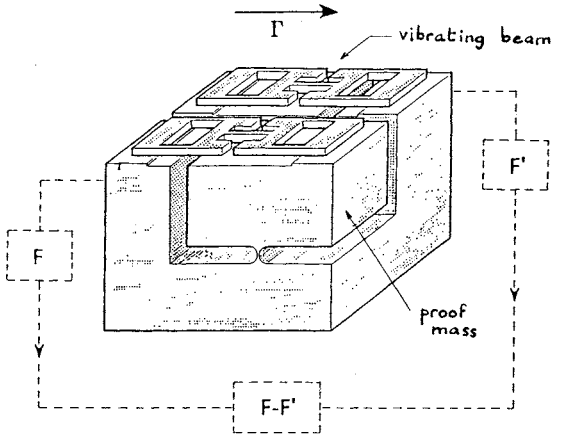


Figure 1 - Typical configuration of a Vibrating Beam Accelerometer.

$$F = F_0 + K_1\Gamma + K_2\Gamma^2 + A(T°)$$
$$F' = F_0 - K_1\Gamma + K_2\Gamma^2 + A(T°)$$
$$F\text{-}F' = 2K_1\Gamma$$

This is why the concept of the V.B.A. has been reviewed leading to the design of a monolithic configuration which can be obtained by chemical etching already used in watch crystal industry.

MONOLITHIC ACCELEROMETER CONFIGURATION

The geometry of the Monolithic V.B.A. has been determined in order to satisfy the following requirements:

- Operating frequencies: $F \sim F' \sim 50$ kHz
- Scale factor $\sim$ 50 Hz/G
- Range $\sim$ - 50 G to + 50 G
- Vibration resonance $\geqslant$ 2000 Hz

These nominal values appear to be optimal for the vibrating beam concept.

The interest of the configuration presented in figure 2 has been first demonstrated through an analytical approach.

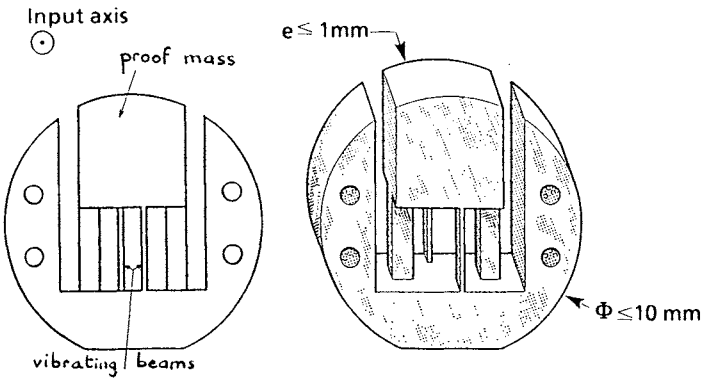


Figure 2 - Monolithic Quartz Vibrating Beam Accelerometer.

Then, this monolithic geometry has been assessed by finite element software (SAMCEF*, see figure 3), and by the experimentation of a large scaled model (scale: × 100).

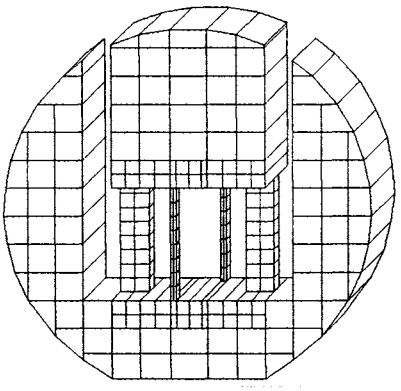


Figure 3 - Finite element model.

MACHINING PROCESS OF THE MONOLITHIC ACCELEROMETER

The optimization of the machining procedure is on going, the chemical etching being especially studied.

The present results are very encouraging, as it is shown in figure 4.

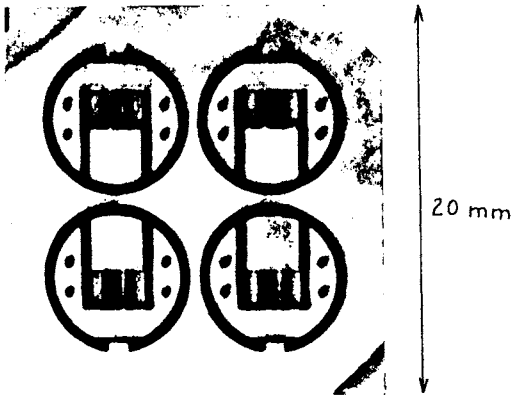


Figure 4 - Four quartz accelerometers wafer.

A quartz wafer with four accelerometers is presented; these four accelerometers have been simultaneously machined.

CONCLUSION

The proposed concept of the Monolithic Quartz Vibrating Beam Accelerometer has been especially designed for manufacturing through a chemical etching process. Its potential cost-performance ratio appears very promising and more generally such an approach can be applied to other types of quartz sensors.

*Système d'Analyse des Milieux Continus par Eléments Finis.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session SA-01 Number 2

RESONANT MAGNETIC-FIELD SENSOR

E. DONZIER - O. LEFORT

SEXTANT Avionique VALENCE - FRANCE

S. SPIRKOVITCH - F. BAILLIEU

ECOLE SUPERIEURE D'INGENIEURS EN ELECTROTECHNIQUE ET ELECTRONIQUE

NOISY-LE-GRAND - FRANCE

Summary:

The detector described below is a mechanical oscillator comprising a monocrystalline silicon vibrating beam driven on its resonance frequency. The Laplace force acting on a current coil in the presence of a magnetic field excites the micro-resonator. The mechanical signal is transduced by a bridge of piezoresistive gauges diffused in the silicon. By using the quality factor of the structure, low fields (less than 10 nT) can be measured on a band of some 10 Hz.

I Introduction

The possibility of measuring the magnetic field offers a variety of applications, such as:

- determining the heading of a vehicle (navigation for land vehicles, boats and aircraft),
- detecting disturbances in the earth's magnetic field (surveillance of an area),
- remote measurement of current.

A number of magnetometers are currently available on the market. The main ones are as follows:

- Fluxgate sensors which use the properties of ferro-magnetic materials. Although they offer highresolution,they are fairly cumbersome and expensive in terms of manpower.
- Integrated sensors of the Hall-effect type, magneto-resistors, magneto-diodes or magneto-transistors. Of fairly poor resolution, the parameters of these field sensitive electronic devices are influenced significantly by temperature changes. Consequently they are not suitable for measuring continuous magnetic fields at very low frequencies.

The quality of a sensor may be judged in terms of the following features:

- performance,
- reliability,
- cost,
- volume,
- power consumption,
- compatibility with the processing system,
- production capacity.

By adapting common microelectronics techniques to the production of silicon-based mechanical micro-structures, it is possible to satisfy the development needs for sensors. (1)

Production techniques are collective and consequently cheap and easy to reproduce.

Photolithographic techniques for shape reproduction attain high performance levels.

The possibilities for thin film deposition are almost unlimited, both in terms of choice of materials and monitoring of deposition parameters.
Silicon machining techniques (chemical, isotropic or anisotropic) combined with assembly techniques (electrostatic bonding of glass/silicon, silicon/silicon bonding) can be used to produce complex structures (beams, bridges, membranes, etc.).

Monocrystalline silicon possesses excellent qualities of elasticity (a very low dislocation rate and consequently excellent resistance to fatigue, reproducible elasticity modules and a low thermal expansion constant).

Monocrystalline silicon is highly piezoresistive. For example, a deformation detector can be connected to the test body simply by diffusing impurities to define a gauge bridge.

Thanks to the compatibility between micro-machining techniques and the manufacture of electronic equipment (MOS and bipolar transistors) local information processing is possible. The advent of integrated signal processing sensors, or "intelligent" sensors, should herald significant improvements in products (increased compensation possibilities, data compression, etc.).

SEXTANT Avionics is active in a wide variety of fields, ranging from space and aeronautics to industrial automation, and attaches particular importance to the quality of its sensors. In terms of innovation, the company devotes considerable energy to the study and development of integrated silicon and quartz sensors (magnetometers, accelerometers, pressure sensors, integrated microphones, etc.).

The study of the vibrating beam micro-magnetometer described below was carried out within the framework of collaboration between SEXTANT and the micro-electronics research laboratory at the E.S.I.E.E. (Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique).

The project is supported by official French bodies, and in particular:

- A.N.R.T. (Agence Nationale pour la Recherche et la Technologie = National research and technology agency)
- M.R.T. (Ministère de la Recherche et de la Technologie = Ministry of research and technology)

II Principles of the magnetic field sensor

The detector is a mechanical resonator comprising a plate fixed along one side and carrying a conduction band. A sinusoidal current i passes through the band at a frequency equal to one of the natural frequency modes of the structure.

As the beam is immersed in the external magnetic field, it will oscillate at its resonance frequency under the effect of the

Laplace force $F = \vec{i}\,\vec{l} \wedge \vec{B}$ (cf. Fig. 1).

The signal is mechanically amplified using a quality factor Q, the value of which may reach several thousand, depending on the size of the plate and the damping conditions. In addition, the static field to be measured is modulated using a high frequency signal. This conversion of the static signal into a frequency eliminates the drift problems associated with the measurement of continuous signals (temperature sensitive offset, 1/f noise, etc.). The mechanical signal is converted into an electric signal using a bridge of piezoresistive gauges. The beam movement induces stress in the material: fibre elongation or compression.

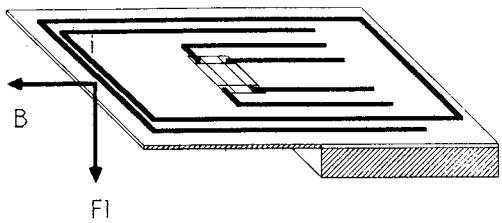


FIGURE 1 : PRINCIPLE

This deformation is transformed into a modification of the resistivity of the conductors integrated in the beam: this is the piezoresistive effect.
For a relative variation in length Δl/l, the relative variation in resistance is $\frac{\Delta R}{R} = G\frac{\Delta l}{l}$, where G is called the piezoresistive gauge factor. Of the order of 1 to 2 for common conductors (conductivity variation mainly related to a geometric effect), G may reach 150 in the case of monocrystalline silicon in particular crystalline directions (modification of the electronic properties of the semi-conductor).

Four gauges are connected in a Wheatstone bridge set-up (in order to avoid power supply drifts) and are polarized by a dc voltage. The gauges are placed at the point of maximum stress and in such a way that the forces are opposing on adjacent branches of the bridge and in the same sense on symmetrical branches .

The beam oscillations will lead to an imbalance in the bridge proportional to the field component along the vibration axis of the plate. The basic principle of the detector is reduced to a succession of elementary physical phenomena, described in fig. 2.

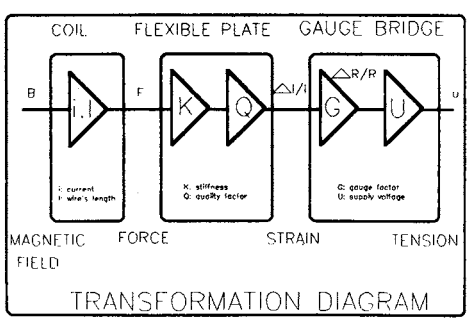


FIGURE 2 :

III Dimensional design of the structure

Once the basic principle has been adopted, the dimensions of the structure need to be defined to satisfy the tender specifications.

The mode frequency of a plate fixed along one side is given by :

$$f_{ij} = \frac{\lambda_{ij}^2}{2\pi l^2}\sqrt{\frac{E.e^2}{12\,\gamma\left(1-\nu^2\right)}}$$

where:

l: length of the plate
L: width of the plate
e: thickness of the plate
i: number of mode half wavelengths on the horizontal axis
j: number of mode half wavelengths on the vertical axis
E: Young's modulus
γ: density
ν: Poisson's ratio

λ_{ij} depends on the length-to-width ratio of the plate. Some numerical results are given in the table below.(2).

l/L	λ_{ij}^2 and ij mode		
	1	2	3
0,4	(11) 3,511	(12) 4,786	(13) 8,115
1	(11) 3,492	(12) 8,525	(21) 21,43
1,5	(11) 3,477	(12) 11,68	(21) 21,56

Assuming a plate of length l, and width 2 l (so as to optimize the area occupied by the conductor), excited on its first flexion mode, the bridge output signal will be expressed as:

$$u = \frac{U \times G \times j \times l \times B}{2 \times \pi\sqrt{E\rho}}\,\Delta f$$

where:

U: bridge supply voltage
G: gauge factor
j: current density in the conductor
B: magnetic field
E: Young's modulus of the material
ρ: material density
Δf: bandwidth

The bandwidth determines the sensitivity of the appliance. This does not depend on the plate area but only on its thickness. The result is favourable to the structural integration.

The detector bandwidth is determined by beam damping. In air at normal atmospheric pressure this damping would be determined by the viscosity of the air.

If the height of the air layer below the vibrating plate is small compared to the plate characteristics, the pneumatic damping coefficient of the air will take the following form:

$\lambda = \frac{3\eta l^4}{h^3}$ and $\Delta f = \frac{\lambda}{m}$

where:

η : air viscosity coefficient
m: weight of plate = 2 r x l^2 x e
h: height of the air layer

If the structure is placed in a vacuum enclosure, damping will be determined by visco-elasticity phenomena (irreversible heat conduction process).

The quality factor Q of the structure is related to the pass-band by the conventional relation:

$Q = \frac{f_o}{\Delta f}$ where f_o is the resonance frequency

These damping parameters are difficult to control. As the bandwidth is generally specified in the sensor specifications, a method of controlling the bandwidth using a servo-loop system will be described in the pages below.

The detector resolution is limited by the noise that may occur in the sensor transfer function.

The beam movement as a result of Laplace's force must first of all be greater than the thermal fluctuations of the beam position. This equivalent noise force is introduced by the dissipating elements of the system (damping terms)

It may be expressed in simple terms by:

$$F_{noise} = \sqrt{4kT\lambda\Delta}$$

where:

k: Boltzmann's constant
T: temperature

This should be compared with F_{signal} = i.l.B.

The signal/mechanical noise ratio is thus obtained, depending on the physical parameters of the structure.

$$S/B = \frac{i.B}{\sqrt{12\eta kT\Delta f}} h^{3/2}$$

This ratio depends only on the air-gap, in other words the base height.

A further limiting factor arises from the electrical noise of the gauge. Only the noise within a Δf band around the resonance frequency affects the output signal. In view of the orders of magnitude involved(Δf = some 10 Hz, f_o = some 10^4 Hz), this noise may be assimilated to the thermal noise of the resistors $.Ub = \sqrt{4kTR\Delta f}$ where R: gauge resistance.

From the foregoing, it may be deduce that the optimisation of detector resolution depends on a minimisation of the beam thickness.

The critical parameters to be checked for a particular specification can be determined from the laws of dimensional change of the various detector characteristics

The power consumption of the unit is proportional to the surface area of the structure (if bridge consumption is low compared to Joule effect energy dissipation in the conductor).

However, technological considerations will tend to impose certain limits on the choice of structure dimensions:

* machining method
* maximum current density in the conductor (electromigration phenomena)
* area of the strain gauges
* encapsulation techniques
* etc.

IV Experiments

The structure described above may be obtained easily using a silicon wafer. The gauge bridge is set into the front face of the wafer. The conductor and bridge connections are obtained by vacuum aluminium evaporation. They are separated from the silicon by a layer of silicon dioxide.

The conductor is made up of n wound turns in order to supply the detector with low current.

This gives: $i = \frac{i_{total}}{n}$

The flexible plate is then obtained by thinning down the wafer from the rear face. An engraving mask aligned on the gauges is used to define the base. This is then fused onto a glass substrate or simply bonded to a ceramic package.

A photograph of the detector and its spectrum show the first three resonance modes.

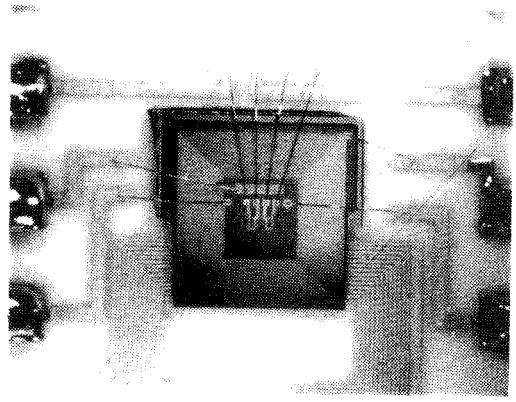


FIGURE 3 : ONE PLATE DETECTOR

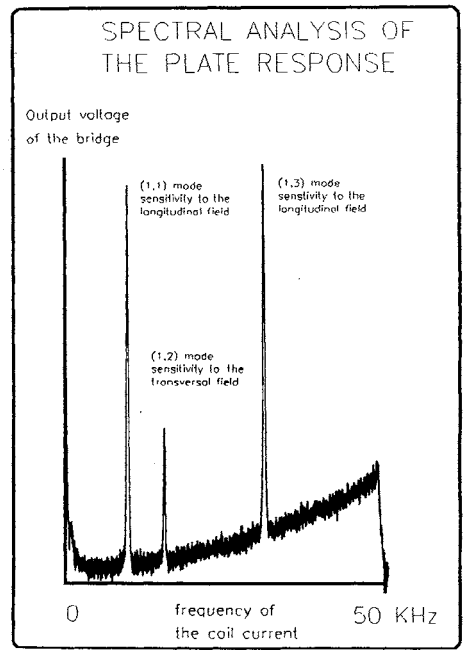


FIGURE 4 : PLATE RESPONSE

In order to reduce disturbances produced by possible acceleration, a structure with two symmetrical beams was produced. Subjected to a field effect, the beams vibrate in phase opposition. The signal caused by accelerations is rejected by determining the difference between the signals from the two detection bridges (cf. photo).

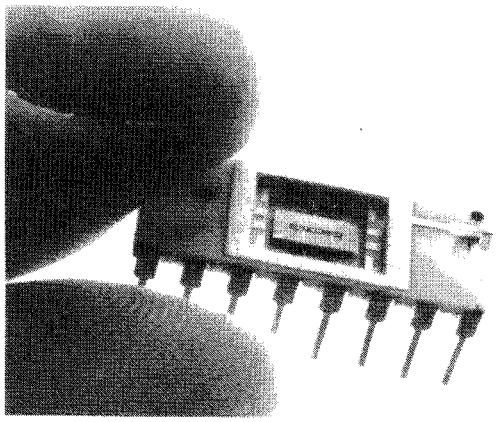


FIGURE 5 : TWO PLATES DETECTOR

V Application

By placing a conductor close to the vibrating beam (figure 3) and maintaining a static current in this conductor, it is possible to cancel the external field by detecting the zero position of the mechanical system. This type of servo-loop control enables a linear response to be obtained which is not affected by possible variations in detection module sensitivity. It also allows effective monitoring of the bandwidth.

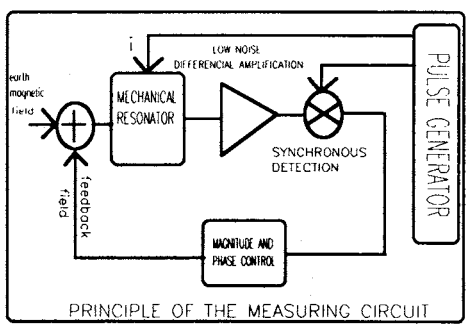


FIGURE 6

The principle of the measurement system associated with the detector is described in fig. 6. The system includes:

- a shaping stage for the excitation signals applied to the vibrating beam and generated using an external reference oscillator,
- a low noise amplification stage for the signals produced by the piezoresistive gauge bridge,
- a synchronous demodulation stage,
- a corrector circuit for checking the gain and control loop phase

The system response to a particular field is shown in the following figure 7.

The sensitivities obtained experimentally are compatible with the models established in section III.

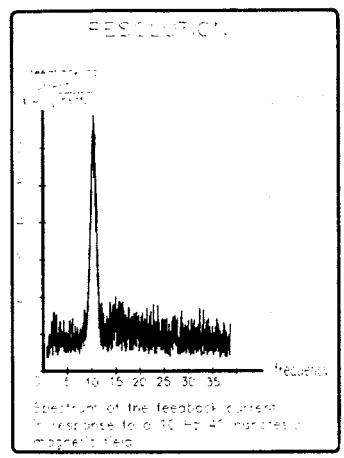


FIGURE 7

VI - Conclusions

The study has proved the feasibility of this type of device. A good level of correspondence has been established between the model built by theoretical analysis and the performance actually obtained.

The technology used to produce the detectors offers the numerous advantages that have contributed to the development of microelectronics.

Another type of magnetometer is under study using a capacitor detector. This should allow even higher levels of integration for detectors (50 µm wide beam, with amplification electronics on the same chip).

References :

(1) : Silicon as a mechanical material K.E. PETERSEN - Proc. of the IEEE, vol. 7 n° 5 May 1982
(2) : Formulas for natural frequency and mode shape
R.D. BLEVINS
R.E. KRIEGER Publishing Company

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session SA-01 Number 3

DOUBLE ACTION PIEZOELECTRIC MOTORS

M. BRISSAUD, J.L. DION *, L. EYRAUD

Laboratoire de Génie Électrique et Ferroélectricité
Institut National des Sciences Appliquées de Lyon, France

* Laboratoire de Recherche en Électronique Industrielle
Université du Québec à Trois-Rivières, Québec

SUMMARY

The operating principles of a few types of piezoelectric motors are presented. Some common features of these are relatively small size, simplicity, relatively high torques and forces, low velocities from zero. The computed possible traction force is reported for a multilayered stator. Some types of motors hold promises for eventual miniaturization and integration in small mechanisms.

RÉSUMÉ

Nous présentons les principes de fonctionnement de quelques types de moteurs piézoélectriques. Leurs traits communs: compacité, simplicité, couples et forces relativement élevés, faibles vitesse à partir de zéro. Nous avons calculé la force de traction possible pour un stator multicouches. On peut espérer l'intégration éventuelle de certains moteurs dans des microméca-nismes.

INTRODUCTION

Various types of piezoelectric motors have been studied in the last few years which appear quite promising for applications where small dimensions, variable low velocities and relatively high forces and torques are required [1-4]. These new motors use the high frequency elliptical vibrations of piezoelectric ceramics assemblied as a stator, simply pressed on a movable part or a rotor. The movement is produced by friction forces exerted by the stator on the movable part. They are powered by at least two controlled current sources at a few tens of kilohertz, with a phase difference such that an elliptical mechanical vibration is produced. They can be made as linear or rotating motors. The effective friction force is exerted between the stator and the rotor, usually on a limited surface, during part of a cycle. Consequenty, the published efficiencies are less than 40% for most designs. They are used to produce a rectilinear or rotational motion with controllable velocities between 0 and about one meter per second. Some operating principles of existing motors are presented and the theoretical model we use to design an efficient system and predict its performances. We computed possible traction forces well over 1 kgf/cm² for a unit being constructed.

SOME OPERATING PRINCIPLES

1- The piezoelectric effect

All piezoelectric motors make use of the *piezoelectric effect* in certain dielectric materials, discovered in quartz by the Curie brothers in the 1880's. It can be briefly described as in fig. 1. A piece of piezoelectric ceramic polarised (P) in direction 3 (fig. 1a), submitted to voltage V produces a longitudinal stress T_3 (N/m²) along 3:

$$T_3 = c_{33}S_3 - h_{33}D_3 \qquad (1)$$

If polarization P is perpendicular to the electric field as in fig. 1b, a shear stress T_5 is produced:

$$T_5 = c_{55}S_5 - h_{15}D_1 \qquad (2)$$

where c_{33} and c_{55} are elastic constants, S_3 and S_5 are longitudinal and shear strains (relative deformations), h_{33} and h_{15} are piezoelectric constants, D_3 and D_1 are the electric induction (or displacement) fields, equal to the electrical charge densities on the electrodes.

2- "Vibrating finger" motors

As shown in fig. 2 [1], the end of a metal rod or "finger" can describe a *Lissajous figure* when excited resonantly by a set of piezoceramics, separately inducing longitudinal and transverse vibrations. If the end of the rod is then pushed on the periphery of a wheel, this wheel can then rotate: piezoelectric energy is converted to rotational energy and work. In this particular case, a torque of 3 cm N and an efficiency of 5% is reported [1] at a few rev/s. We can also imagine the structure shown in fig. 3, where a set of piezoceramics 2 perpendicular to set 1 is excited by voltage V_2 in quadrature with V_1. Due to the Poisson coefficient of the ceramics, flexion stresses are induced which produce an elliptical movement at the tip of the rod. A force perpendicular to the rod can therefore be exerted on a movable body. One apparented device is described in reference [3].

3- Progressive wave motors

The largest number of piezoelectric motors now produced are apparently of this type, one embodiement of which is shown in fig. 4 [4-7]. Stator in shape of a ring 43 is composed of piezoceramic(s) 25 with a metallic dented ring on top, and a rotor 5 above. Excitation of the piezoelectric elements by two electrical generators at frequency f with a phase difference of about 90° can give rise to a progressive wave around the ring, producing at each point in contact with the rotor an elliptical movement in the opposite direction as shown in fig. 5. Therefore, the rotor can move in the direction opposite to that of wave propagation. How this wave is produced can be explained using fig. 6a showing a piezoelectric ring with uniform polarization perpendicular to the page. On top, there are two sets of independent electrodes of such length that the perimeter can contain an integer number of them. A bottom electrode not shown is uniform. If the set at the right is connected as shown to

generators $+\cos(2\pi ft)$ and $-\cos(2\pi ft)$, the piezoelectric stresses will be of opposite signs in adjacent zones, and can give rise to a resonance as shown in fig. 6b representing the perimeter of the ring. This is a *stationary wave* which displacement can be described by:

$$Y_1 = 2A \cos(2\pi ft) \cos(2\pi s/\lambda)$$

or,

$$Y_1 = A \cos(2\pi ft - 2\pi s/\lambda) + A \cos(2\pi ft + 2\pi s/\lambda) \qquad (3)$$

which represents the superposition of two progressive waves travelling in opposite directions. The length of each top electrode is one half wavelength; s is the position measured along the perimeter, and A is the amplitude of vibration. The other set of electrodes at the left (fig. 6a) is displaced by $\lambda/4$ along the perimeter, and excited with a phase difference of $\pi/2$ rd (90°). They produce another stationary wave which superposes to the first one:

$$Y_2 = 2A \cos(2\pi ft - \pi/2) \cos(2\pi s/\lambda + \pi/2)$$

$$Y_2 = -A \cos(2\pi ft - 2\pi s/\lambda) + A \cos(2\pi ft + 2\pi s/\lambda) \qquad (4)$$

So, because of the displacement and the phase difference, the first travelling wave cancels out, leaving the one propagating in the negative direction of s, as the sum of Y_1 and Y_2.

The resulting motors can be small enough to be used in autofocusing cameras. The torque produced can be some 0.1 m N at 1 rev./s, with a 20% efficiency [7]. Some are apparently more powerful [5].

4- <<Push and twist>> motors

Some motors use the double action of a set of extensional piezoceramics and a set of shearing piezoceramics, as the one shown in fig. 7 [8,9]. Appropriate excitation produces simultaneous axial and torsional vibrations of the stator, with a phase difference of 90°, such that each point point in contact with the rotor has the same elliptical circumferential movement, moving the rotor by friction forces. This type of motor gives the highest torque, about 8 cm kgf at less than 1 rev./s, or 3 cm kgf at 2 rev./s.

5- Elliptically oscillating module

We became interested in the last type of motors and decided to work on a "building block" using a pile of rectangular piezoceramics of two types: thickness and shear vibrators. These are combined with other solid layers (brass, *laiton*, etc, fig. 8) to achieve simultaneous resonance in the two vibrations. For the analysis, we used a composite transducer model based on the analogy with transmission line theory, assuming that compressional and shear waves propagating perpendicular to the electrodes were nearly plane waves, non interacting with each other.

Then, we assumed in the first place that the rotor (not shown) was a propagating medium in rigid contact with the stator surface. It is then possible to compute the normal and shear stresses at the common surface for given applied applied voltages. This provides a value of the minimum contact force which must be maintained to hold the stator and rotor together. We can also compute the electrical impedance of each part of the unit, which makes possible to evaluate the required power.

RESULTS

We designed a unit as in fig. 8 to operate at a frequency of less than 100 kHz. It is composed of 8 shear vibrating (2.02 mm thick each), and 4 thickness vibrating piezoceramics (4.27 mm thick each), with brass layers of 0.1 and 4 mm. The moving part (on the lower side of fig. 8) is made of aluminium and the contact surface is 5.2 cm^2. Fig. 9 shows the acoustic shear and longitudinal velocities at this surface for $V_1 = V_2 = 0.5$ volt. We see that velocity maximums almost coincide at two possible operating frequencies: 87.5 and 188.5 kHz. We chose the first frequency. Setting the open-circuit voltages V_1 and V_2 of the generators to an amplitude of 10 volts, we compute a compression/traction oscillating force of 159 N (16.2 kgf) and a shear force of 86.5 N (8.83 kgf) at the aluminium surface. Therefore, we may conclude that if the applied normal force is less than 159 N, the aluminium part should begin to slip. The power dissipated in the compression and shear piezoceramics is then respectively 1.75 W, and 1.5 W. Setting this force to only 40 N (4.1 kgf) and assuming an effective friction coefficient of 0.25, we see that the resulting traction force could be at least 1 kgf. In such a unit, the applied voltage and traction force can easily be at least four times higher. The traction velocity should then be around 10 cm/s.

CONCLUSION

We have seen that piezoelectric motors are interesting new devices which can develop appreciable forces and power at low velocities. Their operating frequency is normally inversely proportional to their size. If a device such as the last one should be reduced to a length of 4 mm, it should be excited with currents ten times smaller, at about 900 kHz. In any case, the requirements on the surface finish of moving parts would be much more stringent to allow controlled slip. The combination with integrated circuit techniques should eventually allow quite appreciable miniaturization.

REFERENCES

[1] M. FLEISCHER, D. STEIN, H. MEIXNER, New type of piezoelectric motor, *IEEE Trans. Ultrasonics, Ferroelectrics, Freq. Control*, vol. 36, No. 6 (1989), 614-619.

[2] Ultrasonic Motors now in Vogue, *J. of Electronic Engineering* (Japan), vol. 26, no. 265 (jan. 1989), 84-85.

[3] M. KUROSAWA, K. NAKAMURA, T. OKAMOTO et S. UEHA, An ultrasonic motor using bending vibrations of a short cylinder, *IEEE Trans. Ultrasonics, Ferroel. Freq. Control*, v. 36, no 5 (1989), 517-521.

[4] T. TAKANO, Y. TOMIKAWA, T. OGASAWARA, S. SUGAWARA et M. KONNO, Ultrasonic motors using piezoelectric ceramic multi-mode vibrators, *Proc. 8th Symp. on Ultrasonics Electronics, Tokyo 1987, Jap. J. Appl. Phys.*, v. 27 (1988), suppl. 27-1, 192-194.

[5] O. KAWASAKI, K. TAKEDA, T. NISHIKURA (Matsushita El. Ind. Co.), *Ultrasonic motor*, European Patent No. 0 315 933 A2 (1988).

[6] T. SASHITA (Shinsei Kogyo K.K.), *Motor for utilizing ultrasonic wave oscillation*, Jap. Patent No. 61-277384 (A) (1986).

[7] *Centre Technique de l'Industrie Horlogère*, note No 90.2002: Analyse du moteur piézoélectrique à ultrasons Shinsei Corp. modèle USR-45, janvier 1990.

[8] K. NAKAMURA, M. KUROSAWA et S. UEHA, Characteristics of a hybrid transducer type ultrasonic motor, *13rd Internat. Congress on Acoustics*, Belgrade, Yugoslavia, 1989.

[9] O. MYOGA (NEC Corp.), *Ultrasonic motor*, Jap. Patent No. 1-238477 (A) (1989).

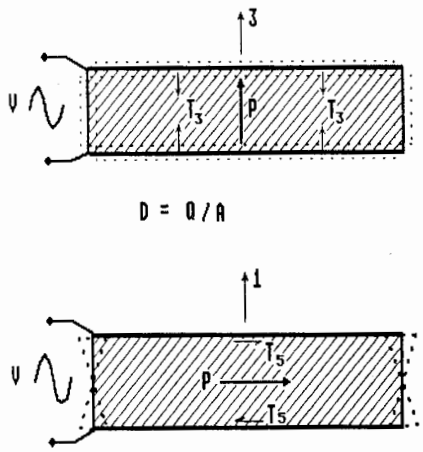


Figure 1 The piezoelectric effect

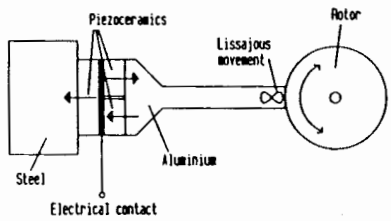


Figure 2 "Vibrating finger motors". End acting.

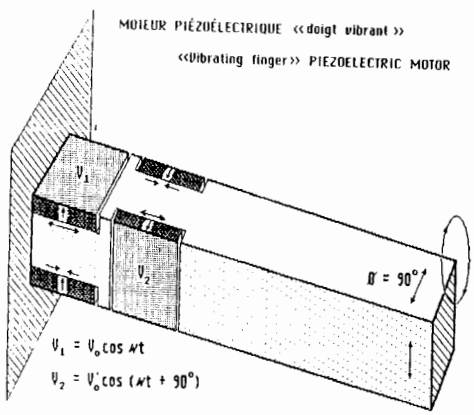


Figure 3 Vibrating rod motor. Side acting.

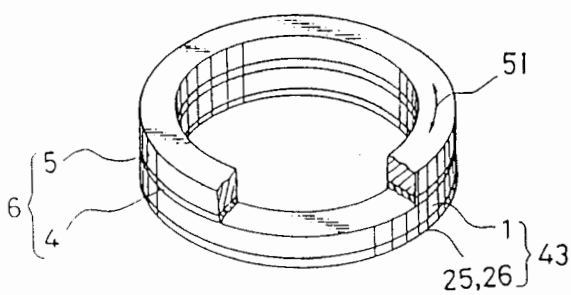


Figure 4 Progressive wave motor

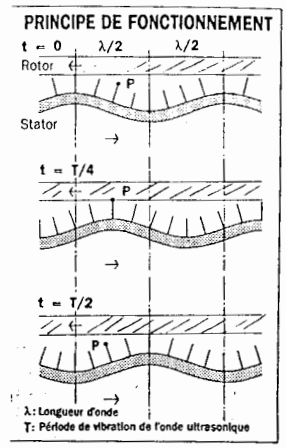
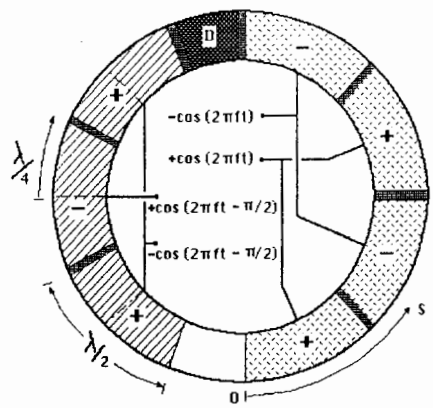


Figure 5 Action of the teeth of a progressive wave motor.



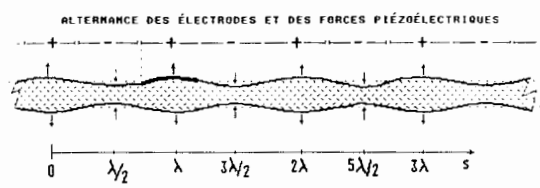


Figure 6
(a) Electrodes on a progressive wave motor.
(b) Perimeter of the rotor, showing the action of piezoelectric stresses under the electrodes.

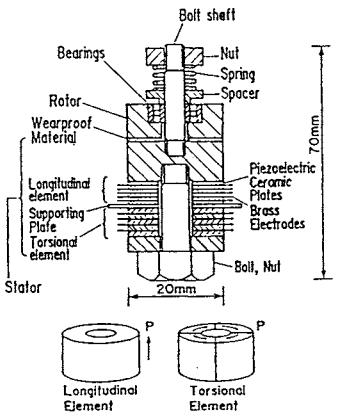


Figure 7 "Push and twist motor" (Pousse-tourne)

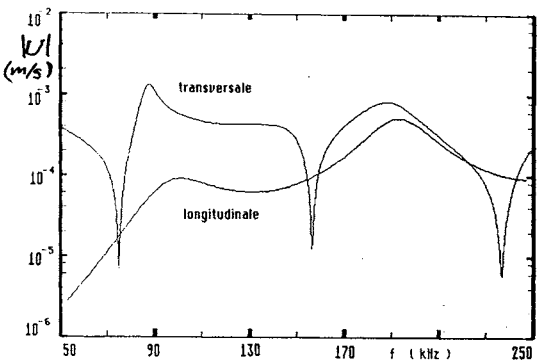
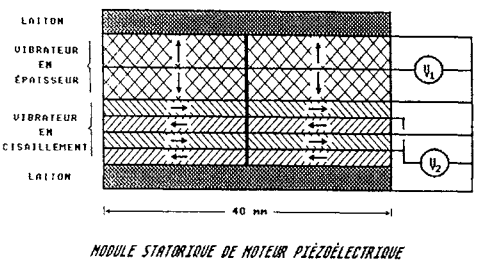


Figure 9 Computed shear and longitudinal acoustical velocities at the moving part interface.



MODULE STATORIQUE DE MOTEUR PIÉZOÉLECTRIQUE

Figure 8 Elliptically oscillating module.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session SA-01 Number 4

NEW BULK WAVE QUARTZ ACCELEROMETER

R. BOURQUIN, B. DULMET, R. BESSON

ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Route de Gray - La Bouloie - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

High stability bulk waves quartz resonators are operated as force-frequency transducers in a new accelerometer design. The implied physical mechanism is non linearity of stress versus strain in quartz crystal. The range of measurement of here presented devices might reach 50 G. A precise caracterization has yet occured in the +/- G range. A third order fit of the frequency output exhibits a residual of a few 10^{-4}. Thermal sensitivities of bias and scale factor are also reported together with time stability. Provisions are made to insure the monoaxiality of the accelerometer and its low thermal sensitivity by taking advantage of an original structure.

I - INTRODUCTION

Frequency output accelerometers are mostly designed from dual piezoelectrically driven tuning forks (VBA : vibrating beam accelerometers) (1). Their frequency output sensitivity arises from a change in boundary conditions at their ends. In devices presented here, the transducers consist of quartz plates vibrating in high frequency shear modes. The frequency shift of those resonators under the effect of acceleration results from a small change in the corresponding wave propagation celerity, as a consequence of the applied static stresses inside the crystal. The relative frequency shift is comparatively smaller than VBAs one, while this may be compensated by the better time stability of shear modes, which largely outperforms flexure modes one. In addition, using plate shaped resonators instead of beams greatly simplifies machining and assembly processes. Thermal compensation is achieved through a differential mounting which also improves monoaxiality.

II - A DIFFERENTIAL FORCE-FREQUENCY TRANSDUCER

The transducer is a thin and parallel quartz plate including two vibrating areas. Vibrations are energy-trapped in those areas by means of electrodes mass loading. First we consider that the rectangular plate is held at both ends. If a force is exerted in the central region, while oriented along the length of the rectangle, it becomes obvious that both halves of the plate (the vibrating areas) are subjected to plane compression stresses of opposite signs. The effective output signal is given by the resonators frequencies difference. f_1 and f_2 keep very near together but the difference does not vanish when the force is not present, in such a way that $f_1 - f_2$ never vanishes in the whole operating range of the transducer. The frequency perturbation of a resonator under the effect of any static stresses is accurately predicted by the formula (3) :

$$\frac{\Delta f}{f_0} = \frac{1}{2f_0^2} \frac{\int_{V_0} \hat{C}_{ijkl}\, u_{l,k}\, u_{j,i}\, dV_0}{\int_{V_0} \rho_0\, u_i\, u_i\, dV_0} \qquad (1)$$

where ρ_0 denotes the mass density of the crystal, whereas u_j denotes a component of the dynamic displacement, with $u_{j,i} \equiv \partial u_j / \partial x_i$ indicating its partial derivative with respect to x_i coordinate. The $\hat{C}_{ijkl}$ coefficients (dimension of elastic stiffnesses) indicate the effective elastic coefficients variations arising from $T_{i\,j}$ static stresses :

$$\hat{C}_{ijkl} = \delta_{jl}\, T_{ik} + C_{jisk}\, w_{l,s} - C_{ipkl}\, w_{j,p} + C_{ijklmn}\, w_{m,n} \qquad (2)$$

where :

$$T_{ij} = C_{ijkl}\, w_{l,k}$$

$$S_{ij} = \frac{1}{2}\,(w_{i,m}\, w_{j,m} - \delta_{ij}) = s_{ijkl}\, T_{kl}$$

In these relations w_m is a component of the static displacement, S_{ij} is the associated strain and C_{ijkl} is a usual, second-order, elastic modulus, whereas C_{ijklmn} stands for a non linear, third order, elastic modulus.

III - SHAPE OF RESONATORS

We only retained two possibilities among various shapes for the plate :

- either it includes two circular resonators, isolated by bridges (cf. Fig. 1)
- or its active region is rectangle shaped (cf. Fig. 2)

Properly using the above-mentionned formula (1), we can predict and compare the sensitivities of each kind of structure, assuming a resonator is submitted to a force F which is oriented along the sensitive axis of the device. In the case of AT cut plates the sensitive axis of which is truly oriented along cristallographic X axis, we get the following expression for the frequency shift, evaluated at the plate center :

$$\sigma_c = \frac{\Delta f}{f} = \frac{F}{\pi D e C_{66}}\,[(3s_{11} - s_{13})(C_{166} - 2C_{66}) + (3s_{12} - s_{23})C_{266} + (3s_{13} - s_{33}) + 3s_{14}C_{466}] \qquad (3)$$

This is for a circular resonator (diameter D, thickness e) **(4)** and assuming the static stresses distribution is isotropic in the disk submitted to a diameter compression **(5)**.

If we look at a rectangular resonator, of width L, we get the following expression :

$$\sigma_R = \frac{F}{2LeC_{66}}\,[s_{11}(C_{166} + 2C_{66}) + s_{12}C_{266} + s_{13}C_{366} + s_{14}C_{466}] \qquad (4)$$

Numeric calculations show that a disk structure is about 1.7 times more sensitive than a rectangular one, if D = L. Nevertheless, it should be pointed out that such a comparison is somewhat idealistic since practical realizations of disks exhibit bridges of finite width, thus reducing their sensitivity.

Although it is less sensitive, the rectangular shape offers some advantages :

- easier machining of the slots,
- better resistance to normal forces,
- less sensitivity to misplacement of the vibrating areas.

The latter feature deserves a special comment : as a matter of fact, vibration amplitude distribution in trapped energy resonators depend on the shape, the thickness and the mass density of the electrodes. But, if for any reason, the major surfaces of the plate present a small lack of parallelism, this results in a misplacement of the vibrating area from electrodes precise center. Because the static stresses are largely more homogeneous in the rectangular structure than in the disk, the former structure is far less sensitive to such a misplacement, thus improving efficiency of differential effect.

IV - STRUCTURE ORIENTATION AROUND THE CRYSTALLINE PLATE NORMAL

To discuss the optimal orientation of the rectangle around the plate normal, we have to consider frequency sensitivity to transverse forces normal to and in the plate plane, while perpendicular to the sensitive axis that can be drawn between the centers of the vibrating areas. When the sensitive axis is truly cristallographic X axis, the transverse force sensitivities are nearly zero in both cases. When the sensitive axis is chosen along AT cut's Z' axis, a transverse sensitivity ten percents of the compressional main sensitivity should be expected for a transverse force normal to the plate plane. In this configuration, the transverse sensitivity almost equals zero for an in plane transverse force.

A leading rule for the conception of here presented differential accelerometers consists of ensuring the compensation of any residual transverse sensitivities with help of differential effect. A detailed analysis showed serious reasons to prefer X oriented structure in the frame relative to AT-cut. Fig. 3 is a schematic diagram of a X oriented structure, submitted to transverse forces respectively acting along x_2 (fig. 3a) and x_3 (fig. 3b) axes.

Let $w(-x_1, x_2, x_3)$ and $w_{i,j}(-x_1, x_2, x_3)$ be the displacement and deformation gradient fields in the left part of the differential structure, whereas $w^*(x_1, x_2, x_3)$ and $w^*_{i,j}(x_1, x_2, x_3)$ are the corresponding quantities in the right part. All those quantities are related by :

$$w^*_i = w_i \text{ and } w_{i,j} = w^*_{i,j}$$

with the following exceptions :

$$w^*_1 = -w^*_1 \; ; w_{1,2} = -w^*_{1,2} \; ; w^*_{2,1} = -w^*_{2,1} \; ; w_{1,3} = -w^*_{1,3} \; ; w_{3,1} = -w^*_{3,1} \qquad (5)$$

If we consider a Z' oriented structure (fig. 4), the non equal terms from left to right part are :

$$w^*_3 = -w^*_3 \; ; w_{1,3} = -w^*_{1,3} \; ; w^*_{3,1} = -w^*_{3,1} \; ; w_{2,3} = -w^*_{2,3} \; ; w^*_{3,2} = -w^*_{3,2} \qquad (6)$$

Thus, transverse sensitivities behavior of the differential mounting can be known from (1) and (2). Table I summarize obtained results.

Main sensitivity axis	X	Z'
X sensitivities	add	compensate
Y' sensitivities	compensate	add
Z' sensitivities	compensate	add

V - ACTUAL DEVICES DESCRIPTION

The present accelerometer design has gained benefits from several realizations, some of which used disk resonators and unguided seismic masses. In its actual state, our device uses guided seismic masses while vibrating areas are located in a rectangular zone which is defined between two rectilign slots machined in a circular AT cut quartz plate. The length of the rectangle is along X axis (see fig. 5).

Both resonators operate at 40 MHz, 5th overtone. The vibrating plate thickness is near 200 microns. The cut angle is chosen to minimize the total frequency temperature variation in a - 20, + 80° C range.

A small central region of the differential plate is bonded to a rigid support, thus achieving a firm solidarization to the external holder. Two cylindrical seismic masses are bonded near the circumference of the disk, on each side. The guidance system consists of two rectangular thin plates whose plane is normal to the sensitive axis and the holder plane. One end of those plates is bonded to the holder whereas the other end is bonded to the lower seismic mass. To avoid most of differential thermal stresses arising between materials of different thermal expansions, the whole system, seismic masses + guiding thin plates + rigid support is made of quartz, with the same crystalline orientation as the sensitive plate. The device is put under vacuum in a cover whose volume does not exceed 10 cubic centimeters **(6)**.

This device architecture features several benefits :

- Planeity of the very thin plate regarding to its lateral dimensions is granted by its bonding to rigid seismic masses on the whole circumference.
- Quartz crystalline structures are known to be very stable over large periods of time.
- The machining of the primary parts of the system is relatively simple.
- Assembly process is simple.

The driving electronic subsystem uses surface mounted devices, directly delivers the frequency output $f_1 - f_2$, while keeping the power consumption under 100 mW.

The total seismic masses amount yet to about 15 g.

The guiding plates specifications do not reduce main sensitivity more than one percent of the scale factor.

In these conditions, the breaking acceleration is about 600 G, if we admit an extensional breakings stress nearing $10^4 N/cm^2$ in quartz.

Numerical computations using third order elastic coefficients values of Ref **(7)** predicted a theoretical sensitivity about 118 Hz/G, with help of Eq(14). This result could be improved taking into account the amplitude distribution of trapped energy vibrations **(8)**.

Planeity and parallelism of major surfaces of the plate were especially taken care of, the defects being kept under 0,1

micron. The frequency difference between both resonators is then lower than 20 kHz. It can be adjusted in further achievements by a more precise control of electrode thickness. As consequence, the operating range of the actual device is ± 50 G.

VI - OUTPUT LINEARITY

The output signal F_a is analyzed through a third order polynomial fit, as follows :

$$F_a = K_0 + K_1(A_s + K_2A_s^2 + K_3A_s^3 + K_pA_p + K_{sp}A_sA_p + \delta) \qquad (7)$$

A_S and A_p are acceleration components respectively along the sensitive axis and a normal axis. δ is the model residual. Fig. 6 shows a plot of this model residual during a linearity test by rotating the device in the gravitational field ($-1G \leq As \leq 1G$). The parameters are obtained from a least square treatment of the model given by (7). Typical results have been obtained from a limited series of 6 devices, and are shown at Table II.

Scale factor K_1	114 Hz/G
2nd order coefficient K_2	- 30 $\mu G/G^2$
3nd order coefficient K_3	200 $\mu G/G^3$
Coupling axes coefficient K_{sp}	< 200 $\mu G/G^3$
Standard error δ	150 μG

Table II

VII - THERMAL SENSITIVITIES

Since both resonating areas are not absolutely identical and never will be as the identicity criterion becomes more and more severe, there is left a small residual thermal sensitivity of the output signal. Some tests performed on a score of raw sensitive plates indicate a mean scaled thermal sensitivity less than 0,3 mG/°C over the operating range - 20, + 80°C. The way the plate is mounted in the structure may unfortunately aggrave this thermal sensitivity if not enough attention is paid to the elimination of static stresses arising from temperature variation, and especially under the action of unmatched thermal expansions.

The actual thermal sensitivity of the scale factor appears to lay between 160 and 800 µG/°C. A theoretical prediction of this effect, which is not differentially compensated, is not possible as long as the temperature coefficients of third order elastic constants of quartz are not known.

The thermal sensitivity of the main axis orientation error K_p is lower than 20 µrad/°C.

VIII - TIME STABILITY

Parameters stability has been obtained from a series of measurements at the four cardinal points (up, down, right, left). The dispersion of the results indicated a r.m.s. error about 150 µG for the output bias K_0 and 25 ppm for the main sensitivity K_1. Some long term measurements of the output bias K_0 indicated a stability better than 5 mG/month.

IX - ACKNOWLEDGEMENTS

The authors wish to thank D.R.E.T. for its support during the complete development of the accelerometer.

(1) A.M. GOGIC, R.B. PETERS, "Description and test methods for a frequency output accelerometer". Proc. of the National Technical Meeting, Vol 84, p 108-117, San Diego (1984).

(2) W.C. ALBERT, "Force sensing using quartz crystal flexure resonators", Proc. 38th An. Freq. Contr. Symps., p 233 - 239, Philadelphia (1984).

(3) H.F. TIERSTEN - Perturbation Theory for linear Electroelastic Equations for small fields superposed on a bias. J.A.S.A., 64, n 3, p 832-837, (1978).

(4) D. JANIAUD, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", thèse D.I., Besançon, (1978).

(5) S.P. TIMOSHENKO, "Théorie de l'Elasticité", Béranger, Paris, (1935).

(6) R. BESSON, R. BOURQUIN, B. DULMET, P. MAITRE, "Accéléromètre différentiel à résonateurs piézoélectriques", French patent n° 9000252

(7) THURSTON, ANDREATCH, MC SKIMIN, "Third order elastic coefficients of quartz", J. Appl. Phys., 37, N1, p 267-275, (1966).

(8) B. DULMET, R. BOURQUIN, "Sensibilité d'un résonateur à quartz plan-convexe à la compression diamétrale", Proceedings of the first European Time and Frequency Forum, p 99-103, Besançon, (1987).

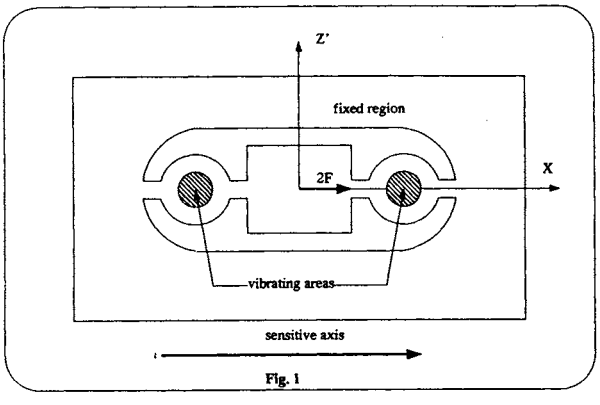


Fig. 1

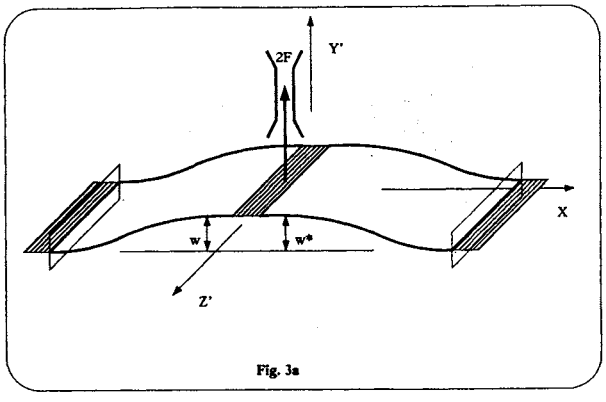


Fig. 3a

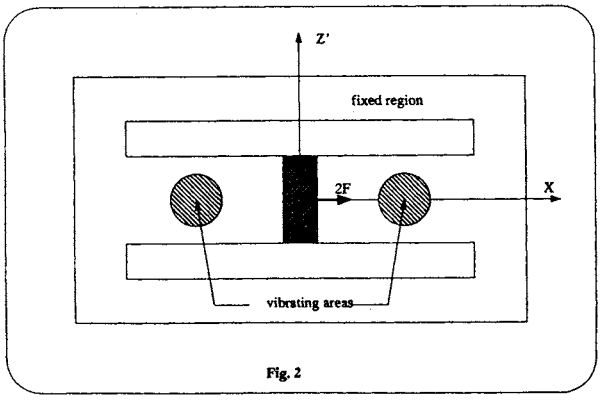


Fig. 2

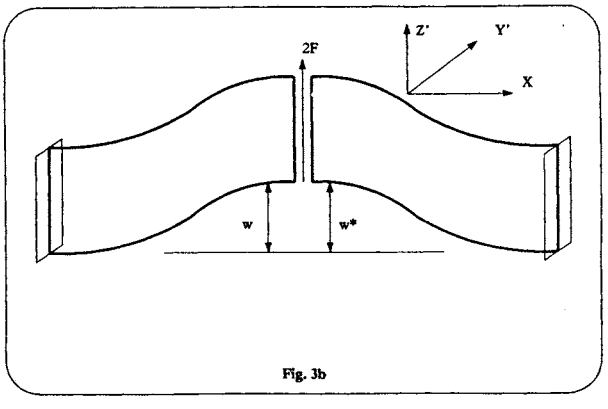


Fig. 3b

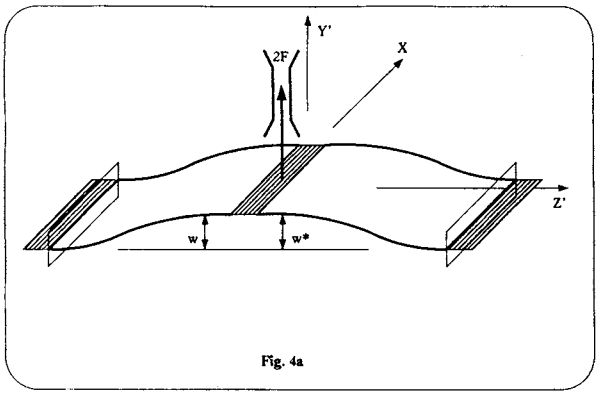


Fig. 4a

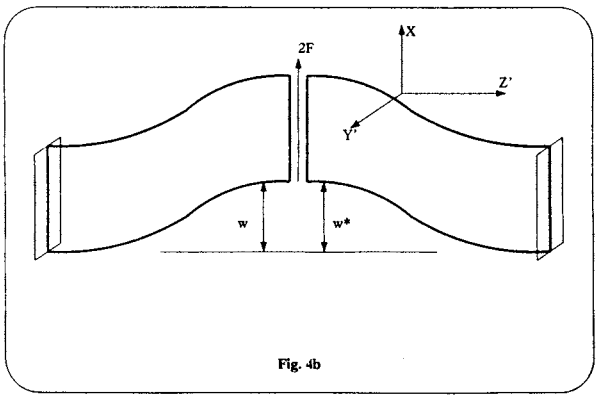


Fig. 4b

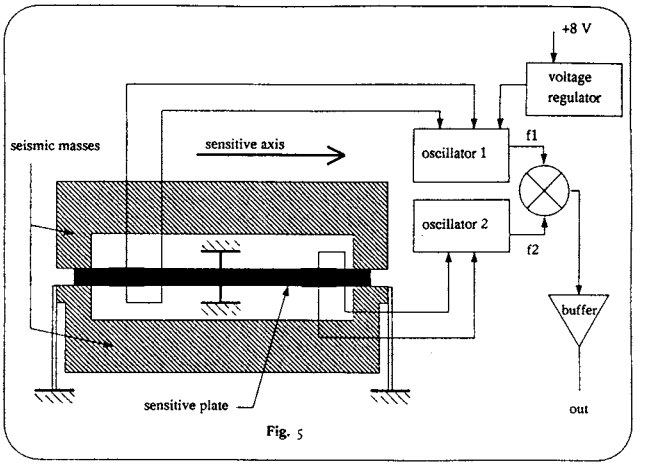


Fig. 5

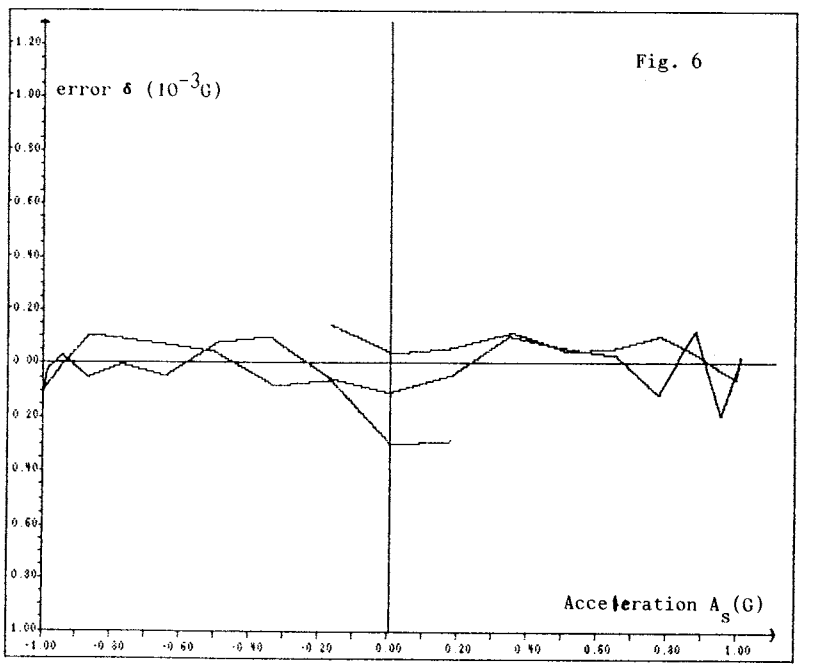


Fig. 6

Frequency Standards

FS-04 Advanced Frequency Standards

Chairman: M. Granveaud

Session FS-04 Number 1

Laser frequency stabilization with applications to atomic frequency standards

K.Nakagawa and M.Ohtsu

Graduate School at Nagatsuta, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 227, Japan

Abstract
Recent works on frequency stabilizations of semiconductor lasers are introduced and several applications are demonstrated. The stability of the Rb atomic clock is expected to be improved by using these stabilized lasers. Several works on optically pumped Rb atomic clock are shown and further improvements are discussed.

1. Introduction

Frequency stabilizations of lasers are very important factors in many applications using lasers. High stability and wide wavelength range of stabilized lasers should be required in many applications of these lasers. Recently, semiconductor lasers are developed and widely used especially for optical communications and optical information systems. Very recently, oscillation wavelength range is extended to 670nm, which is close to that of a red HeNe laser. Available output power is higher than 1W, which is comparable to that of the Ar^+ laser. The frequency fluctuations of semiconductor lasers have been reduced to be reached to stable gas lasers. These progresses make it possible to use semiconductor lasers in many applications such as high resolution spectroscopy, precise measurement for fundamental physics[1].

In the field of frequency and time standards, optical frequency standard is important and required to be more accurate. The coherent optical communication system requires absolute frequency reference in near-infared region[2]. The frequency stabilized semiconductor lasers will play an essential role in this field. Another important application of semiconductor lasers is optically pumped atomic frequency standard. The usage of stabilized semiconductor lasers can improve further stability and contribute to improve the realiability[3)-6)].

In consideration of frequency standards, we show the recent investigations of frequency stabilization of semiconductor lasers and an optically pumped Rb atomic clock of our group.

2. Frequency stabilizations of semiconductor lasers

Magnitude of frequency fluctuation of free running semiconductor lasers is generally larger than that of other stable gas lasers. However, the characteristics of frequency tunability and fast response by the injection current make it possible to stabilize its frequency by feedback techniques. There are mainly two methods of negative feedback for frequency stabilization. One method uses electrical feedback by controlling the injection current and the other uses the optical feedback from an external reflector or cavity.

In the case of electrical feedback method, an external cavity is used as the frequency discriminator and the reflected light from the cavity is detected and fed back to the injection current of the laser[7](Fig.1). In this method, the high feedback gain and desirable feedback bandwidth is available. Figure 2 shows the improvement of the frequency fluctuation by using this electrical feedback. The obtained linewidth is

560Hz which is further less than the Schallow-Townes' limit of linewidth of 55kHz in free running condition. The advantages of this method are that direct frequency modulation is available and the instability induced by the optical feedback is avoidable. This electrical feedback method was also employed to realize a homodyne-type optical phase locked loop. The residual phase variance as low as 0.15 rad^2 was obtained by expanding the feedback of bandwidth of 100 MHz[8].

Another method is to use a optical feedback from an external cavity[9]. This method can reduce frequency fluctuation in wide frequency bandwidth which is determined by the linewidth of the external cavity. The linewidth of less than 10kHz can be easily achieved[10]. We develped compact semiconductor laser modules with using this method for practical applications[10](Fig.3). This method was also applied to the heterodyne type optical phase locked loop[11](Fig.4). The residual phase error variance was reduced to 0.02 rad^2(Fig. 5).

Recently, semiconductor laser oscillations have reached to visible region(<670nm). In the near future, the HeNe laser(633nm) will be replaced by these visible semiconductor lasers in many fields. However, the spectral characteristic of these lasers is still worse than that of the near-infrared semiconductor lasers. The typical linewidth of this laser is order of several hundred MHz. We have improved the linewidth of this visible semiconductor laser of less than 50kHz with using the optical feedback method[12]. This result suggests the potential capability of optical frequency standard of using visible semiconductor lasers. Now we are trying to stabilize these laser frequencies to atomic transitions of Li at 670nm.

In the case of the optical feedback method, a distance between a laser facet and an external reflector should be controlled to adjust the phase of reflected light. This technical inconvenience can be avoided by replacing the external mirror by a phase conjugate mirror(PCM). We have demonstrated the linewidth narrowing by the PCM which use the four-wave mixing process in the semiconductor laser medium[13]. The noncolinear light from the master laser is incident into the slave laser and the phase conjugate light is fed into the master laser. The linewidth of the master was reduced to be less than fifth of that of the free running and was not affected by varying the distance between two lasers.

Frequency stabilization of the semiconductor lasers has been carried out by using atomic and molecular transitions as frequency references, and the stability as high as 2×10^{-12} was obtained at the integration time of 100s[14]. A novel method of using an internal second harmonics of an InGaAsP laser of 1560 nm was proposed to stabilize its frequency to the Rb absorption spectral line at 780 nm[15](Fig. 6). This method will provide frequency chain between the 1560nm laser and 780nm laser by locking simultaneously to the Rb absorption line. To increase an efficiency of second harmonics of semiconductor lasers, we are trying to use an external SHG from nonlinear material.

3. Application to optically pumped Rb atomic clock

It has been mentioned that the usage of stabilized lasers for optically pumped atomic clock can improve the stability of the atomic clock[3]. We have improved the stability of a Rb atomic clock by employing a frequency stabilized semiconductor laser[4](Fig. 7).

One of the advantage of using a laser instead of a Rb lamp is that an extremely narrow linewidth of 200Hz can be obtained in a optical-microwave double resonance signal and this result contributes to improve a short-term frequency stability of the Rb atomic clock[4],[16](Fig. 8). The S/N ratio of the double resonance signal is mainly determined by the FM noise of the semiconductor laser. By employing the electrical feedback method, the FM noise of the laser was reduced at the microwave modulation frequency of 1 kHz and the S/N ratio of the double resonance signal was improved to be more than 10 dB(fig. 9).

The long-term stability of the clock is improved by employing the stabilized lasers[17]. Although the light shift causes the drift and error of the microwave clock frequency, this error is canceled by the self-tuned frequency stabilizing of the semiconductor laser. The slight frequency detuning of the laser from

the Rb transition causes light shift and asymmetry of the line shape of the signal. By employing a computer controlling system, this asymmetry was estimated and the effect of light shift was compensated by the feedback of the injection current of the laser. Figure 10 shows the result of the improvement and the drift was about forty five lower than that of the free running.

The transition frequency of the present clock is affected and shifted by the pressure shift of buffer gases. Now we are constructing a Rb-beam type optically pumped atomic clock.

4. Summary

We have investigated the frequency stabilization of the semiconductor laser. The experimental results are summarized as follows.

center frequency stabilization	2×10^{-12}
linewidth reduction	560 Hz
frequency tracking	1.1×10^{-18}

These stabilization and controlling technics of semiconductor lasers can be directly applied to optical frequency standard in visible and near-infrared region.

The optically pumped Rb atomic clock was stabilized by using the stabilized semiconductor lasers. The short-term stability was improved to be $7.9\times 10^{-13}\tau^{-1/2}$. Further improvement can be expected by employing the FM noise reduction of more than 60 dB, which can be achived by employing the electrical feedback method(Fig.2). The frequency drift induced by light-shift was reduced by self-tuned controlling system and the long-term stability was obtined to be as low as 6.3×10^{-13}/h.

References

1) C.E.Tanner and C.E.Wieman, Phy.Rev.A **38**, 162-165(1988).
2) S.Sudo, Y.Sakai, H.Yasaka, and T.Ikegami, IEEE Photon. Technol. Lett., Vol.1 281-284(1989).
3) J.L.Picque, Metrologia **13**, 115-119(1977).
4) M.Hashimoto and M.Ohtsu, IEEE J.Quantum Electron. **QE-23**, 446-451(1987).
5) S.Ohshima, Y.Nakadan and Y.Koga, IEEE Trans. Instrum. Meas. **37**, 409-413(1988).
6) D.W.Sesco and C.E. Wieman, Opt. Lett. **14**, 269-271(1989).
7) M.Ohtsu, M.Murata and M.Kourogi, IEEE J.Quantum Electron., **QE-26**, Feb. issue (1990).
8) M.Kourogi and M.Ohtsu, prepared for submission.
9) B.Dahmani, L.Hollberg, and R.Drullinger, Opt. Lett. **12**, 876-878(1987).
10) C.H.Shin, M.Teshima, M.Ohtsu, T.Imai, J.Yoshida, and K.Nishide, IEEE Photonics Technol. Lett., **2**, March issue, (1990) to be published.
11) C.H.Shin and M.Ohtsu, IEEE Photonics Technol. Lett. **2**, April issue, (1990) to be published.
12) H.Suzuki, K.Nakagawa and M.Ohtsu, prepared for submission.
13) I.Koshiishi, H.Suzuki, and M.Ohtsu, prepared for submission.
14) M.Ohtsu, Optical and Quantum Electron. **20**, 283-300(1988).
15) M.Ohtsu and E.Ikegami, Electron. Lett. **25**, 22-23(1989).
16) M.Hashimoto, and M.Ohtsu, J.Opt.Soc.Am.B **6**, 1777-1789(1989).
17) M.Hashimoto, and M.Ohtsu, IEEE Trans. Instrum. Meas. May issue, (1990) to be published.

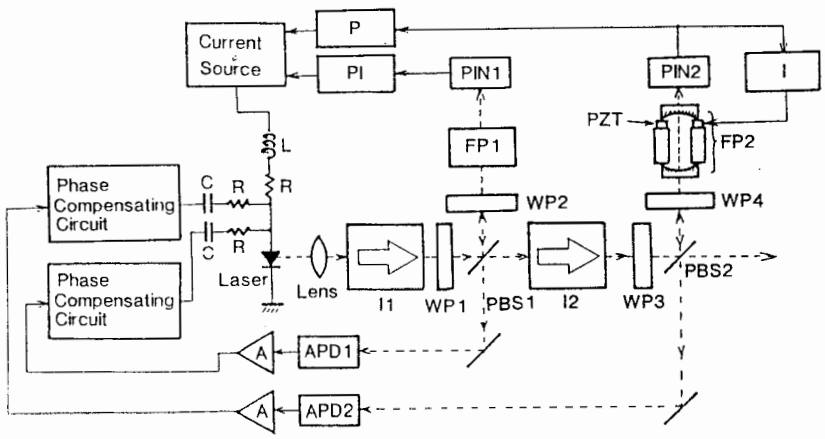


Figure 1. Block diagram of an electrical feedback semiconductor laser(Ref.7).

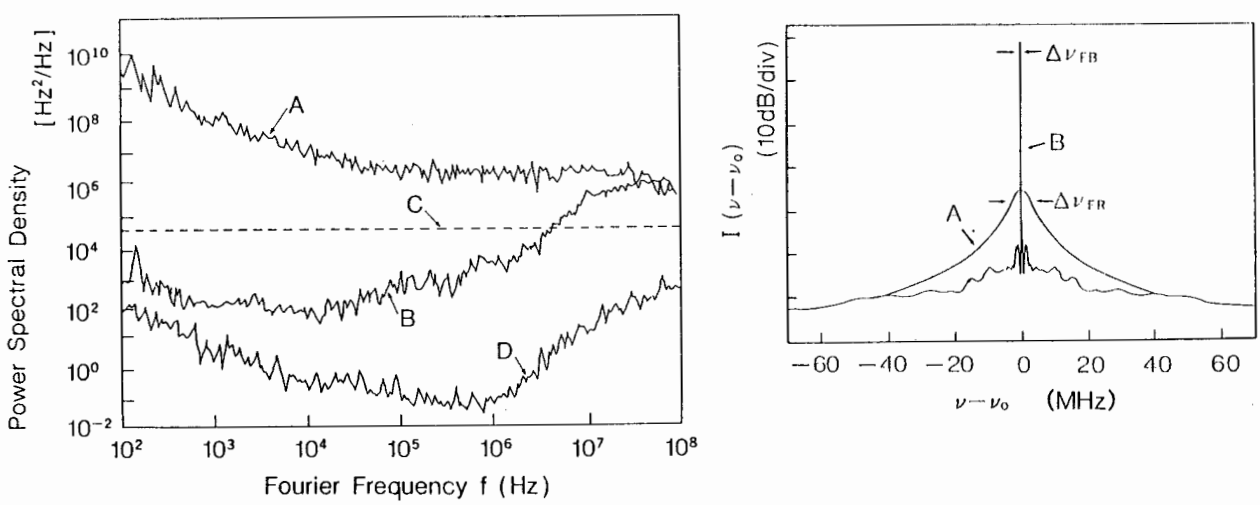


Figure 2. Progress of the experiments on linewidth reduction by electrical feedback(Ref. 7). Power spectral densities of frequency fluctuations are shown in (a). Profiles of the field spectrum are shown in (b). A: Free running. B: With electrical feedback. C: Shallow-Townes' limit of the free running laser. D: Detection noise limit of the electrical feedback laser.

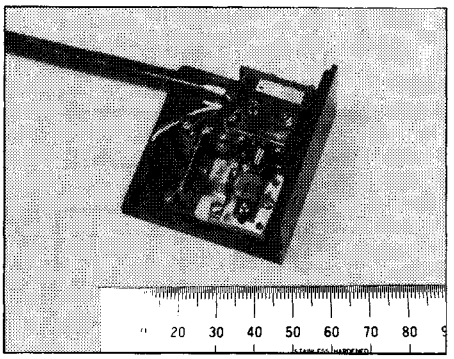


Figure 3. Compact module of an optical feedback semiconductor laser (Ref.10).

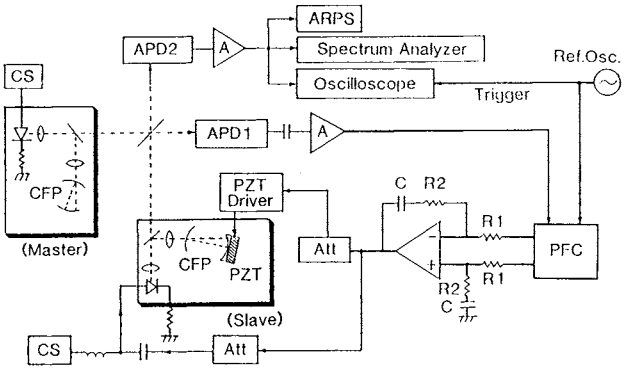


Figure 4. Block diagram of the heterodyne optical phase-locked loop(Ref.11).

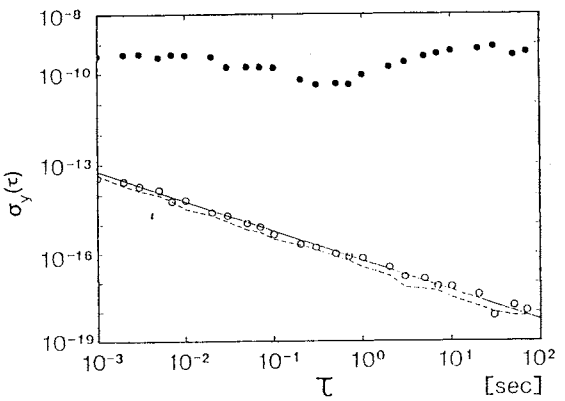


Figure 5. Measured results of Allan variances of residual frequency fluctuations of the heterodyne signal(Ref.11). ● and ○ represent in free running and phased-locked condition, respectively.

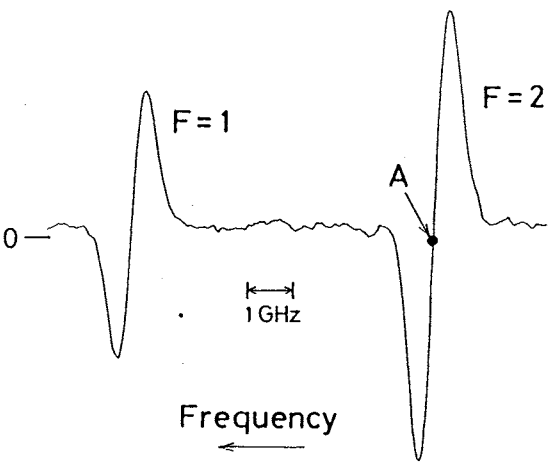


Figure 6. First derivative signal of ^{87}Rb-D_2 spectral lines. The internal second harmonics frequency was locked to point A(Ref.15).

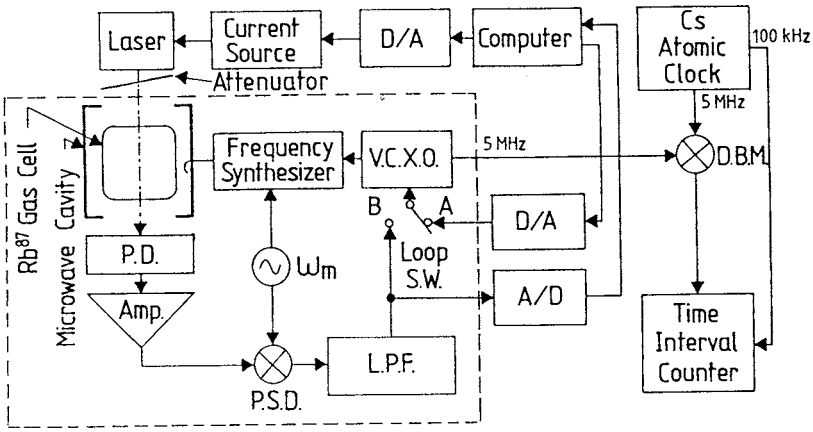


Figure 7. Block diagram of a semiconductor laser pumped Rb^{87} atomic clock(Ref.17).

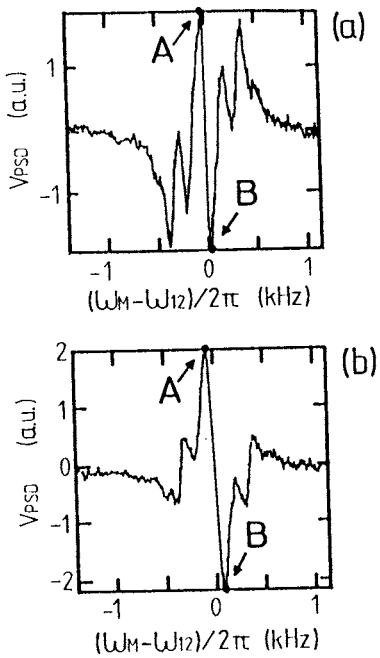


Figure 8. Typical spectral lineshapes of double resonance signal (Ref.16).

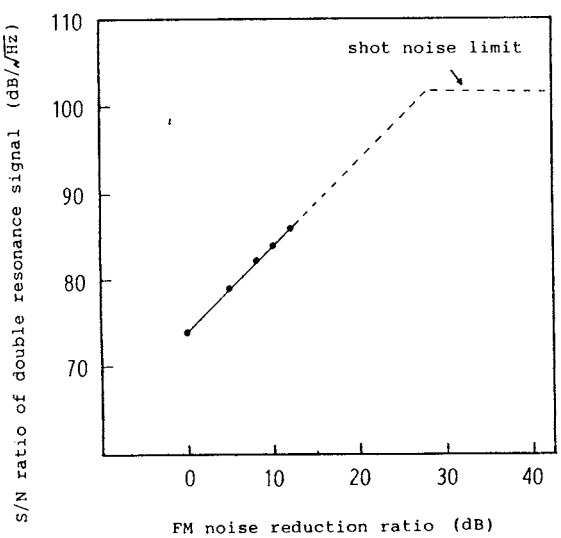


Figure 9. Improvement of the S/N ratio of the double resonance signal.

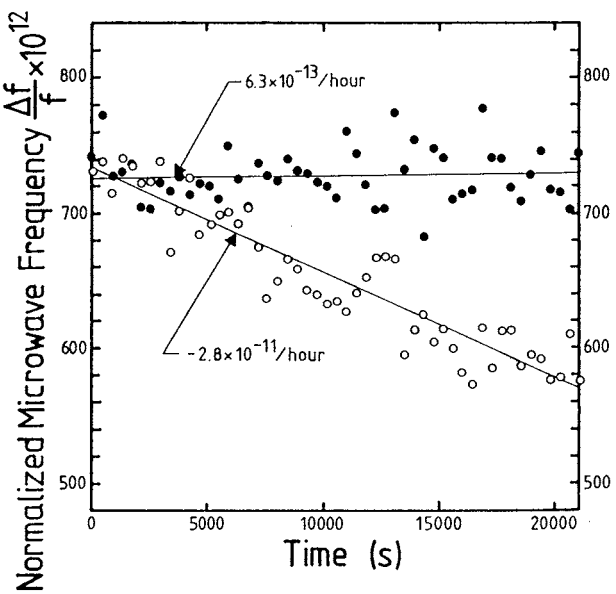


Figure 10. Drift of the microwave frequency. ● and O represent the results obtained with and without the self-tuning method, respectively (Ref.17).

Session FS-04 Number 2

Operation of a cesium beam resonator optically pumped with cesium D_1 line.

N. Dimarcq, G. Théobald and P. Cérez
Laboratoire de l'Horloge Atomique
Unité Propre de Recherche du CNRS associée à l'Université Paris Sud
Bâtiment 221, Université Paris Sud, 91405 Orsay Cédex, France

ABSTRACT.

In this paper, we have first investigated both from the theoretical and experimental point of view the operation of a cesium beam resonator optically pumped with the D_1 resonance line. We have then compared the performances of the resonators operating with D_1 and D_2 lines. The most efficient excitation schemes of these lines are $3 \mapsto 4\Pi$ and $3 \mapsto 3\sigma$ respectively. They lead to comparable high clock signal to noise ratio values. However, the $3 \mapsto 4\Pi$ (D_1) scheme seems to be more promising than the $3 \mapsto 3\sigma$ (D_2) scheme for the two following reasons : i) the efficiency of the optical processes performed with the $3 \mapsto 3\sigma$ (D_2) scheme depends on the magnetic field amplitude. It is not the case with the $3 \mapsto 4\Pi$ (D_1) scheme. ii) the microwave spectrum obtained from the $3 \mapsto 3\sigma$ (D_2) scheme may exhibit a sligth asymmetry which does not exit in the case of he $3 \mapsto 4\Pi$ (D_1) scheme. We show that this asymmetry is mainly linked to the locking of the laser frequency at the top of the fluorescence profile.

I - INTRODUCTION.

The interest of using laser diode for state preparation and clock signal detection in an Optically Pumped Cesium Beam Resonator (OPCsBR) has been already demonstrated. Up to now, all the experiments have been performed with a laser diode tuned to the cesium D_2 line at 852 nm [1,2]. New laser diodes emitting about 894 nm are now commercially available. So it is of interest to investigate the performances of a cesium beam resonator optically pumped with D_1 line at 894 nm. In this paper, we first present in Part II the theoretical predictions relative to the pumping efficiency and to the fluorescence signal features when the atomic beam is irradiated by a broadband laser diode tuned to the D_1 line. The Hanle effect which may critically affect the clock signal to noise ratio [3] is investigated in details. We also show that the anisotropic character of the fluorescence intensity, which has been previously pointed out for the D_2 line, does not exist any more with the D_1 line. Part III presents the experimental results relative to an OPCsBR operating with D_1 line. The excitation schemes leading to the best signal to noise ratios are derived and compared to the theoretical predictions of Part II. Finally, we compare in Part IV the capabilities of the best excitation schemes of the D_1 and D_2 lines to optimize the operation of an OPCsBR.

II - CESIUM D_1 LINE - THEORETICAL APPROACH.

We consider the interaction between a monokinetic atomic beam and a light beam respectively directed along the x and y axis (Fig.1). A dc magnetic field $\vec{B}$ directed along the x axis fixes the quantization axis. The energy diagram of the cesium atom is shown in Fig.2.

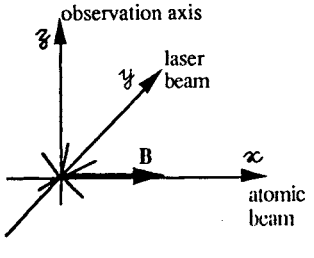


Fig.1 Excitation configuration.

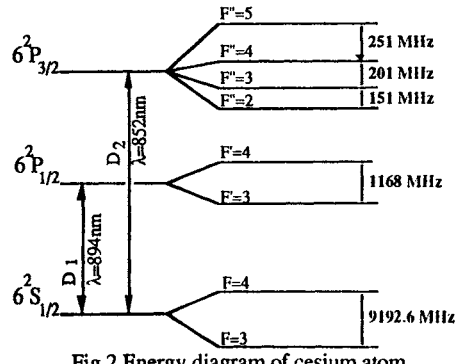


Fig.2 Energy diagram of cesium atom.
(Not to scale)

The optical excitation is performed by a broadband laser diode tuned to the D_1 line at 894 nm. According to the selection rules $F'-F=0,\pm 1$, all the transitions between the $6^2S_{1/2}$ and $6^2P_{1/2}$ levels are allowed, namely the $3\mapsto 3$, $3\mapsto 4$, $4\mapsto 4$, $4\mapsto 3$ transitions. Each of them perform hyperfine optical pumping. As a consequence, the detection of the clock signal using a cycling transition with high photon yield is not possible when D_1 line is considered. Therefore, we only study the performances of an OPCsBR operating with the same laser diode for the state preparation and the detection of the clock signal. The laser beam may be polarized Π or σ depending on whether the polarization vector is parallel or perpendicular to the static magnetic field. An optical excitation scheme is then defined by both an atomic transition and a laser polarization. A theoretical model based on atomic density matrix formalism has been developed in order to determine the best excitation scheme. As the laser excitation is broadline, the evolution of the atomic density matrix elements can be derived from appropriate "rate equations" [4]. In these equations, the laser radiation couples the atomic populations $n(F,m_F)$ of the $|F,m_F\rangle$ Zeeman states to the $\Delta m=\pm 2$ Zeeman coherences . These coherences are only generated when the polarization is coherent. They lead to the detrimental Hanle effect which may decrease both the pumping efficiency and the fluorescence intensity when the static magnetic field amplitude becomes too low [3]. As a consequence, the clock signal to noise ratio decreases because the signal amplitude decreases whereas the background and the noise associated to the unpumped atoms increase [5]. We are thus interested in the two following quantities:

- the fractional population difference Δn between the $m_F=0$ ground state Zeeman sublevels.

$$\Delta n = \frac{n(F=4,m_F=0) - n(F=3,m_F=0)}{n}$$

where n is the total population of the cesium atom. Δn specifies the efficiency of the light to create a large population difference.

- the yield L_F of fluorescence photons which are collected by the optical detection set-up. L_F generally depends both on the populations of the excited state Zeeman sublevels and on the geometrical characteristics of the optical collection set-up. It has been previously shown that the fluorescence may exhibit anisotropy which can be revealed when the collection is performed inside a limited solid angle [6]. This effect can be taken into account through a corrective term in the expression of L_F. However, this corrective term is always equal to zero in the case of the D_1 line. Consequently, there is no anisotropy effect in the fluorescence signal and the detected fluorescence yield L_F is directly proportional to the populations of the excited state and to the collection efficiency of the fluorescence photons.

The rate equations have been numerically solved in order to predict the evolution of Δn and L_F as a function of the interaction time T_i and the magnetic field amplitude B. Fig. 3 shows the evolution of Δn versus T_i for B=500 milligauss and a laser intensity equal to 3 mW/cm^2. It may be noticed that the evolution times of Δn are comparable for the D_1 and the D_2 lines because the lifetime of the excited states $6^2P_{1/2}$ and $6^2P_{3/2}$ have the same value τ=30 ns. Table I reports the values of Δn and L_F for the four transitions of D_1 line and for each polarization Π or σ. The magnetic field amplitude is choosen large enough to cancel the influence of Zeeman coherences. The efficiency of an excitation scheme is characterized by the pumping efficiency in the preparation region and the fluorescence yield in the detection region. It allows us to classify the schemes from the best (number 1) to the worse (number 8).The $4\mapsto 3\sigma$ scheme seems to be the most efficient. Nevertheless, this scheme has a very large pumping time (Fig.3) and is affected by Hanle effect. The $3\mapsto 4\Pi$ scheme is apparently less efficient but has a short pumping time and is not affected by Hanle effect due to its Π polarization.

The dependence of Δn and L_F on B value are presented in Fig.4a and Fig.4b respectively for the σ polarized schemes. It clearly appears that the pumping processes are fully efficient for magnetic field values higher than 300 milligauss.

Excitation scheme	Δn (%)	L_F (ph/at)	efficiency classification	Hanle effect
$4\mapsto 3\sigma$	-16.5	2	1	yes
$3\mapsto 4\Pi$	14.4	1.2	2	no
$4\mapsto 4\sigma$	-15.3	0.85	3	yes
$3\mapsto 4\sigma$	9.5	1.2	4	no
$4\mapsto 3\Pi$	-10	0.95	5	no
$3\mapsto 3\sigma$	11.9	0.65	6	yes
$4\mapsto 4\Pi$	0	0.6	7	no
$3\mapsto 3\Pi$	0	0.5	8	no

Table 1 Performances of the excitation schemes of D_1 line.
Interaction time T_i =20 µs;
Laser intensity = 3mWcm^{-2};
Detection efficiency ρ=0.5.

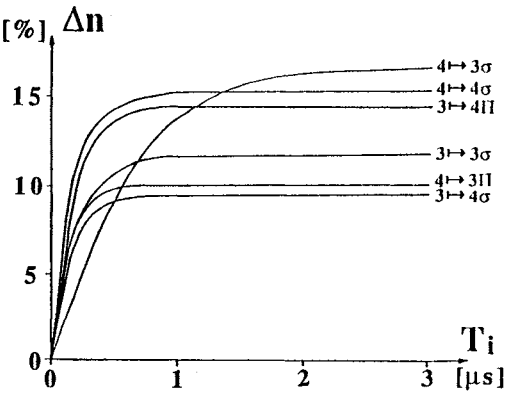


Fig.3 Evolution of the fractional population difference Δn as a function of the interaction time T_i.

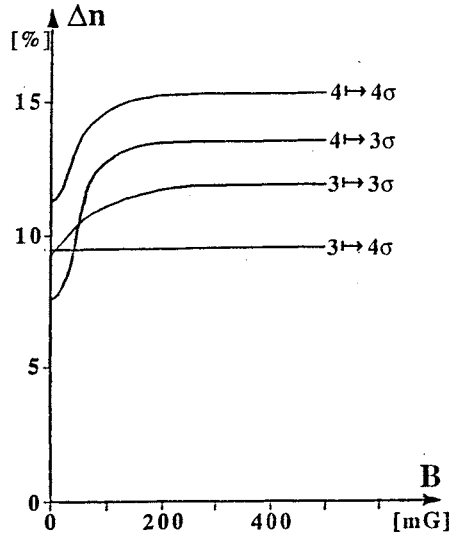


Fig.4a Variations of Δn versus the magnetic field amplitude.

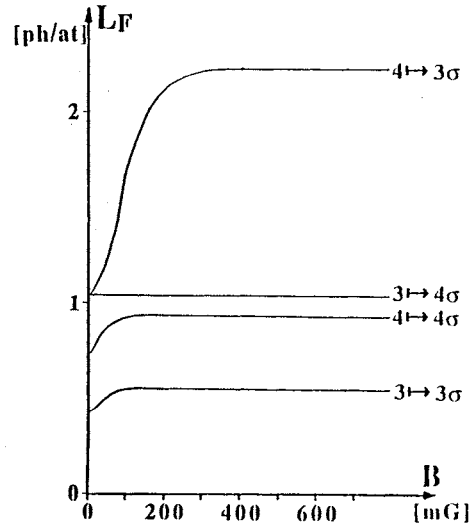


Fig.4b Variations of L_F versus the magnetic field amplitude.

III - CESIUM D_1 LINE - EXPERIMENTAL RESULTS.

The experiments have been realized on the OPCsBR described in Ref.[5]. The laser diode emitting at 894 nm is ML2701 type from Mitsubishi. Its output power is about 3mW. The best excitation scheme has been experimentally found to be 3↦4Π. Fig.5 shows the Ramsey fringe obtained when this excitation scheme is considered. The clock signal to noise ratio is S/N=15,000 in one hertz bandwidth. As the FWHM of the central part of the fringe is about 500 Hz, this would lead to a short term frequency stability of the order of $2 \cdot 10^{-12} \tau^{-1/2}$.

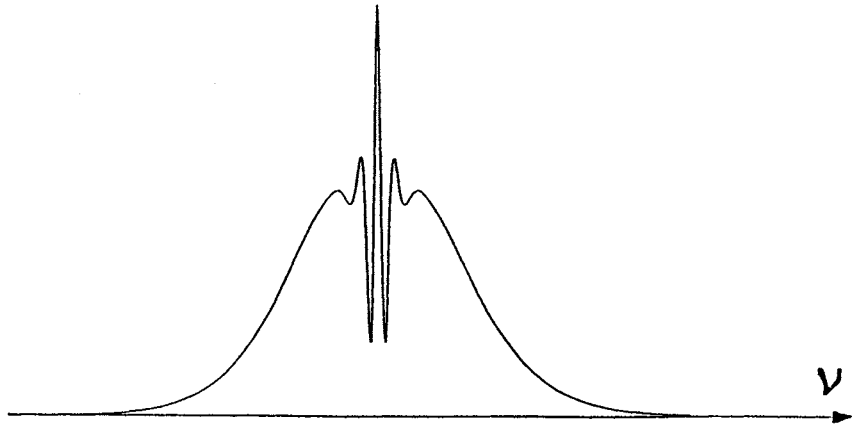


Fig.5 Ramsey fringe obtained with the 3↦4Π (D_1) excitation scheme.

The 4↦3σ scheme leads to a clock signal to noise ratio whose measured value is not so good. This is due to an extra noise linked to the two following effects:

- the pumping time for 4↦3σ scheme is rather long. Consequently, a few atoms are unpumped in the preparation region. The fluctuations of the number of these atoms reflect the frequency noise of the laser. They contribute to an extra amplitude noise in the detection region.
- the 4↦3σ scheme partially behaves like a cycling transition due to important values of the transition probabilities between the Zeeman sublevels of the F=4 and F'=3 states. Therefore, the corresponding fluorescence signal L_F increases compared to the signals given by the other excitation schemes (Table I).However, a detrimental amplitude noise occurs in fluorescence signal as it was previously observed when a cycling transition is excited by a broadband laser [7].

Concerning Hanle effect experiments, the variations of L_F versus B value are depicted in Fig.6. We note a good agreement with the theoretical predictions (Fig.4b).

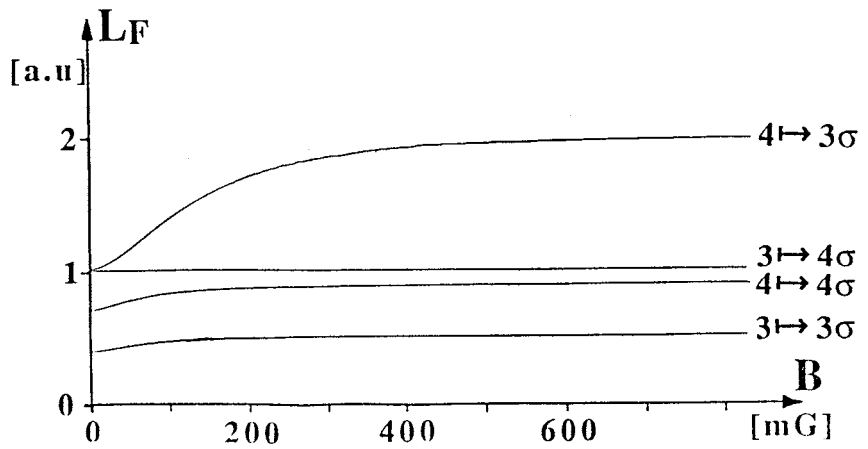


Fig.6 Variations of L_F versus B value.

IV - RESONATORS OPERATING WITH D_1 AND D_2 LINES : COMPARISON.

a - Clock signal characteristics.

The characteristics of the most efficient excitation schemes of D_1 and D_2 lines ($3 \mapsto 4\Pi$ and $3 \mapsto 3\sigma$ respectively) are summarized in Table 2. They lead to comparable clock signal to noise ratio values about 15,000. Nevertheless, in the case of the $3 \mapsto 3\sigma$ scheme, this value is only reached if the static magnetic field value is higher than 300 milligauss. This is the consequence of Hanle effect which affects the S/N ratio through the decrease of the signal and the increase of the unpumped atoms extra noise. The dependence of the clock signal noise spectrum on the magnetic field value is clearly pointed out in Fig.7a for the 3-3σ scheme. However, Fig.7b shows that the clock signal noise spectrum is independent on B value for the 3-4Π scheme. This scheme appears as the most efficient for an OPCsBR operating with a low value uniform magnetic field.

Cesium excitation line	D_2 (852nm)	D_1 (894nm)
Recommended excitation scheme in a one laser configuratiion.	$3 \mapsto 3\sigma$	$3 \mapsto 4\Pi$
Pumping efficiency Δn for laser intensity=3mWcm^{-2} and laser linewidth=25MHz	Δn =12% if the static magnetic field B>300mG	Δn =14% independent on B value
Detected fluorescence yield L_F for a detection efficiency ρ=0.5	L_F=2 ph/at if B>300mG	L_F=1.2 ph/at independent on B value
Measured clock S/N ratio in one hertz bandwidth for a beam intensity I_{beam}= 10^{11} at.s^{-1}	∼ 15,000 if B>300mG	∼ 15,000 independent on B value
Physical properties of the excitation scheme	Δn and L_F are sensitive to Hanle effect in a low magnetic field value	Δn and L_F are independent on B value
Symmetry of the microwave spectrum	Slight asymmetry depending on the locking of the laser frequency	Good symmetry linked to a well defined position of the locked laser frequency
Detection opportunity	Capability of detection of the clock signal using a high photon yield cycling transition	No cycling transition avalaible

Table 2 Characteristics of the most efficient excitation schemes of D_2 and D_1 lines (B is the magnetic field in the optical region).

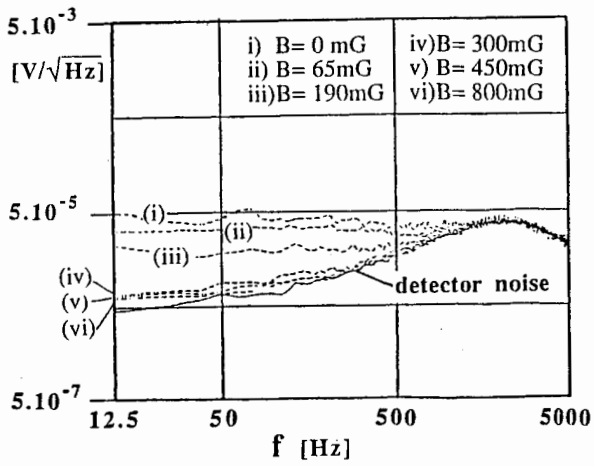


Fig.7a Dependence of the clock signal noise amplitude on B value for the $3 \mapsto 3\sigma$ (D_2) scheme.

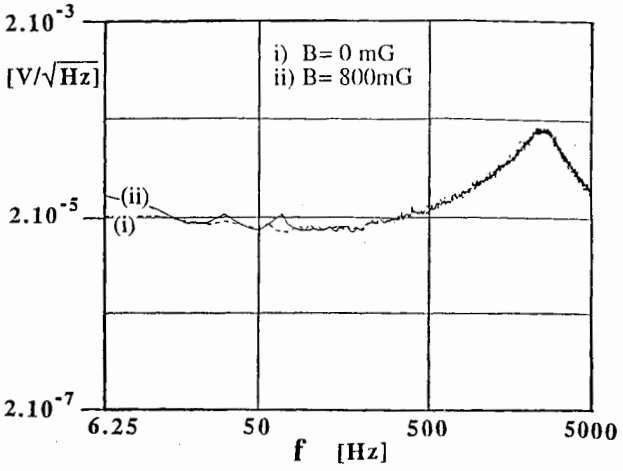


Fig.7a Dependence of the clock signal noise amplitude on B value for the $3 \mapsto 4\Pi$ (D_1) scheme.

b - Symmetry of the microwave spectrum.

It is also of interest to study the symmetry of the microwave spectrum which influences the long term frequency stability through Rabi pulling effect. Fig.8a and Fig.8b show the microwave spectra obtained when the laser is tuned to the $3 \mapsto 3\sigma$ scheme of D_2 line and to the $3 \mapsto 4\Pi$ scheme of D_1 line respectively. The spectrum of Fig.8a exhibits a slight asymmetry which does not occur on the spectrum of Fig.8b. This asymmetry is linked to the locking of the laser frequency at the top of the resonance line. The shape of the $3 \mapsto 3\sigma$ resonance line is strongly affected by the neighbouring cycling transition $3 \mapsto 2\sigma$ (Fig.9a). As a consequence, its top is shifted from its actual position. The laser frequency is no longer locked on the center of the resonance line. This can induce an imbalanced distribution of the Zeeman sublevels populations in the ground state. Moreover, the closeness of $3 \mapsto 2\sigma$ cycling transition induces leakages leading to an asymmetrical evolution of the Zeeman sublevels populations. These effects do not exist in the case of the D_1 line because the different resonance lines are well separated (Fig.9b).

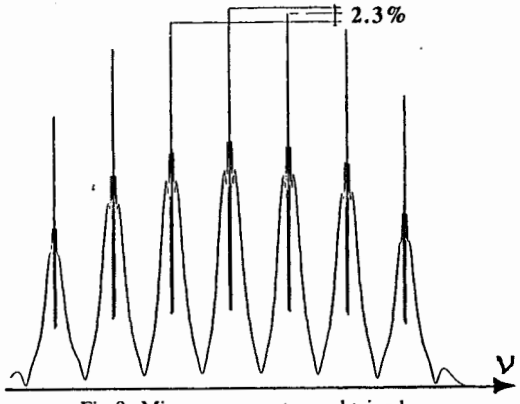


Fig.8a Microwave spectrum obtained with the $3 \mapsto 3\sigma$ (D_2) scheme.

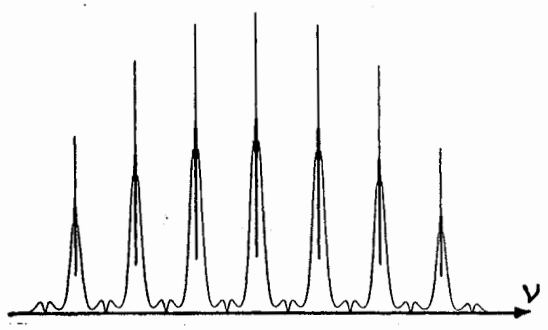


Fig.8b Microwave spectrum obtained with the $3 \mapsto 4\Pi$ (D_1) scheme.

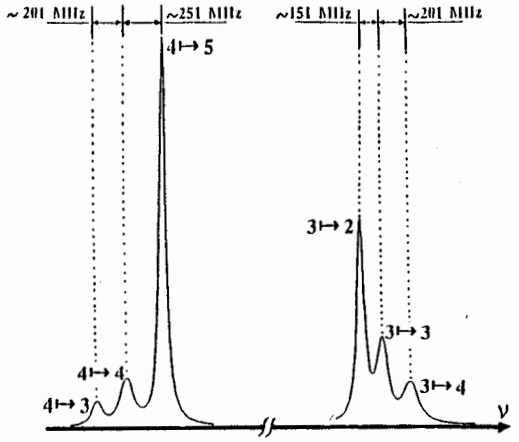


Fig.9a Recorded fluorescence signals for D_2 resonance lines.

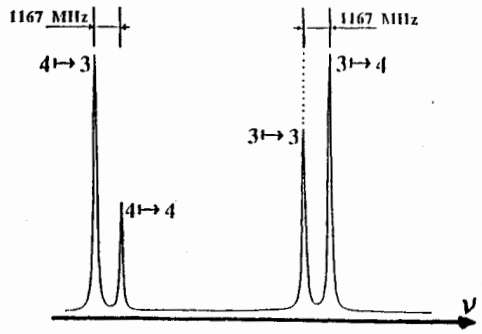


Fig.9b Recorded fluorescence signals for D_1 resonance lines.

It is noteworthy to mention another cause of asymmetry of the microwave spectrum which exists for both D_1 and D_2 lines. The detection efficiency ρ of a fluorescence photon emitted by an atom obviously depends on the position x of this atom in the interaction zone whose length ℓ may be scaled as the "diameter" of the laser beam. Moreover, the fluorescence intensity I_F emitted by a three-level atom also depends on x. The schematic variations of $\rho(x)$ and of $I_F(x)$ versus the position x are shown in Fig.10a for various detunings Δ of the laser frequency. The spatial selectivity of the collection set-up may be characterized by the width W of the curve $\rho(x)$. The overall detected fluorescence intensity I_D is thus the summation over x of the product $\rho(x).I_F(x)$. The measured variations of I_D versus Δ are depicted in Fig.10b for the $3\mapsto 4\Pi$ (D_1) resonance line. They present a characteristic hole at the top of each resonance line. When $\Delta=0$, the main fluorescence intensity is emitted at the beginning of the interaction zone where the collection efficiency is low. When Δ increases ($\Delta=\Delta_1$), the photons are emitted over a higher length and are then more efficiently collected. This leads to a greater signal compared to the signal at resonance $\Delta=0$. If Δ is still increasing, the emitted fluorescence rapidly decreases and I_D decreases.

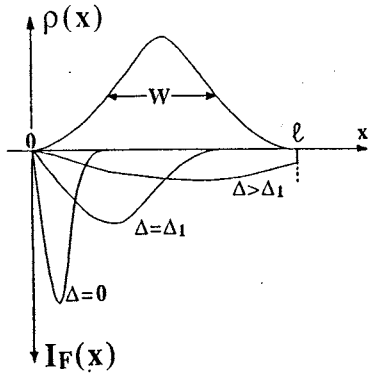


Fig.10a Schematic variations of ρ and I_F versus the position x in the interaction zone.

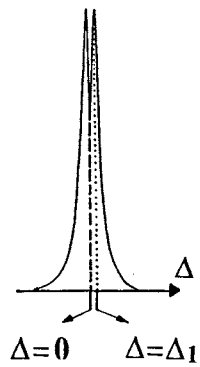


Fig.10b Measured variations of the detected fluorescence intensity I_D versus the laser detuning Δ.

The depth and the width of this "pumping signal hole" depends on:

- the intensity and the linewidth of the laser.
- the collection width W compared to the interaction length ℓ.

They increase when the laser intensity increases or when the laser linewidth decreases. Moreover, when the laser intensity is set to produce a complete pumping, the depth and the width of the hole increases when the ratio W/ℓ increases. Fig.11 illustrates this behaviour.

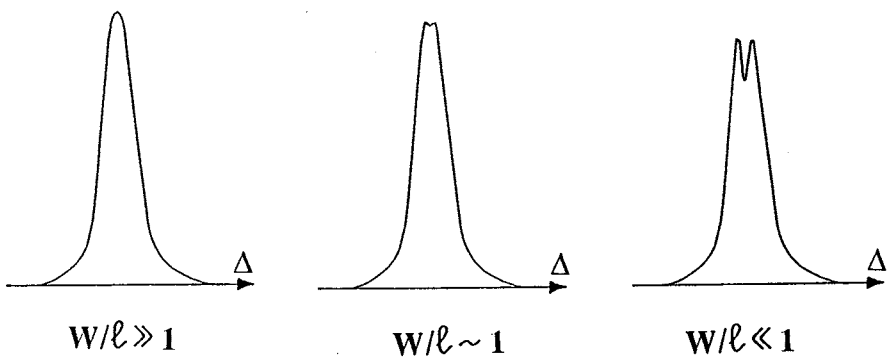


Fig.11 Pumping signal hole profiles for various W/ℓ ratios.

The "pumping signal hole" does not exist if the interaction zone is punctual compared to the collection width W. On the other hand, the fluorescence profile exhibits a strong dip if the interaction length is much greater. An intermediate situation exists when ℓ and W are similar. In this case, the unpronounced existing dip generally leads to a bad locking of the laser frequency at the center of the resonance line.

c - Detection opportunity.

The detection of the clock signal using a cycling transition is only possible with D_2 line. This detection configuration is of interest for the two following reasons:

- the optical signal to noise ratio can be very large if a laser excitation has a good spectral purity [7].
- due to the large emitted fluorescence signal, this ratio has been shown to be independent on the collection efficiency [8]. This situation is not encountered when a pumping transition is used. That is why large, sophisticated and highly efficient collection optics are necessary in resonators using only one laser.

VI - CONCLUSION.

In this paper, we have investigated the operation of a cesium beam resonator optically pumped with the D_1 line. We have then compared the performances of the resonators operating with D_1 and D_2 lines. The most efficient excitation schemes have been determined. They are $3 \mapsto 3\sigma$ for the D_2 line and $3 \mapsto 4\Pi$ for the D_1 line respectively. Both lead to typical clock signal to noise ratio values equal to 15,000. Optical pumping with the $3 \mapsto 3\sigma$ configuration of D_2 line is sensitive to Hanle effect : in this case, it is necessary to use a resonator designed with three magnetic field regions [3]. In the case of the $3 \mapsto 4\Pi$ (D_1) configuration, the magnetic field may be lowered even in the optical regions. Another interesting feature concerns the symmetry of the microwave spectrum. We have investigated the various reasons leading to an imbalance of the ground state populations responsible of the asymmetry. The best configuration using the D_1 line has been found to cancel quite perfectly this asymmetry.

Acknowledgements : The authors want to thank E. de CLERCQ from LPTF, France for his generous loan of laser diodes.

REFERENCES.

1. V. Candelier, V. Giordano, A. Hamel, G. Théobald, P. Cérez and C. Audoin, Appl. Phys. **B49**, 365 (1989)
2. V. Giordano, A. Hamel, P. Petit, G.Théobald, N. Dimarcq, P.Cérez and C. Audoin, Proc. of the 43rd Annual Symposium on Frequency Control (Denver USA 1989) 130.
3. G. Théobald, P. Cérez, N. Dimarcq and V. Giordano, Proc. of the IVth Symposium on Metrology and Frequency Standard (Ancôna,Italy 1988) 110.
4. G. Théobald, N.Dimarcq, V. Giordano and P. Cérez, Optics Comm. **71**, 256 (1989).
5. A. Hamel, P. Petit, G. Théobald, P. Cérez and C. Audoin, Proc. of the 3rd European Time and Frequency Forum (Besançon France 1989) 274.
6. G. Théobald, N. Dimarcq, V. Giordano, A. Hamel and P. Cérez, J. Appl. Phys. **66**, 4581 (1989).
7. M. De Labachelerie, K. Diomandé and N. Dimarcq, Proc. of the 2nd European Time and Frequency Forum (Neuchâtel Switzerland 1988) 547.
8. N. Dimarcq, J. Phys Fr, Colloque, to be published.

Session FS-04 Number 3

FREQUENCY STANDARDS IN THE FAR-INFRARED USING LASER COOLED ATOMIC BEAMS

N. Beverini, E. Maccioni, F. Strumia

Dipartimento di Fisica dell'Università di Pisa

1. THE LASER COOLING TECHNIQUE

Laser cooling has been demonstrated a very effective technique to produce atomic beams with subthermal velocity distribution. The principles of laser cooling is quite simple. When an atom absorbs or emits a photon, the momentum conservation law implies that atomic momentum must change by a quantity equal to the photon momentum hν/c. Let us consider a two level atom, irradiated by a counterpropagating running wave, resonant with the transition frequency. Each absorption-spontaneous emission cycle will lead to an average variation of the atom velocity by a quantity hν/mc. On the contrary, an absorption followed by a stimulated emission, with two identical photons, does not change the atomic momentum. Taking into account also the saturation effects, an atom experiences an effective force, given by the ratio between the photon momentum and the mean time interval between two spontaneous emission, that is

$$F = \frac{h\gamma}{\lambda} \frac{S}{1 + S + (\Delta\omega - kv)^2/\gamma^2} \qquad (1)$$

where γ is the HWHM homogeneous linewidth, $\lambda=c/\nu$ is the resonance wavelength, and $S=I/I_S$ the saturation parameter of the radiation (for S=1 spontaneous and stimulated emissions have equal probability). Thus, an atom with initial velocity v_0 is stopped, in a laser field of constant intensity I, after a distance

$$L(S) = \frac{v_0^2 M\lambda}{h\gamma}\left(1 + \frac{1}{S}\right), \qquad (2)$$

where τ is the upper level lifetime and M the atomic mass.

However, only a small fraction of the atoms over the thermal velocity distribution profile can absorb the laser radiation when a monochromatic single-frequency laser beam is used, namely those atoms whose laser frequency offset is compensated, within the natural linewidth, by the Doppler shift. As soon as the deceleration process starts, the velocity is reduced, the atoms are Doppler shifted out of resonance, and the absorption becomes negligible.

The atoms can be kept in resonance with the radiation during the cooling process, either by chirping quickly the laser frequency [1,2], or by tuning the atomic absorption frequency by the Zeeman effect, produced by an axial magnetic field B(z) of appropriate intensity along the beam [3]. Both the methods have been first demonstrated on Na atomic beams, and then applied also to other atomic species [4,5,6,7]. Frequency chirping requires a frequency laser tuning of the order of some GHz in a fraction of ms; such a performance can be quite easily achieved, if the cooling radiation is produced by a diode laser, like in the case of Cs (λ=851 nm) [4], otherwise sofisticate electronical equipment is required. When the atomic absorption wavelength lies in the blue or in the UV region, the dye laser operation is more critical and less efficient and the magnetic tuning is the most convenient alternative. An additional advantage of the magnetic cooling method is the production of a continuous beam of decelerated atoms, while only pulsed beams can be obtained by laser chirping. The parameters relevant for laser cooling are given in Table I for different atomic species.

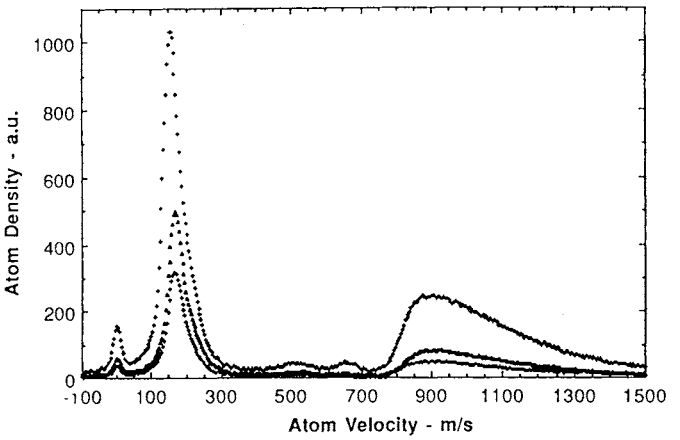


Fig.1.

The velocity distribution obtained by applying the laser cooling technique with Zeeman tuning to an atomic beam of Ca is shown in Fig.1. Three recordings at different oven temperature (560°C, 580°C, 630°C) are superimpose. It is worth to note that the atomic velocities are not simply slowed down, but that an almost monochromatic velocity distribution is produced.

The efficiency of the laser cooling process can be further increased by using the radiation pressure also to collimate the atomic beam. This can be obtained by sending a red shifted standing wave, orthogonally to the beam. In this case, it is easy to demonstrate [8], that an atom in the beam experiences a viscosity-like force

$$F = - \beta M v_t \quad , \tag{9}$$

where v_t is the transverse velocity and β is a function of the detuning and the intensity of the radiation. Impressive increasing in the atomic densities on the axis can be obtained.

The application of atomic beams, decelerated and collimated by radiation pressure, may have a tremendous impact in the development of atomic frequency standards (AFS) with superior performances. In fact the benefits of laser cooling and collimation are crucial in many key points:

1)- By reducing v, the transition Q, which is usually limited by the interaction time, is increased.

2)- By reducing v, a better detection of the transition is possible, thus increasing the S/N ratio. This is of particular relevance for Mg and Ca submillimeter AFS

3)- By reducing v, the RF power for interrogating the atomic transition is reduced as v^{-2}. This will strongly attenuate the technical problems for extending the synthesis of the interrogating radiation toward higher frequencies.

4)- By reducing v, the relevance of most of the small perturbations that are responsible for the maximum attainable precision and reproducibility in the passive primary AFS will be reduced.

5)- By reducing the transverse velocity by laser collimation, the first order Doppler effect line broadening can be reduced to be smaller or equal to the homogeneous line width even for the very high Q expected for the reference transitions in laser cooled atomic beams.

2. APPLICATION OF THE LASER COOLING TO MG AND CA ATOMIC FREQUENCY STANDARDS.

The metastable triplet states $^3P_0$, $^3P_1$,and $^3P_2$ and the associated magnetic dipole transitions at 0.6 and 1.2 THz for Mg, and 1.5 and 3.2 THz for Ca, respectively, have a set of properties that are of particular metrological interest. A scheme of the energy levels, for Mg and Ca respectively, is given in Fig..2 and 3. The radiative decay of the ^3P levels towards the ground $^1S_0$ state is strongly forbidden with the exception of the $^3P_1$ level, which has a lifetime of about 10^{-3} s as a consequence of a small mixing with the singlet $^1P_1$ state of the same configuration. As first proposed in 1972 [9], it is then possible to obtain a spontaneous time of flight states selection and the submillimeter $^3P_0$–$^3P_1$ transition (or the $^3P_2$–$^3P_1$ transition) can be monitored by the fluorescence in the visible from the $^3P_1$ - $^1S_0$ transition. In this way, the submillimeter transition can be detected with a good S/N ratio. Moreover, these fine structure transitions has a Q factor about two order of magnitude larger than that of the hyperfine structure transition, used in the Cs

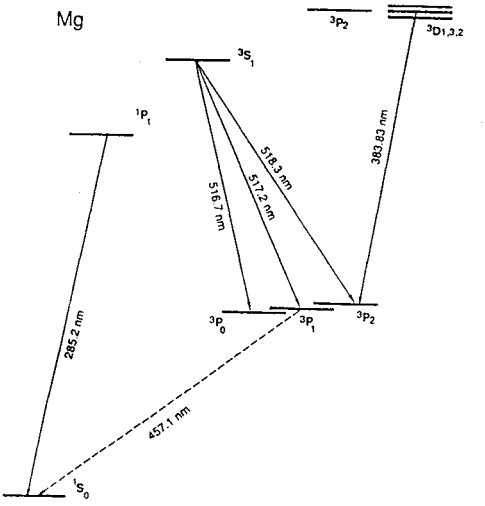


Fig.2

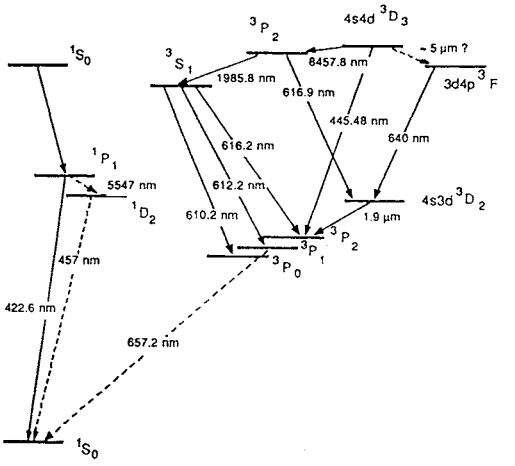


Fig.3

beam atomic frequency standard, and its $m_J=0$ — $m_J=0$ components has a dependence on the residual magnetic field which is about four orders of magnitude smaller, because the quadratic Zeeman coefficient is inversely proportional to the square of the energy splitting. All these factors would promise an atomic frequency standard of better stability, precision , and reproducibility.

A prototype of atomic frequency standard, based on the Mg transition at 601 277 157 860 Hz was effectively realized [10] in the past few years at the IEN (Istituto Elettrotecnico Nazionale "G.Ferraris", Torino), and in collaboration with the Pisa University, by using a thermal effusive atomic beam and the optical Ramsey interrogation scheme. A linewidth of 1.2 kHz ($Q \approx 5x10^8$) and a S/N = 250 was demonstrated with a distance of 30 cm between the two oscillating fields. The stability was measured by comparison with a commercial high performance Cs AFS and was estimated to be $\sigma(\tau) = 8 \times 10^{-12} \times \tau^{-1/2}$, or better. This short term stability is better than that of the present commercial Cs AFS and inferior only to that of the large laboratory Cs AFS of the National Boureau of Standards - USA and National Research Council - Canada . The Q is already larger than that of all the present Cs AFS, and can be easily further increased (L=30 cm for Mg, while L=375 and 213 cm for the NBS IV and NRC V respectively).

As stated above, the use of cooled beams can improve dramatically the performance of an atomic frequency standard. The reduced requirements in microwave power can allow the direct synthesis of the transition frequency, may be up to the 3.2 THz of the Ca transition, in spite of the phase-locked BWO presently used in the Mg standard. To apply this technique to an atomic clock, based on the transitions between the metastable levels of Mg or Ca, the problem of the production of a low velocity metastable beam must be solved. Two different methods can be used. In the first method a ground state thermal atomic beam is slowed down by using the resonance $^1P_1$ — $^1S_0$ transition and then is excited from ground state to the metastable triplet by electronic collisions. In the second one a discharge excites the thermal beam in the metastable levels and then laser cooling is applied directly to the metastable atoms. Both ways are now explored in our laboratory and will be discussed in detail in the following.

A) Cooling in the ground state

A thermal beam of Mg or Ca in the ground state can be slowed down by using the resonance $^1P_1$ — $^1S_0$ transition at 285 nm and 422 nm, respectively [11].

Laser cooling of Mg is difficult because the resonance line lies in the deep UV and has a quite high saturation intensity (444 mW/cm^2). It is then necessary to use an intracavity frequency doubled dye laser. Preliminary results, with a low UV power, have demonstrated the possibility of obtaining effective laser deceleration if a larger laser power were available [5].

Effective laser cooling of Ca has been demonstrated by using magnetic tuning [6]. The velocity distribution in typical experimental condition is shown in Fig 1. Final velocity between 10 and some hundreds m/s can be obtained by properly tuning of the cooling laser beam. Population trapping into the $^1D_2$ metastable level has been demonstrated negligible, measuring the $^1P_1$ — $^1S_0$ versus the $^1P_1$ — $^1D_2$ branching ratio[12]. Without any collimation, final density of low velocity atoms greater than that on the maximum of the unperturbed maxwellian has been obtained also at the lowest final velocities. Further increases in the efficiency can be expected by laser collimation. First attempts have been made by using a portion of the cooling radiation for collimation and confirm the prediction.

To excite the Ca atoms into the metastable levels, an electronic gun has been constructed, which excites orthogonally the atomic beam. In such a design the perturbation of the atomic motion due to the exciting anelastic scattering can be controlled, resulting in a deviation of the metastable beam by an angle of about 1°. First measurements have been made on thermal beams. The efficiency of the excitation was measured by detecting the fluorescence from the $^3P_1$ level downstream. With a current of the order of 6 mA, an efficiency of $3.2x10^{-4}$, corresponding to an excitation cross section of $3x10^{-16}$ cm^2, has been obtained. The efficiency should obviously increase for low velocity atoms, proportionally to the increasing in the interaction time. The perturbation to the atomic motion due to the inelastic scattering between the atoms and the electrons can be controlled, if a collimated electronic beam is used.

B) Cooling in the metastable level

An alternative way to produce low velocity metastable atomic beam is to excite in advance the thermal beam by electron collisions (more than 30% efficiency in metastable production has been obtained by a careful experimental design [13]) and to cool the atoms directly on the metastable states. The cooling transition must be chosen in order to behave, at the best approximation, as a two-level system, in oeder to avoid losses of atoms in the cooling cycle. This is fortunately possible by using the transition connecting the metastable $^3P_2$ level to a $^3D_3$ level. If a circularly polarized light beam is used, the $m_J=+2 - m_J=+3$ Zeeman component is an effective two-level system, provided that no other decay channel is available for the upper $^3D_3$ level.

In Mg the transition lies at 383.83 nm, a wavelength that is now directly available from the new Exalite dye laser, or by frequency doubling high efficient Ti-sapphire laser. A problem in the magnetic tuning may arise because of the peculiar fine structure of the

(3s,2d)^{3}D multiplet . The quadrupole term is effectively of the same order of magnitude of the dipole one, so that the Landé interval rule is not respected, and the J=3 level lies between the J=1 and J=2 levels. As a consequence, the frequency of the $^3P_2$–$^3D_3$ (m_J=+2 – m_J=+3) component at a field of about 7.5 mT is coincident with the frequency of the $^3P_2$–$^3D_1$ (m_J=+2 – m_J=+1) component. At this field, any spurious σ^- component in the laser beam may deplete the $^3P_2$ level. The oscillator strength of $^3P_2$–$^3D_1$ transition is however reduced by a factor 84 with respect to the $^3P_2$–$^3D_3$ transition, so that the problem can be easily overcome by reducing the region of magnetic field around this field value.

In Ca the lower ^{3}D multiplet (4s,3d) cannot be used, because the correspondent $^3P_2$-$^3D_3$ transition lies in the infrared at 1.97 μm; but the following $^3P_2$-$^3D_3$ (4s,4d) transition can be used. The wavelength of 445.48 nm is directly available from the high efficiency Stilbene 3 dye laser. In this case, other decay channels are allowed, through the $^3P_2$(4s,5p) level or, possibly, the ^{3}F(3d,4p) levels. However, the corresponding decay rates are depressed, because of the long wavelength (8.5 μm and 5 μm respectively) and the change in the principal quantum number, and can be estimated to be sufficiently small for allowing an efficient cooling process. Experiment are now in progress to exploit this possibility.

The present paper has been developed under the "Progetto Finalizzato" on Electrooptical Technologies of the National Research Council (C.N.R.) of Italy

Table I
Numerical data on laser cooling (N is the number of photons needed for cooling starting from the velocity v_0 , and I_S is the saturation intensity of the resonance transition)

	^{23}Na	^{133}Cs	^{24}Mg $^1S_0$	^{24}Mg $^3P_2$	^{40}Ca $^1S_0$	^{40}Ca $^3P_2$
λ (nm)	589.0	852.1	285.1	383.8	422.6	445.5
τ (ns)	16	32	2.02	5.95	4.57	11.6
γ HWHM (MHz)	5	2.5	39	13.4	17.5	6.84
Δv (cm /s)	2.29	0.35	5.84	4.29	2.36	2.23
v_0 (m /s)	800	300	1000	1000	800	800
N (x10^3)	35	85.7	17.1	23.3	33.9	35.9
L_m (cm)	45	82.3	3.46	6.13	12.4	33.3
v_t (m /s)	4.3	1.02	7.6	6.55	4.34	4.2
v^t_m (m /s)	0.41	0.12	1.13	0.66	0.58	0.37
I_S (mw /cm^2)	6.4	1.05	444	61.7	60	20.2

REFERENCES

[1] V.S. Letokhov, V.G. Minogin, B.D. Pavlik : Opt. Commun. **19**, 72 (1976)

[2] W. Ertmer, R. Blatt, J. Hall, M. Zhu : Phys.Rev.Lett. **54**, 996 (1985)

[3] W.D. Phillips, J.V. Prodan, H.J. Metcalf: J.Opt.Soc.Am.B **2**, 1751-1767 (1985)

[4] R.N.Watt, C.E. Wieman : Opt.Lett. **11**, 291 (1986)

[5] N. Beverini,S. De Pascalis, E. Maccioni, D. Pereira, F. Strumia, G. Vissani, Y.Z. Wang, C. Novero : Opt. Lett. **14**, 350 (1989)

[6] N. Beverini, E. Maccioni, D. Pereira, F. Strumia, G. Vissani : Ital. Phys. Soc. Conf. Proc., **21**, 205 (1989)

N. Beverini, E. Maccioni, D. Pereira, F. Strumia, G. Vissani : " Frequency Standards. and Metrology", A. De Marchi ed., Springer Verlag, pag. 282 (1989)

N. Beverini, E. Maccioni, D. Pereira, F. Strumia, G. Vissani : "Advance in Laser Science -IV", Conf. Proc. **191**, pag.533-535, Am. Inst. Phys., New York,1989

[7] F. Shimizu, K. Shimizu, H. Takuma : Jap.J.Appl.Phys. **26**, L1847-49 (1987)

[8] V.G. Minogin, V.S. Letokhov: "Laser light pressure on atoms", Gordon and Breach Sc. Publ., New York 1987

[9]. F.Strumia : Metrologia **8**, 85-90,(1972)

[10] E.Bava, A.Godone, C.Novero : Appl.Phys.B **48**, 495, (1989)
E.Bava, A.Godone, C.Novero : Metrologia **24**, 133, (1987)

[11] F. Strumia: "Application of laser cooling to the atomic frequency standards", in "Laser Science and Technology", Ed. da A.N. Chester e S. Martellucci, Plenum Press, N.Y. 1988, pagg. 367-401.

[12] N. Beverini, F. Giammanco, E. Maccioni, D. Pereira, F. Strumia, G. Vissani : J.Opt. Soc.Am. B **6**, 2188-2193 (1989)

[13] G.Giusfredi, P.Minguzzi, F.Strumia, M.Tonelli : Z. Physik, **A274**, 279-287, (1975)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-04 Number 4

THERMALIZATION OF IONS IN A PENNING AND IN A PAUL TRAP

F. PLUMELLE, M. JARDINO, M. DESAINTFUSCIEN and J. YU*

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
*Laboratoire Aimé Cotton Bât. 505 Université Paris-Sud Orsay France

Abstract

We describe some thermalization processes of ions stored in a Penning and in a Paul trap. In particular we study the effect of the resonant coupling of the magnetron and axial motions in the Penning trap and of the background pressure in both traps, when the ions are continuously created in the trap.

Introduction

An ion cloud, confined in a Penning or a Paul trap is submitted to various mechanical excitations which result from confinement fields, electric interactions between the ions inside the cloud, collisions with background molecules (1) and interactions with photons (2) when an optical experiment is performed on the ions. This leads to very complicated physical situations. We can mention, among the most recently described phenomena, the order-chaos transitions observed on a few number of ions cooled to a low temperature by laser radiation (3), the effect of the micromotion on the efficiency of laser cooling (4),...

In this paper, we describe some systematic observations of the effect of physical parameters on the dynamical properties of ions stored in a Penning and in a Paul trap. Unlike the experiments mentioned above, we are concerned with large number of ions.

Evaporation Rate of the Ions

We have studied the evaporation rate of the ions as a function of their temperature.

Experimental Set-up

It has been described elsewhere (5). The same electrod configuration can be used as a Penning or a Paul trap. The temperature and the density of the ions (Mg^+) are measured by monitoring the 1st order Doppler broadened $^2S_{1/2}$ - $^2P_{3/2}$ transition, excited by a laser beam in the equatorial plane of the trap. In order to evaluate the ion creation rate we assume that, for a given value of the Mg pressure, it is proportional to the electron beam intensity, and that for a given value of the electron beam intensity, it is proportional to the Mg pressure in the trap. We define the evaporation rate as the ion creation rate divided by their total number.

Experimental results

First of all, the experimental variations of the ion creation rate have been obtained by changing the electron gun temperature. Figures 1 and 2 show the observed and calculated values of the evaporation rate in the Penning and Paul trap respectively. The numerical calculation (solid line) assumes that the evaporation rate for a given temperature is proportional to the ion number whose total mechanical energy is larger than the well depth D. For the Paul trap, the well depth D is defined as the smallest value of the pseudo-potential well in the three space dimensions ; for the Penning trap, D is defined as the electrostatic depth in the z direction.

The results show that for the highest temperatures, the evaporation rate agrees with the calculated variation and varies according to an exponential law in 1/T. In the case of the Penning trap, the variation of the evaporation rate for lower values of the temperature is much smaller than the calculated one. However, since the laser beam is in the equatorial plane of the trap, we only measure the cyclotron temperature, which is very close to the axial temperature. On the over hand, the magnetron motion is only weakly coupled to

the cyclotron and axial motions and does not broaden the observed line. Therefore these experimental results demonstrate that the ion losses, when the axial / cyclotron temperature is low (kT < D/20), are mainly due to the magnetron motion.

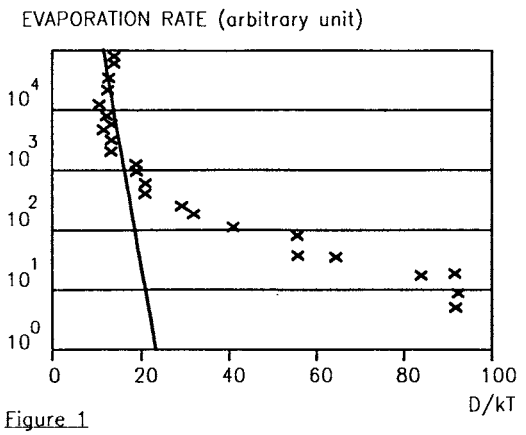


Figure 1

Penning trap. The evaporation rate is modified by varying the electron gun temperature.

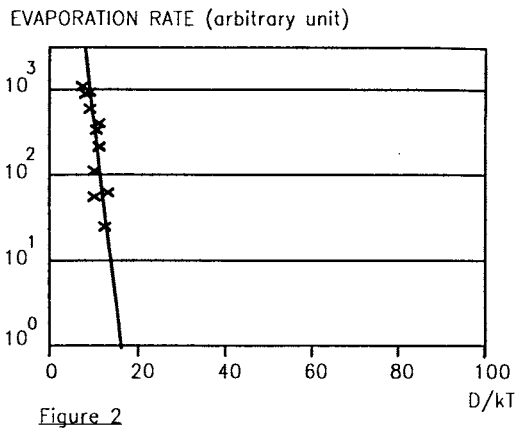


Figure 2

Paul trap. The evaporation rate is modified by varying the electron gun temperature.

In the case of the Paul trap, the all set of experimental points agrees with the (adjusted) calculated law, which validates the pseudopotential well model of the Paul trap. Furthermore we can notice the fast variation of the evaporation rate with the temperature, which explains that the values of kT is always of the order of magnitude of D/10 for any usual creation rates.

In a second step we have studied the variation of the creation rate by changing the oven temperature. In this case, we must keep in mind that another important parameter varies at the same time : the background pressure (mainly neutral Mg), which increases from 1.10^{-10} T to 3.10^{-8} T in these measurements. The results for the Penning and the Paul trap are shown in fig. 3 and fig. 4. In both cases, the evaporation rate is a decreasing fonction of the temperature. This can be explained by the ion-neutral collisions which eject the ions, mainly via the magnetron motion in the Penning trap, while cooling the cyclotron and axial motion.

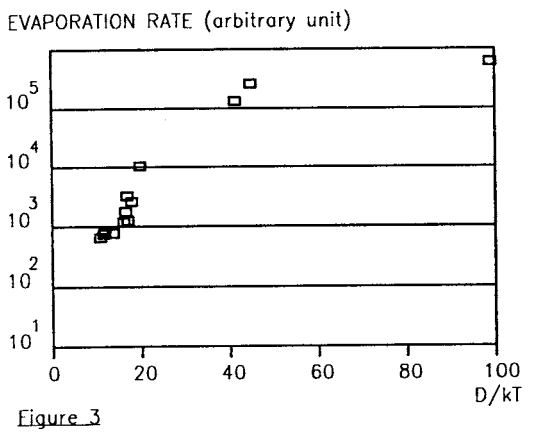


Figure 3

Penning trap. The evaporation rate is modified by varying the oven temperature.

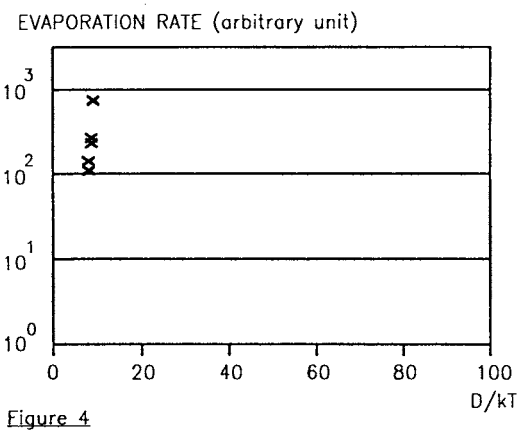


Figure 4

Paul trap. The evaporation rate is modified by varying the oven temperature.

Effect of the Asymmetries of the Field on the Evaporation Rate

Theoretical model

It is known (6) that electric fields asymmetries and misalignment of the magnetic field can produce a coupling between the axial and equatorial motions of the ions.When a resonant coupling exists (i.e. $\omega'_z = \ell\omega'_m$ (equ.1), ℓ integer, ω'_z and ω'_m the axial and

magnetron frequencies modified by space charge) the ion trajectories expand, leading to a rising of the temperature and a loss of ions which limits their density.

We introduce the two dimensionless parameters α and γ :

$$\alpha = \omega_p^2/\omega_z^2$$

where ω_p is the plasma frequency and ω_z the axial frequency (for one ion in the trap). α is proportional to the ion density.

$$\gamma = \frac{\Omega}{\sqrt{2}\omega_z}$$

where Ω is the cyclotron frequency. γ is proportional to the magnetic field.

Experimental observation for a continuous loading of the trap

The effect of the resonant coupling is clearly demonstrated in fig. 5, where the equilibrium value of the ion fluorescence (proportional to the ion density or α) is recorded versus the parameter γ. For this experiment, the trap was continuously loaded with ions and the magnetic field was swept very slowly from 0.58 T (γ = 1.05) to 1.53 T (γ = 2.75). The laser freqency was near the center of the optical transition and its intensity (< 7 μW) relatively weak in order to avoid cooling or heating effects. In fig. 5, the theoretical curves for different values of ℓ are solutions of equ.1 expressed in the plane α , γ. A more detailed analysis is given in (7). The good agreement between the experimental recording and the theoretical curves shows that, in the Penning trap, the resonant coupling is the principal reason of the ion loss, providing the creation rate is sufficient to reach this limitation. It has to be noticed also that the density cannot exceed a given value (α = 0.31, i.e.1.10^7 cm^{-3} for the density). For this value which is far below the limit imposed by space charge (α = 3), we observe a rapid and important loss of ions. This cannot be explained by our simple model, but could be the result of electrostatic modes allowed by the asymmetries of the fields and propagating through the cloud.

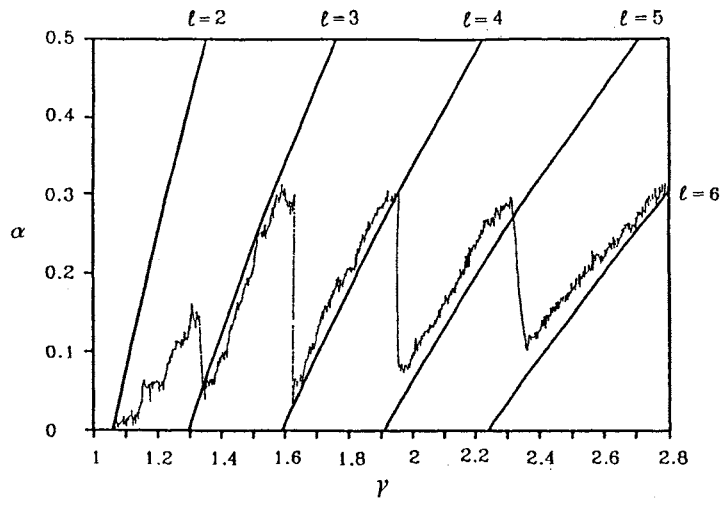


Figure 5
Experimental variation of the ion density in a Penning trap as a function of the parameter γ

Other evidence for resonant coupling can be observed in figs. 6 and 7, where we have plotted the previous measurements (see figs.1 and 3) in the temperature-density plane.In the case where the creation rate is modified by the electron current, Figure 6 shows that the temperature increases as the density comes near the value for which there is a resonance.When the creation rate is further increased, the measured density decreases, which might be due to a spreading of the magnetron motion resulting from collisions with the background gas. Indeed the background gas pressure may increase due to local outgassing and our fluorescence collecting system measures mainly the density near the center of the trap, therefore a deformation of the cloud can result in a change of the fluorescence.

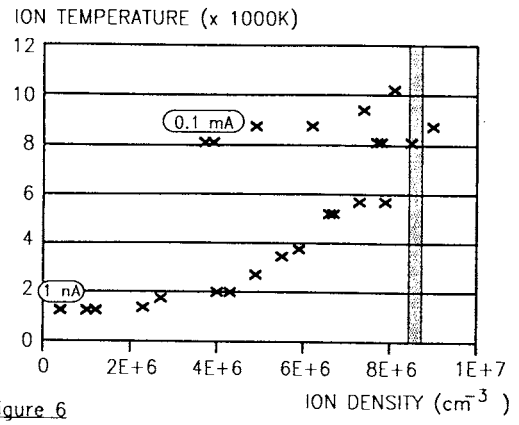


Figure 6

Penning trap. The ion density is modified by varying the electron gun temperature.
The vertical line shows the maximal value of the density due to asymmetry and the labels give the electronic current

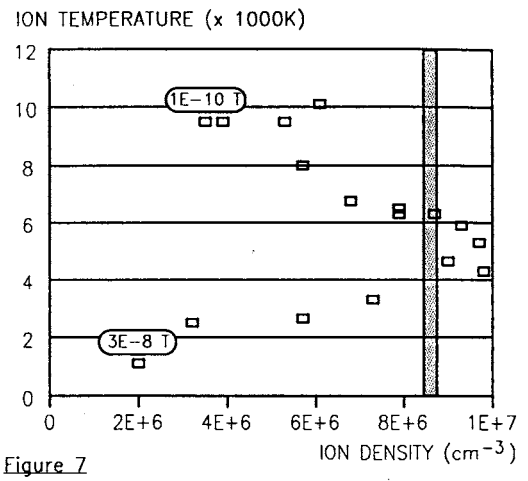


Figure 7

Penning trap. The ion density is modified by varying the oven temperature.
The vertical line shows the maximal value of the density due to asymmetry and the labels give the background pressure

In the case where the creation rate is modified by the oven temperature (and where the background pressure varies), the higher temperatures occur when the pressure is low. For the higher values of the creation rate, the ion density is limited by the collisions with the background gas, as explained in the preceding section.

Experimental observation for an isolated cloud

In order to get a more precise idea of the resonance coupling effects, we have reproduced the experiment of fig. 5, in the case where the creation rate equals zero. Figures 8 and 9 show the ion density and their corresponding temperature when the value of the

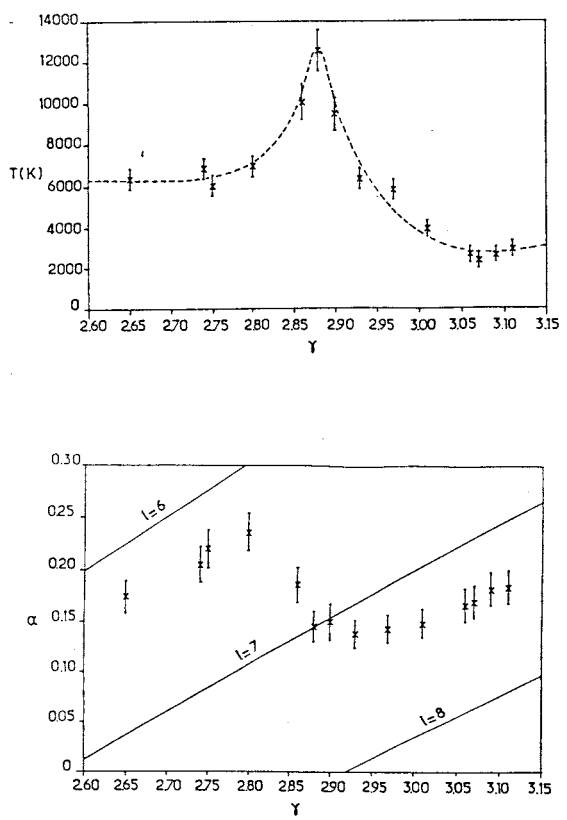


Figure 8

Density and temperature variation when the magnetic field is increased

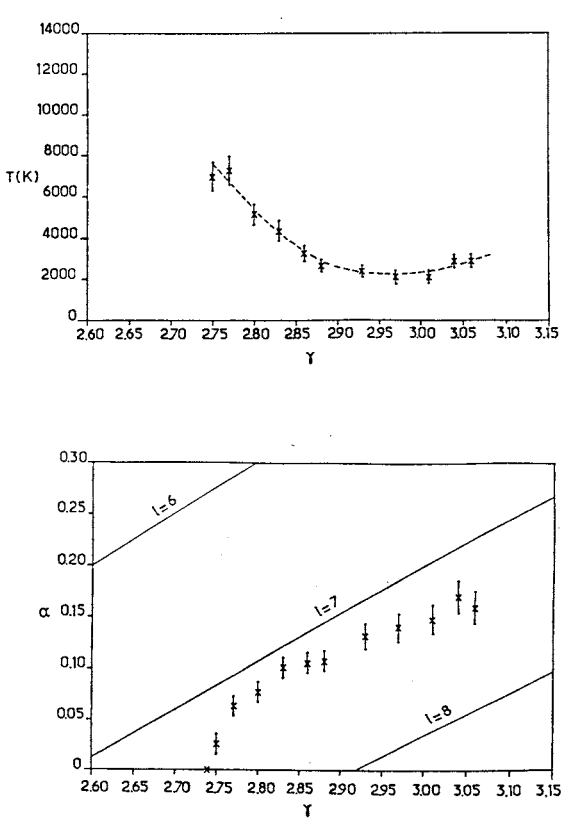


Figure 9

Density and temperature variation when the magnetic field is decreased

magnetic field is slowly increased and then decreased. When the magnetic field varies, so does the resonance condition and we can observe the behaviour of the ion cloud when this resonance condition is realized. The experimental procedure was the following : as soon as the density of the ion cloud had reached a sufficient value (below the resonance), the ion creation rate was set to 0. Then we swept the magnetic field and for each value corresponding to an experimental point, we recorded the ion absorption line by sweeping the laser frequency. The proportionality factor between the fluorescence signal and the value of α has been determined by assuming that the highest temperature in Fig. 8 indicates the resonance. We observe in Fig. 8 that the density increases when the conditions move away from the resonance showing a contraction of the cloud since there is no ion creation. The temperature increases strongly near the resonance (Figs. 8 and 9). In Fig. 9, γ is continuously decreased in order to keep the ions near the resonance condition : we observe the total loss of the ions and the continuous rise of the temperature.

Conclusion

We have studied the density, the temperature and the evaporation rate of ions confined in a Penning and a Paul trap. We have shown the main effect of the background gas pressure for large ion creation rate which reduces the density without temperature increase,whereas in the case of the Penning trap the resonant coupling due to the asymmetries of the fields leads to a temperature increase.

References

(1) H. G. Dehmelt, Adv. in At. and Mol. Phys. 3, 53, (1967), 5, 109, (1969)

(2) D. J. Wineland, J. J. Bollinger, W. M. Itano, J. D. Prestage, Opt. Soc of Am. B 2, 1721,(1985)

(3) See for instance : D. J. Wineland, J. C. Bergquist, W. M. Itano, J. J. Bollinger and C. H. Manney, Phys. Rev. Lett. 59, 2935 (1987) ; F. Diedrich, E. Peik, J. M. Chen, W. Quint and H. Walther, Phys. Rev. Lett. 59, 2931 (1987) ; J. Hoffnagle, R. G. De Voe, L. Reyna, J. Rosenkranz, and R. G. Brewer, Proc. of the 9th Intern. Conf. on Laser Spectr., June 1989, p.270

(4) R. G. De Voe, J. Hoffnagle and R. G. Brewer, Proc. of the 9th Intern. Conf. on Laser Spectr., June 1989, p.270

(5) F. Plumelle, M. Desaintfuscien, M. Jardino, P. Petit, Appl. Phys. B41, 183 (1986)

(6) J. Yu, F. Plumelle and M. Desaintfuscien , 4th Symp. on Frequency Standards and Metrology , Ancona,1988,p.449

(7) J. Yu, M. Desaintfuscien and F. Plumelle, Appl. Phys. B 48 ,51 (1989)

Resonators, Oscillators and Materials

ROM-02 SAW Devices

Chairman: G.F. Riha

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-02 Number 1

STATE OF THE ART OF SAW RESONATORS

Philippe DEFRANOULD, Louis PENAVAIRE, Serge CALISTI and Frederic RAYMOND

THOMSON SINTRA ASM/DTAS
399, route des Crêtes - BP 38 - Parc de Sophia-Antipolis
06561 VALBONNE CEDEX FRANCE

Abstract

The SAW resonators which appeared in the early seventie's, have been improved for years to become now good candidates to make low phase noise oscillators and narrow bandpass filters, both in the VHF-UHF range. After a brief history this paper describes the latest techniques in the design and manufacture of high performance SAW resonators realized on quartz. The emphasis is put on the all quartz package, first developped by Raytheon in the beginning of the eightie's. These devices, very easy to handle, show long term stability less than 1 ppm/year with unloaded Q-factor higher than 17000 at frequency close to 500 MHz. These characteristics lead to very high performance SAW oscillators : phase noise floor as low as -182 dBc/Hz has been achieved with special design. Such data has been obtained with accelerometric sensitivity in the 10^{-10}/g range so that on board equipments are now on development for military and space applications.

1.- INTRODUCTION AND HISTORY

SAW resonators are constituted of two distributed reflectors made of periodic discontinuities of the surface which are generally grooves or metallized strips. The electrical coupling with the resonant cavity is obtained with one or more frequently with two transducers placed between the reflectors.
Since 1970, date of the first reported work on SAW resonators, they have stired up the interest as they can cover the frequency range from 100 MHz up to more than 1 GHz, knowing that the upper limit of conventionnal bulk wave quartz resonators is between 100 and 200 MHz and that the lower one of dielectric resonators is around 1.5 GHz.

From 1970 to 1975, several theoretical studies and experimental works were done in order to establish a satisfying model for grooves or metallized strips, and to determine the basic possibilities and limitations of SAW resonant cavities. All these works were presented in a lot of papers, one can cite those of R.C.M. Li et al. concerning the model of grooves arrays, and of G.L. Matthaei et al. [3] giving a model of the electrical coupling with SAW resonant cavities.

Of course, in the same time than SAW resonators possibilities were shown (Q-factor of 20000 at 500 MHz, low sensitivity to high power level), difficulties appeared, the main of them and the more difficult to solve being the resonant frequency dispersion in the manufacturing process and also due to the mounting and encapsulation techniques.

In spite of these difficulties, from 1975 to 1982, a lot of companies developped successfully SAW resonators for several applications going from RF converter in video cassette recorder [4] (mass produced devices), to noise suppression filter for a satellite use [5], as well as oscillators for measuring equipments like spectrum analyzer [6]. They found how to reduce resonant frequency dispersion [7] and to avoid transducers degradations due to the high power level applied to the device.

But at the beginning of the eightie's, although SAW resonator was a suitable component for many applications, and though it was demonstrated that it allows to build low phase noise oscillators [8] because of its low sensitivity to high power levels [9], it was not used to make high quality sources, the reasons being the not enough accurate frequency adjustment and the generally poor and/or unreproducible aging characteristics of this device [10].

Then, the most interesting works of the last five years, which have been carried out by the Raytheon team working on SAW resonators and oscillators, are based on the use of the all quartz package (AQP) and have resulted in devices having low aging, low acceleration sensitivity, low flicker noise, and specially designed to admit high power level. Although the main application of these AQP resonators is the fabrication of very low phase noise sources needed by the next generation of military radar systems, they are also interesting for many applications because they can be mounted directly on an hybrid circuit without risk regarding to their electrical preformance and allowing miniaturization.

Now, after twenty years of life, the SAW resonator seems to have reached its maturity and is now used in applications ranging from high volume consumer garage door transmitters to low phase noise frequency sources for military radar systems. Then we are going to have a review of the latest works done and of the best performances which can be obtained.

2.- IMPROVEMENTS IN SAW RESONATORS ANALYSIS AND DESIGN

The models published 15 years ago and particularly that of transducers are simplified ones ; they were satisfying enough when the SAW resonator was at the beginning of its life.

Today, SAW resonators for low phase noise oscillators must have very large resonant cavity and wide aperture at the expense of additionnal longitudinal and transverse modes and increased ohmic losses (reducing the Q-factor). Then an accurate and reliable analysis program is necessary to do the right compromise between insertion loss, loaded Q-factor and resonant cavity dimensions.

Probably, most of the SAW manufacturer have developped such programs, but a few were reported. For instance, THOMSON-SINTRA DTAS presented at the last IEEE Ultrasonics Symposium the work done by its team [11].

Three analysis program modules were developed : the first one is for gratings, the second one to analyse transducers and the third one to compute the transfer function of the entire resonator.
Based on the grating theories of R.C.M. Li et al [2] and H.A. Haus et al [12], a grating analysis program was developed to compute the two-port S-matrix for even and odd transverse modes in a periodic groove and/or metallic strip grating structure.
The transducer analysis program analyses a general three phase periodic apodized transducer structure having 2N acoustic ports corresponding to N transverse modes. Modeled effects include electrostatics and acoustic coupling between electrical ports and acoustic mode interaction. The program generates the complete (2N+2)*(2N+2) matrix description for the transducer.
The third program module combines transducer and grating analysis with structure separations and acoustic port boundary conditions, outputing the resulting electrical port matrix of the complete resonator.
The resulting resonator analysis is then combined with external electrical networks using

existing network analysis software.

Comparisons were made between experimental results and computations, they showed a good agreement and are illustrated by the plots of the measured (fig. 1) and calculated (fig. 2) S21 of a resonator near 500 MHz, the higher curve in each figure being the amplitude and the lower one the group delay which is related to the loaded Q-factor by the formula :

$$Q_L = \pi * f_0 * \tau$$

In this case, only the first three symmetrical transverse modes have been taken into account.

These results show that with such programs, it is possible to design a resonator having the optimized characteristics for a given application.

3.- SAW RESONATOR PACKAGING AND MOUNTING

As soon as saw resonators began to be manufactured, a lot of efforts were concentrated on the mounting and encapsulation to be used for these devices because any saw manufacturer knows that they are key points of the fabrication.
At first, and when high performances in terms of frequency accuracy and q-factor are not required, standard mounting techniques for saw devices in dil or to packages and encapsulation under dry nitrogen can be used. Sometimes, because of the available facilities, manufacturers fixed and sealed saw resonators exactly like mass produced baw resonators [13].

But for applications needing good frequency setting, high q-factor and low aging rate, a mounting free of stresses, free of organic components inside the package and an encapsulation under high vacuum are required. Moreover a careful processing and manufacturing of the resonator and an outgassing before sealing must be done. Techniques issued from those employed for high quality baw quartz resonators were very often utilized : coldweld sealing under high vacuum after a several hours outgassing at high temperature (300°c) as described by W.R. Shreve [10]. More recently, a sealing at 370°c using an eutectic gold germanium solder preform to assembly a gold plated kovar lid on a metallized alumina base, was reported by R.C. Bray [14]. In this last example, the saw resonator is attached on the ceramic base with gold ribbons to avoid stresses.
By these methods, the q-factor is as high as possible regarding to the resonator design and to the quality of the material used, and the measured agings are of the order of 1 ppm/year after a pre-aging period.

But, in fact, probably the best and surely the most attractive solution appears to be the so-called "all quartz package" (AQP) illustrated by figure 3 : the base of the package is the substrate on which is etched the SAW resonator; the cover is a quartz plate having the same crystalline orientation than the resonator ; the two parts are assembled with a low temperature glass frit under vacuum.

Although the AQP concept is not new, it was mentionned the first time by M. Gilden et al. in 1978 [15], it has been successfully developped by the RAYTHEON team working on SAW resonators and oscillators and the first results were published in 1985 [16]. Since 1985, the same team have published a lot of papers on the impressive work they have done on SAW resonators and oscillators and all based on the AQP [17,18,19].
Interested in the results obtained by RAYTHEON, and convinced that the AQP is the best solution to the problems encountered with resonators, THOMSON-SINTRA DTAS has developed too this encapsulation and described it at the last European Frequency and Time Forum in

1989 [20].

We will see in the next paragraph what are the performances obtained recently on SAW resonators, AQP devices.

4.- PERFORMANCES OF AQP SAW RESONATORS

4.1.- Frequency adjustment

Before the AQP, a lot of frequency adjustment processes have been described. They are generally based on the addition or suppression of a material deposited inside the resonant cavity or on a selective etching changing the step height for the transducers, this results in a SAW velocity change and then in a frequency variation.
As these adjustments are made before sealing and as they are not accurate enough, the frequency accuracy which can be obtained is a few tens of ppm.
The method described by T.E. Parker et al. [19] and based on the AQP allows frequency trimming of more than 50 ppm with an accuracy better than ± 1 ppm. It consists in transporting a material from the cover to the resonant cavity by evaporating it with a laser (see figure 4). In addition to its accuracy, the advantage of this method is that the adjustment is done after sealing and eventually after a pre-aging period.

4.2.- Aging

All the reported aging test results obtained for AQP resonators, with low applied level (<10 dBm), show long term frequency variations less than 1 ppm/year and sometimes near 0.1 ppm/year [16,20], comparable to those obtained with other high quality encapsulation techniques like cold weld sealing.
In the case of low phase noise oscillators, resonators are submitted to a high incident power (~ +20 dBm) which generally degrades the aging characteristic. But the use of very large resonant cavities or/and dual cavities allows to obtain aging rate still less than 1 ppm/year [19].
It was also shown that laser frequency trimming of SAW resonators increases the aging rate but not more than 1 ppm/year.

4.3.- Acceleration sensitivity

The advantage of the AQP is that the mounting of the resonator can be done with any convenient material in order to reduce the acceleration sensitivity without any other requirement.
Raytheon have worked a lot on the subject and they have found that by choosing properly the dimensions of the AQP, it is possible to reduce the acceleration sensitivity of SAW resonators below $1.10^{-9}/g$ and to reach a few $10^{-10}/g$. But, these results are very dependant upon the dimensions of the AQP and its mounting.
More reproducible results have been obtained with the AQP dual resonator shown by figure 5 but this structure needs a complicated manufacturing process and a laser frequency trimming in order to obtain the same resonant frequency for the two resonators [19].

4.4.- Phase noise of SAW resonator oscillators

To build low phase noise oscillators, the resonator must have a low insertion loss, a high loaded Q-factor and the power in the oscillator loop must be as high as possible. As we have seen before, a compromise must be done between the resonant cavity dimensions which must be large and the electrical characteristics. This can be done by using accurate and reliable analysis program as described in §2.

Several works have been reported showing resonators near 500 MHz with an insertion loss lower than 6 dB, an unloaded Q-factor greater than 17000 and a power handling in excess of +20 dBm [20,21].

With such resonators, high spectral purity oscillators have been built and measured. Due to the characteristics of the SAW resonators and to a careful oscillator design and choice of components, very good results in term of phase noise can be obtained -133 dBc/Hz at 1 KHz from the carrier and a phase noise floor at -170 dBc/Hz as shown by figure 6 (for 2 identical oscillators with an output amplifier).

More impressive results have been obtained by Raytheon after a long work to improve SAW resonator flicker noise [22] and in applying incident power as high as +26 dBm : less than -140 dBc/Hz at 1 KHz from the carrier and a phase noise floor at -184 dBc/Hz (for one oscillator without buffer amplifier).

5.- CONCLUSIONS

After 20 years of work done on SAW resonators, it has been demonstrated that this device can have excellent electrical characteristics and mainly a high Q-factor (20000 at 500 MHz). By a proper design, it can be used with high incident power (~ +20 dBm) without prohibiting degradation of its aging which is generally less than 1 ppm/year.
The use of the All quartz Package is one of the packaging techniques which guarantees good and reproducible aging characteristics but it also allows a laser frequency adjustment with an accuracy better than 1 ppm, and mounting having low acceleration sensitivity (a few 10^{-10}/g).

Acknowledgements

As the authors have themselves worked for years on SAW resonators, they appreciate the efforts made by all people who have worked on this subject and they would like to acknowledge them and especially those named in the bibliography hereafter.

* This work was partly supported by D.R.E.T.

BIBLIOGRAPHY

[1] E.A. ASH - "Surface wave grating reflectors and resonators" - Digest of 1970 IEEE Symposium on Microwave Theory and Techniques.

[2] R.C.M. LI, J.A. ALUSOW and R.C. WILLIAMSON - "Surface wave resonators using grooved reflectors" - 1975 Frequency Control Symposium Proceedings.

[3] G.L. MATTHAEI, F. BARMAN, E.B. SAVAGE and B.P. O'SHAUGHNESSY - "A study of the properties and potential application of acoustic surface wave resonators" - 1975 IEEE Ultrasonics Symposium Proceedings.

[4] Y. EBATA, K. SATO and S. MORISHITA - "A LiTaO3 SAW resonator and its application to video cassette recorder" - 1981 IEEE Ultrasonics Symposium Proceedings.

[5] W.J. TANSKI, M. BLOCH and A. VULCAN - "High performance SAW resonator filters for satellite use" - 1980 IEEE Ultrasonics Symposium Proceedings.

[6] S. ELLIOTT, M. MIERZWINSKI and P. PLANTING - "The production of surface acoustic wave resonators" - 1981 IEEE Ultrasonics Symposium Proceedings.

[7] W.E. BULST and E. WILLIBALD - "Ultrareproducible SAW resonator production" - 36th Annual Frequency Control Symposium Proceedings - 1982.

[8] L. PENAVAIRE, D. SEGUIGNES, C. LARDAT, J.J. BONNIER, J.Y. CHEVALIER, Y. BESSON - "A 120 MHz SAW resonator stablized oscillator with high spectral purity" - 1980 IEEE Ultrasonics Symposium Proceedings.

[9] M. PLANAT, J.J. GAGNEPAIN, C. LARDAT, L. PENAVAIRE - "Non linear characteristics of SAW grooved resonators" - 1980 IEEE Ultrasonics Symposium Proceedings.

[10] W.R. SHREVE - "Active aging of SAW resonators" - 1980 IEEE Ultrasonics Symposium Proceedings.

[11] L. PENAVAIRE, S. CALISTI, M. SOLAL AND P. DUFILIE - "State of the art SAW resonators in France" - 1989 IEEE Ultrasonics Symposium Proceedings (to be published).

[12] H.A. HAUS and R.V. SCHMIDT - "Modes in SAW grating resonators" - Electronic Letters, Vol. 13, nø1, 6 jan. 1977, pp. 12-13.

[13] E.J. STAPLES and R.C. SMYTHE - "Surface acoustic wave resonators on ST-quartz" - 1975 IEEE Ultrasonics Symposium Proceedings.

[14] R.C. BRAY - "Applications of SAW resonators in high performance instrumentation" - IEEE Transactions on Ultrasonics, ferroelectrics, and frequency control, vol. 35, n°3, may 1988.

[15] M. GILDEN, G.K. MONTRESS, D.E. CULLEN, R.A. WAGNER and T.M. REEDER - "Design of stable SAW oscillators operating above 1 GHz" - 1978 IEEE Ultrasonics Symposium Proceedings.

[16] T.E. PARKER, J. CALLERAME and G.K. MONTRESS - "A new all quartz package for SAW devices" - Proceedings of the 39th Annual Frequency Control Symposium - 1985.

[17] G.K. MONTRESS, T.E. PARKER and J.CALLERAME - "A miniature hybrid circuit SAW oscillator using an all quartz packaged resonator" - 1985 IEEE Ultrasonics Symposium Proceedings.

[18] J.A. GREER and T.E. PARKER - "Improved vibration sensitivity of the all quartz package surface acoustic wave resonator".

[19] T.E. PARKER, and G.K. MONTRESS - "Frequency stability of high performance SAW oscillators" - 1989 IEEE Ultrasonics Symposium Proceedings (to be published).

[20] S. CALISTI, Ph. DEFRANOULD, L. PENAVAIRE, D. HAUDEN, and A. D'ALMEIDA - "All quartz packaged SAW resonator" - 3rd European Frequency and Time Forum - 1989.

[21] G.K. MONTRESS, T.E. PARKER and M.J. LOBODA - "Extremely low phase noise SAW oscillator design and performance" - 1987 IEEE Ultrasonics Symposium Proceedings.

[22] M.J. LOBODA, T.E. PARKER, J.A. GREER and G.K. MONTRESS - "Reduction of close to carrier phase noise in surface acoustic wave resonators" - 1987 IEEE Ultrasonics Symposium Proceedings.

FIGURE 1 : Measured S21 of a FRH 901 resonator

FIGURE 2 : Calculated S21 of a FRH 901 resonator

FIGURE 3 : Schematic representation of an all quartz packaged SAW resonator

FIGURE 4 : Method for the frequency trimming of an AQP using laser radiation.
From "Frequency stability of high performance SAW oscillators" T.E. PARKER, G.K. MONTRESS [19]

FIGURE 5 : All Quartz Packadged dual resonator
a) cross section
b) top view
From "Improved vibration sensitivity of the all quartz package surface acoustic wave resonator" J.A. GREER and T.E. PARKER [18]

FIGURE 6 : Measured phase noise spectrum of 2 oscillators with output amplifier. The center frequency of the resonators are near 500 MHz.

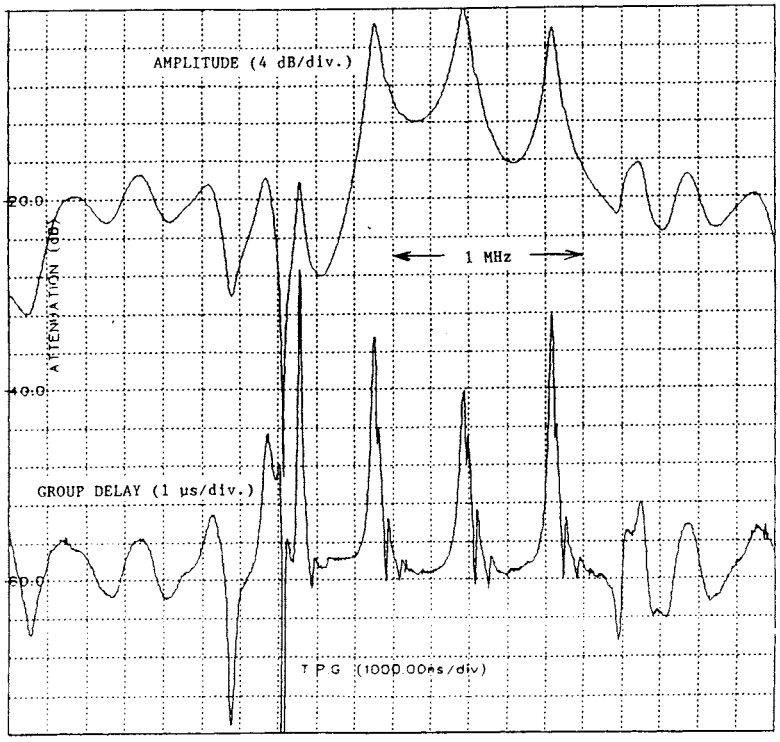


FIGURE 1 : Measured S21 of a FRH 901 resonator

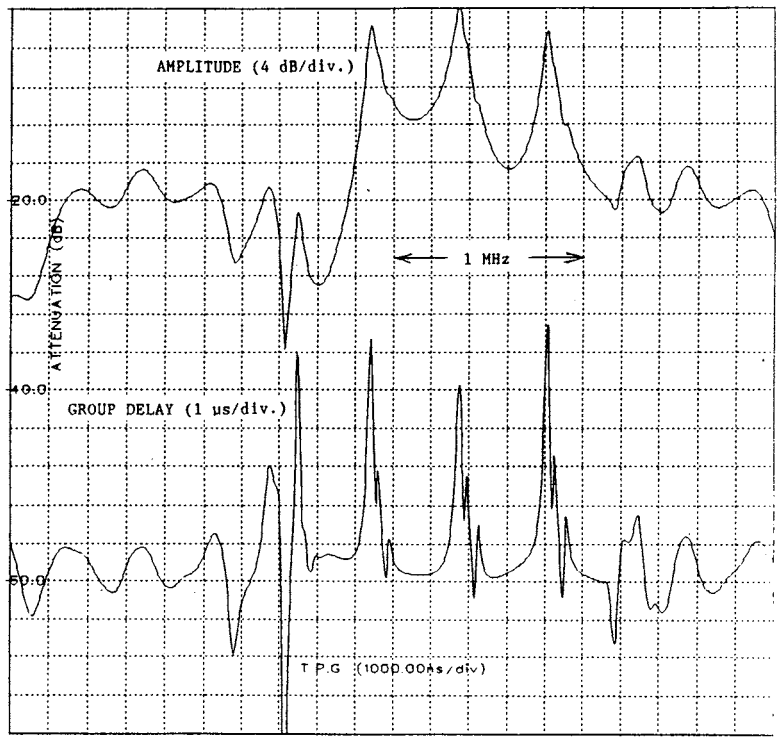


FIGURE 2 : Calculated S21 of a FRH 901 resonator

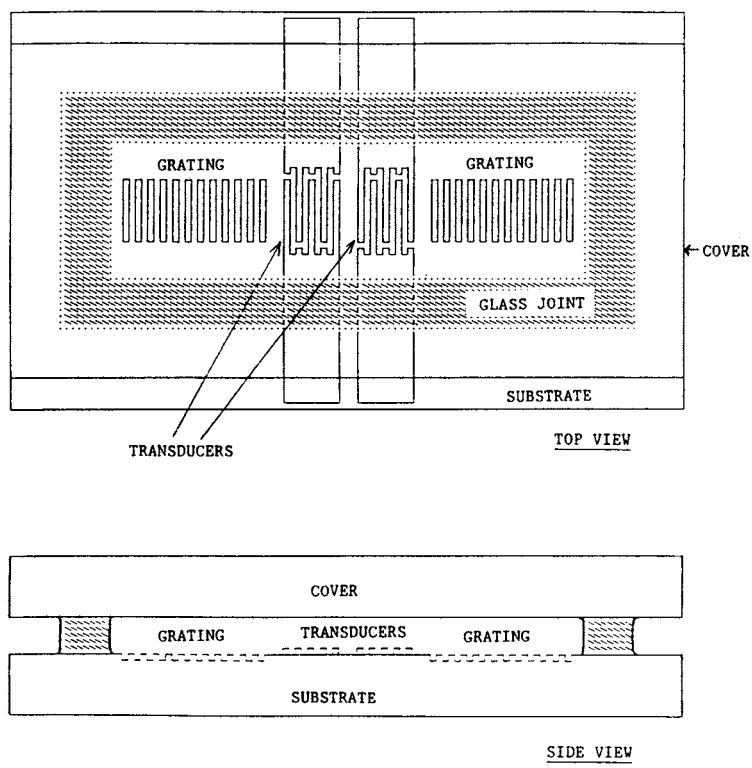


FIGURE 3 : Schematic representation of an all quartz packaged SAW resonator

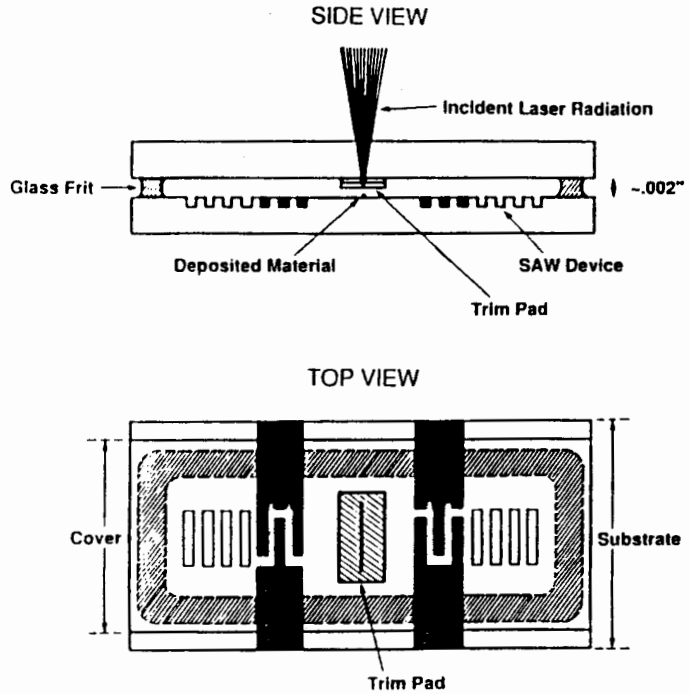


FIGURE 4 : Method for the frequency trimming of an AQP using laser radiation.
From "Frequency stability of high performance SAW oscillators" T.E. PARKER, G.K. MONTRESS [19]

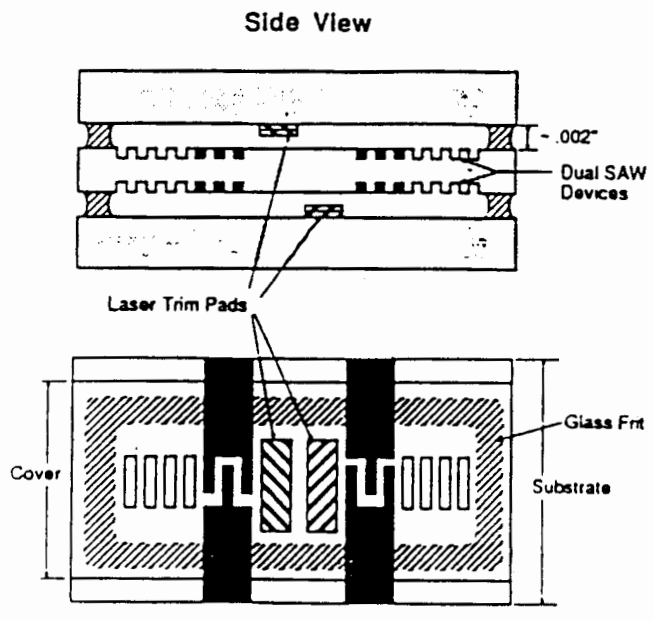


FIGURE 5 : All Quartz Packadged dual resonator
a) cross section
b) top view
From "Improved vibration sensitivity of the all quartz package surface acoustic wave resonator" J.A. GREER and T.E. PARKER [18]

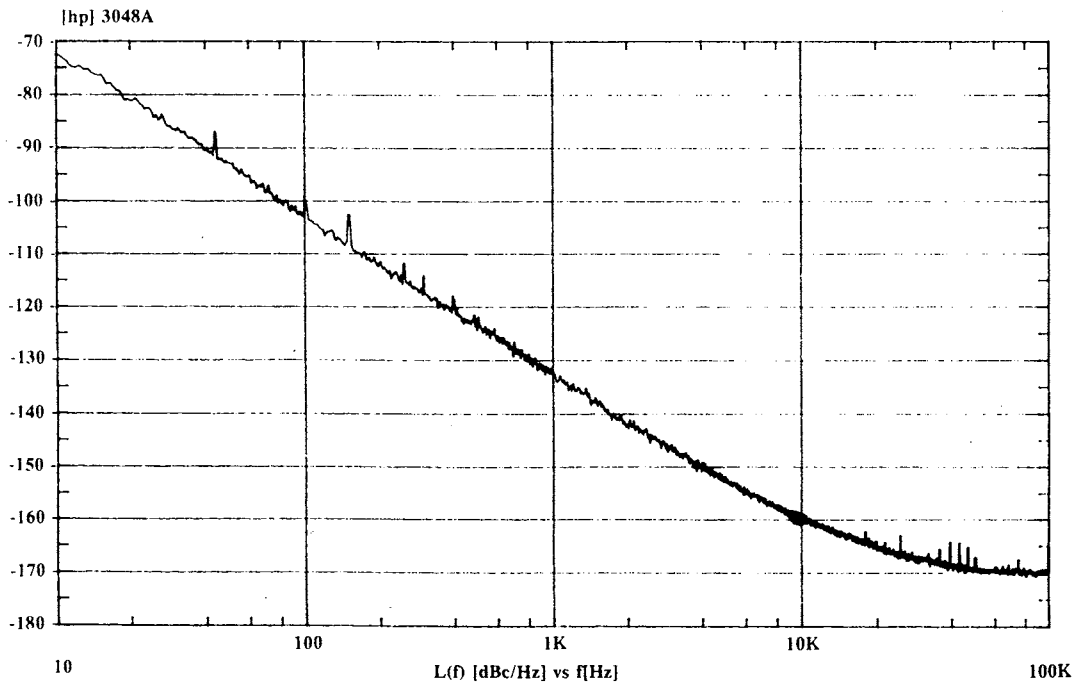


FIGURE 6 : Measured phase noise spectrum of 2 oscillators with output amplifier. The center frequency of the resonators are near 500 MHz.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-02 Number 2

SAW LOW-LOSS FILTERS FOR GHz-APPLICATIONS

K. Anemogiannis, S. Berek

Siemens AG., Corporate Research Laboratories
Otto-Hahn-Ring 6
D-8000 Munich 83, FRG

ABSTRACT

Low-Loss high-selection filters have become key components in modern radio communication systems. The establishment of numerous new system concepts and the demand of miniaturized, high-performance components has led to a challenge for SAW filter-technology toward to new applications in the RF-stages of GHz receivers and transmitters. Conventional SAW components suffer from high insertion loss and operation frequencies below 1 GHz. This paper introduces new low-loss filter techniques suitable to fulfill the requirements of low-loss operation in the GHz-range as well as high stopband rejection and steep passband skirts. A specially developed sub-micron technology using optimized photolithography and projection printing guarantees high volume device production with a high degree of dimensional precision. Realized filters provide a minimum insertion loss of 6 dB at a center frequency of 1.7 GHz, a bandwidth of 60 MHz and a stopband rejection of more than 45 dB. The filter-chip measures only 1.9x1.9 mm^2. Thus, a novel application area opens for mass-producible, high-performance low-loss SAW filters.

I. INTRODUCTION

The recently developed and introduced radio systems in the low GHz-range, e.g. global positioning system GPS, digital european cordless telephone DECT as well as future mobile radio communication networks are characterized by high complexity both in the analog system parts and the digital signal processing. In the analog system-parts high-performance active and passive components in miniaturized form are required. Table 1 summarizes the system requirements for RF filter components.

In the last decade the microstructure-based SAW technology has led to numerous applications of miniaturized, highly-selective filters providing unprecedented precision and unmatched reproducibility. However, the high insertion loss of 15 to 25 dB and maximum operation frequencies of 1 GHz preclude applications in GHz RF-stages. Some filter techniques employing higher spatial harmonics allow center frequencies up to some GHz /1, 2/. However, these techniques lead to high insertion loss of more than 20 dB.

Requirement	Reason	
1. low insertion loss	S/N ratio power consumption	(Rx) (Tx)
2. high selectivity	preselection image frequency suppression noise suppression rejection of spurious mixing signals	(Rx) (Rx) (Tx) (Tx)
3. miniaturization	volume and weight reduction	
4. low temperature coefficient, reproducibility, and reliability	system performance	
5. low cost	system cost	

Table 1 System requirements for RF filter applications (Rx): Receiver part, (Tx): Transmitter part

In order to increase the fundamental operation frequency electron beam lithography /3/ and X-ray lithography /4/ can be advantageously used, but extremely slow exposure speed and process complexity preclude a low cost fabrication.

This paper presents a low-loss configuration satisfying

the above requirements suitable for mass-production. Section II describes the basic mechanisms and design considerations of these devices. Sections III and IV deal with the appropriate piezoelectric crystals as substrate materials and the highly reproducible sub-micron fabrication technology respectively. Finally, section V presents the experimental results of realized filters.

II. FILTER CONFIGURATION

Owing to its three-port nature, the interdigital transducer (IDT) cannot be simultaneously matched at all its three ports. In conventional SAW components using two IDTs the electrical ports are mismatched minimizing acoustic regeneration and hence spurious triple transit signals. Low-loss operation implies electrical matching causing a strong triple transit signal to appear at the filter output. A perfect filter matching, for instance, leads to a passband ripple of about 4.5 dB, which is in general insufficient for an effective usage. In this case the filter provides an insertion loss of 6 dB due to the bidirectionality loss.

In /5/ *Lewis* introduced a structure providing low insertion loss by cascading IDTs connected alternately with the filter input and the filter output as sketched in

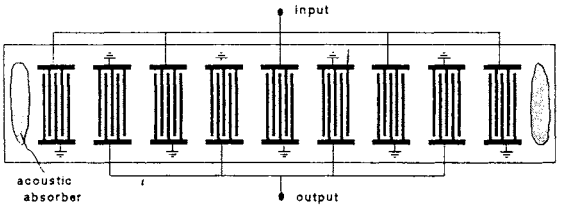


Figure 1 Interdigitated interdigital transducer filter

Fig. 1. Evidently, bidirectionality loss can be mainly avoided by using a high degree of cascading. By a slight impedance mismatch, however, the regenerated wave parts lead to echo signals and high passband ripple. In order to overcome these shortcomings *Hikita* presented in /6/ an extended structure by using two acoustic tracks instead of one (Fig. 2). The track coupling occurs by electrically connected IDT-pairs consisting of two identical IDT each placed in one track. For an appropriate dimensioning of the transducer pattern a self-matching occurs leading to a tight track coupling /7/. Due to the image impedance connection of the two IDTs the matching condition is fulfilled in the frequency region, where the imaginary part of the transducer impedance vanishes. In order to achieve this condition the number of electrodes per IDT is given by

$$N_c \simeq \frac{\pi}{2\ o(s)\ k^2},$$

where k^2 is the electroacoustic coupling coefficient of the substrate material and o(s) is a factor depending on the electrical period of the electrodes and the spatial operation harmonic. For fundamental operation and a periodic electrode connection sequence, o(s) is equal to 0.718 for solid electrodes and 0.275 for split electrodes. In the region of the IDT-passband the imaginary part of the impedance is sufficiently small, allowing an IDT-matching and leading to a tight coupling between the tracks. By taking advantage of internal transducer reflection due to electrical and mechanical surface loading, sharp passband skirts of the track coupling response can be obtained /7/. Since it is proportional to

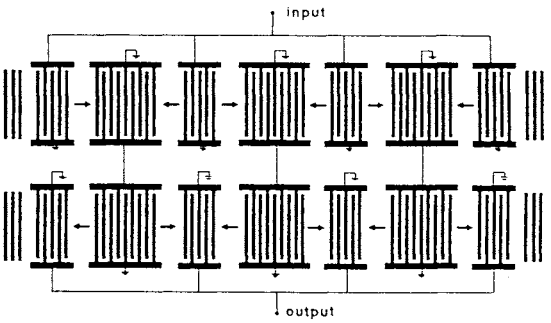


Figure 2 Dual-track low-loss filter

the inverse number of electrodes, the *IDT-bandwidth* and hence the track-coupling increase with an increasing coupling coefficient k^2. Therefore, crystals with high coupling are appropriate for RF applications with fractional bandwidths of a few percent.

For the symmetrical feeding of the coupling-IDTs short uniform IDTs are used, which also determine the filter impedance. Some reflective strips at the track-ends reduce further bidirectionality loss. By means of external elements, in general inductivities, the filter input and output can be matched to the load. Moreover, the feeding IDTs have a broadband characteristic with respect to the track-coupling response. Once matched to the load, they show a weak sensitivity to the external matching condition. Consequently, an effective suppression of the triple transit signal is possible.

Ideally the configuration in Fig. 2 provides a low insertion loss of about 1 dB depending on the number of cascading stages. In practice, however, some second order effects lead to additional losses and increase the insertion loss. As the most of these mechanisms increase at higher frequency, they affect the filter operation at the GHz range. Table 2 summarizes the most important loss

mechanisms and the consequences for the filter design.

Loss mechanism	Consequence
propagation	low surface roughness short propagation paths
ohmic loss	thick metallization
bulk mode excitation	small bandwidth
bulk mode conversion	fundamental operation thin metallization small bandwidth uniform surface perturbation (→ uniform IDT)
end effects	thin metallization uniform surface perturbation
diffraction	high aperture short propagation paths
acoustic mismatch	rectangle radiation profile
tuning loss	simple tuning network

Table 2 SAW filters: loss mechanisms

III. SUBSTRATE

Regarding Table 2, there is a demand of fundamental operation to avoid mode conversion effects. For constant electrode-width the operation frequency increases with the SAW-velocity. A second important substrate parameter is the coupling coefficient k^2 in order to obtain broad bandwidth /7/. Finally, according to Table 1 the temperature coefficient of the SAW-velocity is critical. It should be noted, that the temperature behaviour of the filter has to be compensated in the design by broadening the filter bandwidth. However, this affects the filter selectivity. Table 3 shows the SAW-velocity, the coupling coefficient, and the temperature coefficient of commonly used crystal cuts. Of these materials, quartz is outstanding due to its temperature response. A disadvantage of quartz is, however, the low electroacoustic coupling, which is insufficient for the track-coupling in dual-track filters.

Table 3 demonstrates that there is a compromise between coupling coefficient and temperature behaviour for 36YX-$LiTaO_3$. Another advantage of this cut is the high SAW-velocity sufficient for GHz applications. The wave mode is a leaky surface wave with low propagation loss of theoretically $5 \cdot 10^{-4}$ dB/λ and low sensitivity from the surface roughness. Since this mode is not a true surface wave, energy trapping techniques must be applied. Because of the high coupling coefficient

crustal cut	veloc. [m/s]	k^2 [%]	veloc. temp. coeff. [ppm/K]
STX-Quartz	3158	0.16	0(1)
YZ-$LiNbO_3$	3488	4.5	-94
128YX-$LiNbO_3$	3980	5.5	-75
X112Y-$LiTaO_3$	3300	0.75	-18
36YX-$LiTaO_3$	4216	6.6	-30

Table 3: SAW materials

(1): parabolic temperature dependence

metallized propagation paths are sufficient to support a guided mode. Using this technique a total propagation loss of lower than 10^{-3} dB/λ has been measured.

IV. FABRICATION

The requirement of fundamental operation of low-loss GHz SAW devices leads to both high velocity substrate and sub-micron electrode structures. Therefore, much attention is devoted to the fabrication process of these devices. In the past, electron beam photolithography has been proposed with a resolution down to 0.1 microns. However, extremely low speed, electrical charging of the substrate, and process complexity preclude an efficient low cost mass-production. Exploiting on optimized photolithography process based on optical 10:1 projection printing enables high positioning accuracy and low defect rates. Furthermore, in order to assure a high pattern reproducibility of the sub-micron electrodes the lift off technique has been adopted. In contrast to the various etching techniques, the lift off process guarantees tight dimensional control and does not affect the crystal surface and aging because of the noncorrosive organic solvents being used. The SAW patterning process is illustrated in Fig. 3.

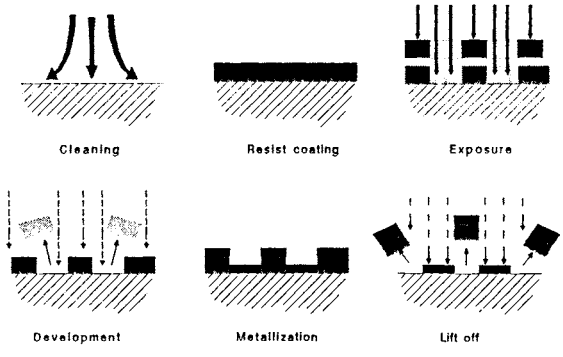


Figure 3 SAW patterning process

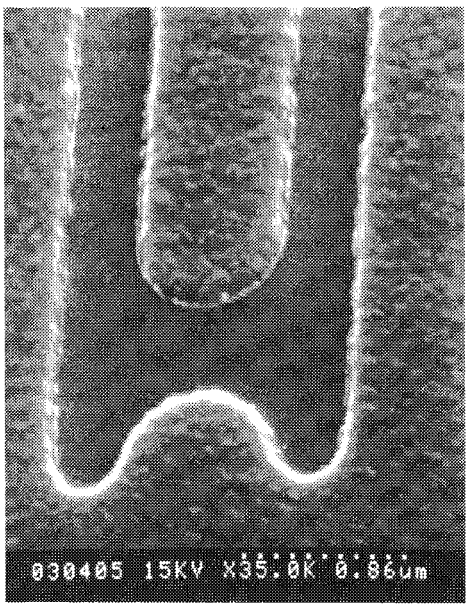


Figure 4 SEM picture of 0.6 micron electrodes

The technological concept concentrates on employing a fully computer-controlled fabrication process to obtain a high degree of reproducibility. In order to achieve the adequate resolution, the 10X reticle is exposed by an electron beam with an absolute accuracy of less than 0.25 microns over the whole reticle field. The wafer exposure is effected by an i-line wafer-stepper exhibiting line-widths down to 0.4 microns corresponding to operation frequencies up to 2.5 GHz on 36YX-$LiTaO_3$. Photoresist- and developing-process have been optimized resulting in one third of the exposure time compared to conventional g-line exposure. After developing, the aluminum for producing the metal pattern is deposited by electron-beam evaporation assuring thickness variations of less than 1%. The subsequent lift off of the remained photoresist does not affect the pattern dimensions. Fig. 4 shows a SEM picture of a 0.6 micron electrode of realized low-loss devices with excellent edge quality.

V. EXPERIMENTAL RESULTS

Fig. 5 shows the layout pattern of a realized GHz filter based on the dual-track structure described in section II. The structure consists of five input IDTs, five output IDTs, and four track-coupling IDT-pairs. The substrate material is 36YX-$LiTaO_3$. It should be noted that special attention has been paid to the layout design in order to obtain optimum transducer interconnections and low ohmic effects. The chip has a size of 1.9x1.9 mm^2 fitting into a specially developed TO-39 package with 8 pins

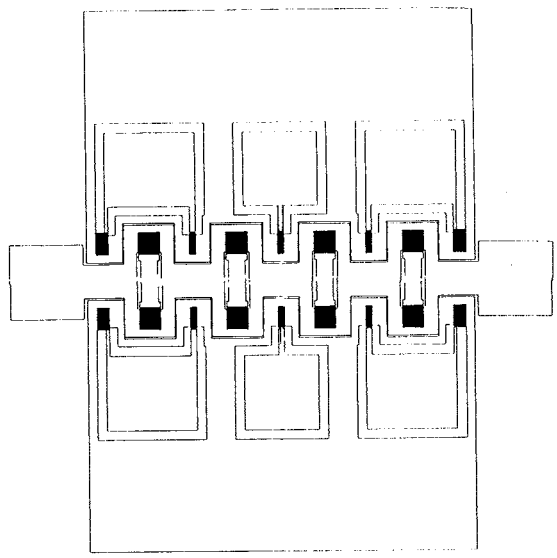


Figure 5 Filter layout

providing an excellent performance with respect to electromagnetic feedthrough suppression.

Fig. 6 shows the measured frequency response. The filter has a center frequency of 1.684 GHz and a 3 dB bandwidth of 59 MHz corresponding to a fractional bandwidth of 3.5%. The minimum insertion loss is 6 dB and the group delay at center frequency is 53 ns. The

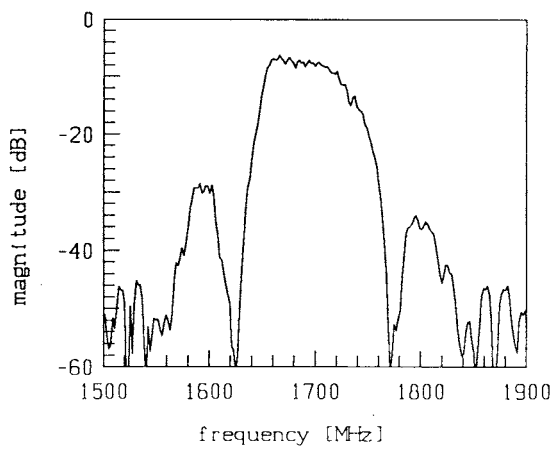


Figure 6 Filter response

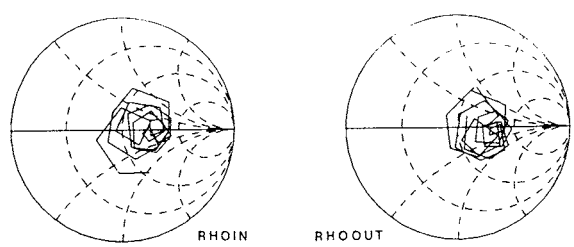


Figure 7 Filter impedance in the region 1.66 GHz to 1.74 GHz

passband ripple is ± 1.5 dB and ± 15ns respectively. The stopband rejection is better than 45 dB in the region 100 MHz away from the center frequency. Fig. 7 shows the filter impedance in the passband. For matching to the load impedance a shunt inductance of 6 nH at the input and output ports has been used.

The remaining loss in the passband occurs primarily due to ohmic losses and spurious bulk mode excitation. The bulk wave radiation losses are typical for 36YX–$LiTaO_3$ substrates which show a small difference between the velocities of the used leaky surface wave and the corresponding bulk wave.

VI. CONCLUSION

Novel SAW filters have been designed and fabricated for GHz applications. The dual-track technique has been applied with the advantage of local separated structure parts responsible for low-loss operation and high selectivity. For fabrication novel optimized techniques allow a highly reproducible mass-production of the electrode pattern with line widths down to 0.4 microns. The measured data exceed results appearing to date in the literature leading to numerous new applications of the SAW technology in the GHz range.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. G. Riha and W. Ruile for helpful discussions and design support, respectively. Furthermore, special thanks go to B. Bienert, A. Lenk and H. Zottl for the assistance of the device fabrication.

REFERENCES

/1/ Fleischmann B., Ruile W., Riha G., "Rayleigh–Mode SAW–Filters on Quartz for Timing Recovery at Frequencies above 1 GHz", IEEE Ultrasonics Symposium Proc. 1986, pp.163–168

/2/ Baer R.L., Flory C.A., "Harmonic Operation of STW Filters", IEEE Ultrasonics Symposium Proc. 1988, pp. 53–56

/3/ Yamanouchi K., Cho Y., Meguro T., "SHF–Range Surface Acoustic Wave Interdigital Transducers Using Electron Beam Exposure", IEEE Ultrasonics Symposium Proc. 1988, pp. 115–118

/4/ Matthews H., *Surface Wave Filters Design, Construction and Use*, Wiley 1977, Chapter 4

/5/ Lewis M., "SAW Filters Employing Interdigitated Interdigital Transducers, IIDT", IEEE Ultrasonics Symposium Proc. 1983, pp. 1–17

/6/ Hikita M., Tabuchi T., Kojima H., Sumioka A., Nakagoshi A., Kinoshita Y., ""High Performance SAW Filters with Several New Technologies for Cellular Radio", IEEE Ultrasonics Symposium Proc. 1984, pp. 82–92

/7/ Anemogiannis K., Müller F., Ruile W., Riha G., "High Performance Low–Loss SAW Filters for Mobile Radio with Improved Stopband Rejection", IEEE Ultrasonics Symposium Proc. 1988, pp. 77–82

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-02 Number 3

Coupled SAW resonators - Theory and Experiment.

P. Durrant *, J. Heighway and D.P.Morgan

Roditi International Corporation Limited,
130 Regent Street,
London W1R 6BR

* MEDL - Microwave Division,
Doddington Road,
Lincoln, England.

Abstract

The advent of high frequency SAW resonators has allowed the implementation of high performance bandpass filters for use in cost sensitive applications. If the SAW resonator is considered as a standard building block existing bulk wave design techniques can be combined with SAW analyses and used directly to implement filter solutions. Straightforward practical design considerations, using only single level metalisation are discussed and their effectiveness demonstrated by implementing an ST-cut quartz 4-pole coupled SAW resonator at 466 MHz.

Introduction

The Surface Acoustic Wave (SAW) resonator was invented by Prof. Sir Eric Ash in 1970 whilst spending a sabbatical year at the IBM T.J. Watson research centre (ref 1). Prof. Ash reported on the reflective properties of SAW's when passing through grating reflecting structures. Although the electrical performance demonstrated at the time was not impressive by today's standards, Q values as high as 80 were obtained!, it was clear that the technique had potential. This potential, particularly for filtering where the resonators were used as the building blocks, was discussed by Ash (ref 2).

This initial work was extended by Staples and Smythe (ref.3) who introduced the two basic methods of coupling resonators; in-line and transverse. Their examples are schematically shown in Figure 1 (taken from figure 6 of ref.3), and since that time a number of authors (refs. 4,5,6) have extended the concept towards the implementation of so-called multipole SAW resonator filters. The advantages of such an implementation are threefold:

1. Well documented resonant element filter design procedures can be used to obtain the desired passband characteristics
2. SAW resonators can be used in order to implement these designs at frequencies of up to a few gigahertz
3. Simpler and cheaper solutions to system design problems will be facilitated.

This paper discusses the first stages of a programme aimed at facilitating the design and manufacture of such coupled SAW resonators and uses devices at 466 MHz to illustrate the techniques.

Basic Considerations

The application of SAW resonators to the mobile communications market area places requirements on the performance which has to be achieved. In the specific area chosen, the performance should be such as to meet a simple template; the principal requirements being low insertion loss and high rejection of the image frequency (55dB at 910 KHz from the carrier) . The system designer would like the rejection of signals at 227 KHz from the carrier to be $>$ 15dB, as an additional advantage.

The relatively high sidelobe level that is required, limits the choice of coupling mechanism somewhat because the in-line approach tends to give a resonant performance that is overlaid on the sin x/x transducer passband (see figure 6 ref.7) which can have an impact on the achievable sidelobe rejection particularly close to the carrier. For this reason, we have chosen to opt for the transversal coupling as depicted in figure 1b. The rejection requirements plus the shape factor implications imposed by the 910KHz and 227 KHz levels, also force a decision regarding the minimum number of poles; in this case the tables

show that a 4 pole design should be sufficient and ultimate rejection will be determined by electrical breakthrough.

Further constraints are introduced by the requirement that the coupled resonators be mass produced at low cost. This imposes several restrictions all of which impact directly on the design:

1. Multilayer metallisation is not possible
2. Etched groove technology, either for the reflective strips or for in-laid transducers is not possible
3. The devices must be monolithic (the assembly of multiple chips, with or without separate packaging, imposes potentially disastrous yield and rework problems onto the manufacturing process)
4. Single finger structures must be used in order to ensure adequate yield at high operating frequencies
5. The package must be hermetic and have excellent electrical isolation as well as being cheap
6. The SAW design should be insensitive to metallisation defects and hence should avoid intricate overlap profiles
7. Optimum use of substrate area must be made and chip sizes kept to a minimum.

It is worth expanding on this last point. The economics of volume production are complicated by the details of the costing structure; the allocation of overhead can be somewhat arbitrary. On first examination it would seem that minimum cost is obtained by minimising the chip size, and this is true if yield is not affected. In practice smaller chips, particularly narrower ones, not only make the design problem more difficult leading to close tolerancing on the electrical performance but also make manufacturing very much more difficult. The real need is to maximise the number of good devices per wafer and this requires the optimisation of the product of yield and chips per slice which can only be maximised for a specific and established production process. This is mainly a matter of experience which for conventional photolithographic operations implies a maximum figure of around 1000 chips for a 3" round wafer.

Finally the environmental considerations, in particular the temperature excursions that the filter will see in a mobile environment, impose the choice of ST quartz as the substrate material.

Thus the basic design criteria are set:

Single layer metallisation
Transverse coupling
4 - pole design (implying 4 resonators)
$\lambda/4$ geometry
ST - cut quartz
Metal, weldable package
Chip size approximately 5 square mm

Elemental design considerations

(a) Transducer

Conventional SAW resonators comprise either two port or one port designs; using either two or one transducer The use of resonators as building blocks and the decision to use transversal coupling enables the use of single port units and all designs considered here are single port type. In addition, the requirement that the coupled resonator be manufacturable at low unit cost means that all the resonator building blocks should have equal, identical spacing (to facilitate photolithographic control) and ideally, this spacing should be the same as in the reflecting array. Unfortunately, the choice of a synchronous placement of the transducer introduces a highly assymmetric resonant characteristic with a low side spurious response of 13dB rejection and a high side null. For filter design, a more symmetric response is advantageous and hence the $\lambda/8$ offset criteria has been adopted in our design (ref.8). This decision was made despite the concern that this choice makes the resonator response more sensitive to metal thickness because of changes in the acoustic reflectivity affecting both the centre frequency and the shape of the response. (In practice, we shall see that extreme control of metal thickness is required anyway to meet the frequency setting accuracy so in hindsight the decision was valid). There is one further complication that could occur, affecting the

choice of placement. In all designs the transducer is assumed to be reflectionless. We shall see that this is not entirely valid but to date the effect of reflections on the choice of placement has not been assessed. The impedance of the application circuit impacts directly on the decision for the overlap of the transducer as does the need for coupling between the adjacent resonators. In this case the circuit impedance is approx. 220 Ω implying a small transducer overlap. In addition the need to eliminate spurious transverse modes imposes a bias towards a low overlap design and in concert with the work of Tanaka (ref. 6) overlaps of around 8 wavelengths have been used. This figure yields a low static capacitance for small numbers of pairs and for this reason as well as to ensure acceptable loss (see below) the number of finger pairs has been increased. The final design trial used 300 pairs which yielded a static capacitance of approimately 3pF and $|Z|$ in excess of 200 Ω. Before arriving at a firm decision with regard to the number of pairs it is instructive to assess the impact of spurious modes. The most immediate effect of using transducers with a large number of finger pairs is an increase in cavity length. This is normally considered to be undesirable. In our application, we can calculate the likely impact through a knowledge of the reflector type (shorted or open circuited reflectors) and the metal thickness and type. The metal will be Aluminium and initially we have chosen connected reflectors on ST - cut quartz. The thickness directly determines the reflectivity and ideally this should be as thin as possible. In order to avoid excess loss we estimate the thickness to be around 1000 Å. For such a choice we can derive an expression for the group delay of the reflected wave and hence arrive at an effective penetration depth for the grating. This yields a group delay equal to $1/(2|r|fo)$ where r is the reflectivity/strip and fo is the centre frequency. By using the relationship $|r| = 0.46\ (h/\lambda)$ (ref.9) where h is the metal thickness and λ the wavelength, we can estimate the penetration for various metal thicknesses (at 460 MHz) and therefore for 1000 Å, $|r| = 0.0072$ and the effective penetration is 36 wavelengths.

The longitudinal mode frequencies for resonators are given by the expression

$$f_n = \frac{n\,V}{2\,L_{eff}} \qquad \text{(ref. 10)}$$

where n is the mode number, V the SAW velocity and L_{eff} the effective cavity length hence, for example, for a 250 KHz bandwidth, a cavity length of 1000 λ would be acceptable before longitudinal modes would impact on the passband. However in our example, where high rejection is required close to the passband a choice significantly less than this would be wise. For this reason, values of 100, 200, 300 and 400 pairs have been assessed.

(b) Reflector array

For ST - cut quartz the reflectivity of a metal strip is dominated by the mechanical loading because the piezoelectric effect is weak. There is thus very little difference between the reflectivity of a shorted or of an isolated strip (ref. 11) and for reasons of yield and isolation we have chosen to use connected strips.

It is interesting to speculate on the situation with respect to other substrates. Wright (ref. 11) calculates the differences for 34° quartz and for Y 112° X lithium tantalate. For the tantalate the difference is quite marked, an effect that might be expected for lithium niobate. For example, the recent work of Mishima etal. shows that for 128Y YX $LiNbO_3$ using Aluminium electrodes the reflection from $\lambda/4$ fingers can be effectively zero if the metal thickness is correctly chosen. In the example given, for $h/\lambda = 0.026$ the mechanical and the electrical effects cancel (they have opposite sign) for connected reflectors. This would imply that for isolated reflectors these effects may enhance one another giving rise to very high levels of reflectivity. Experiments are in hand to test this implication.

The choice of number of reflectors has to an extent been empirically determined. From an initial design with many excess reflectors a reasonable number of between 300 to 400 was determined to ensure low loss and control the length of the chip. This value is satisfactory for metal thickness in the region of 1000 to 2000 Å at frequencies of several hundred MHz.

Experimental results

The initial designs were constructed using 1.6 μ lines on ST - cut quartz, and devices with 100, 200 and 400 pairs with 400 reflector strips were tested. The substrate was $\frac{1}{2}$mm thick and in line metal

package was used for encapsulation. All lines were equal and $\lambda/8$ offset was used to position the transducer within the cavity. The devices were made with 1000 Å Aluminium. The frequency responses for pairs of coupled 8 λ overlap, 2 λ separated devices are shown in Figure 2. These results were obtained using an HP 8753 network analyser and are for a 50Ω termination. It is clear that the 100 pair device, with spurious signals comparable in level to the required response and an insertion loss in excess of 30 dB is not working as an acceptable resonator. The devices using 200 and 400 pairs however show responses that are typical of coupled resonators operating in a mismatch impedance (The conventional 2 - pole coupled L-C filter response can be made to show a strong dip at the centre in a 50Ω system which is completely removed by terminating with an appropriate higher impedance). In addition, the devices both show poor rejection, although 200 pairs has a sidelobe level of 22 to 23 dB and the majority of this is electromagnetic. One further point should be noted and it is particularly apparent on the 200 pairs device, and this is the spurious response to the low frequency side of the main coupled response - marked x on figure 2. Although this does not impact on our current target specification it's origin is of interest. One of the features of the initial batches of devices was a variation in the centre frequency of in excess of 100 KHz. It was considered that this arose from a variation in the metal thickness. Figure 3 shows the results of a series of measurements on devices with varying metal thickness, compared to a theoretical value of $\Delta f/fo = -10\,(h/\lambda)^2$ (in all cases the frequency increases as metal thickness decreases). This graph shows very clearly the extreme sensitivity of the resonant frequency to the precision of metal layer; to achieve a placement accuracy of 25 KHz the Aluminium thickness must be controlled to within 10 Å in a thickness of 1000 Å i.e. 1%. In order to achieve this level of control some modifications to the evaporation process and the use of dry etching was necessary but tolerances in thickness of better than 1% are now being achieved.

To test the precision of the controls imposed on the manufacturing process a series of 2 - pole devices were made based on a 300 pair transducer. Again 8 λ overlap and 2 λ separation were used, all fingers were $\lambda/4$ and a $\lambda/8$ offset was employed. Results were obtained for a similar number of metal thicknesses and frequency placement to within $\pm$ 25KHz was obtained. The insertion loss of the units varied as shown in Figure 4 where the minimum loss is plotted as a function of metal thickness; the optimum thickness for this device resulted in $h/\lambda \simeq$

These 300 pair coupled resonators were laid out so as to facilitate interconnection and a number of monolithic 4 - pole coupled units were made and tested. The results, tested in a standard 50Ω system showed a peaky response with a large spurious signal on the lower frequency side; a typical example is shown in Figure 5. The cause of this spurious ripple / sidelobe has been investigated.

Conventional resonator analyses have focussed their attention on the effect of internal reflections in two ways (i) By analysis of the reflectivity of the transducer compared to it's conversion loss (ref.10 p. 90) or (ii) By determining the conductance of the transducer when reflections are included. We have adopted the latter approach and the result of our analysis is contained in Figure 7. Here the conversion loss and the reflectivity are plotted for a 300 pair transducer matched at the maximum of G_A (the acoustic conductance). This shows a highly assymetric profile with the peak of conversion loss occuring at a minimum of the reflectivity (a similar result is obtained by using the approach of section (i) above).

This result suggests that improved performance will be obtained if the transducer spacing is changed so that the " reflectionless" condition pertains. Such a change, of 0.025μ in the wavelength destroys the "neat" constancy of finger width which was one of the original constraints and imposes a heavy burden upon the processing facility. However, by the careful use of dry etching, the initial results, plotted in Figure 6 as a dotted line, do remove the spurious lobe and in addition reduce the loss by around 1.3 dB to a level of 3.4 dB. There is however a slight broadening of the response, not discernible in the figure.

The latest results were obtained using a matched fixture which employed an impedance converter and a single matching coil. The removal of the spurious lobe by moving the transducer spacing to the peak of the acoustic conductance agrees with the conclusions of Wright (ref 11) with respect to the symmetry of the basic resonator and is an important consideration if a symmetric passband is the design goal.

Conclusion

Transverse coupling of multiple SAW resonators can enable a cost effective solution to filtering requirements at high r.f. frequencies if metal deposition and etching can be controlled sufficiently. In order to obtain a high setting accuracy of better than $\pm$ 25 KHz metal thickness must be controlled to within $\pm$ 1%.

This degree of control has been demonstrated for sample filters that couple 4 resonators on ST - cut quartz at a carrier frequency of 466 MHz. With such an arrangement, 55dB sidelobe and 4 dB insertion loss has been obtained.

References

1 Ash E.A. Int. Microwave Symposium, p.385 (1970).

2. Ash E.A. Private Communication.

3. Staples E.J. & Smythe R.C. Proc. Freq. Control symposium, p.322 (1976).

4. Matthaei G.L. etal. IEEE SU-25, p.72 (1978).

5. Desbois J. Proc. U/S Symp., p.841 (1979).

6. Tanaka M. etal. Proc. Freq. Control Symp., p.286 (1984).

7. Horine B.H. MTT-S Symposium, (1985).

8. Haydl W.H. etal. Proc. U/S Symp., p.287 (1976).

9. Morgan D.P. Little Red Book, Elsevier, (1985).

10. Thorvaldsen T. PhD Thesis, ETH Zurich, p.74 (1988).

11. Wright P.V. Proc. Freq. Control Symposium, p.596 (1989).

12. Mishima N. etal. Proc. U/S Symp. (1989).

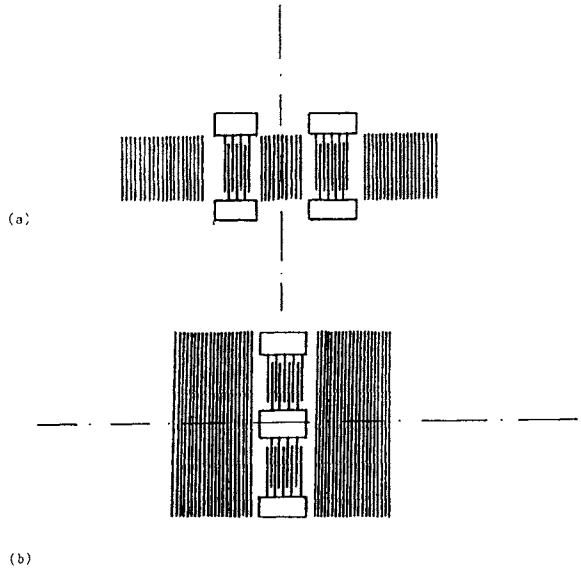
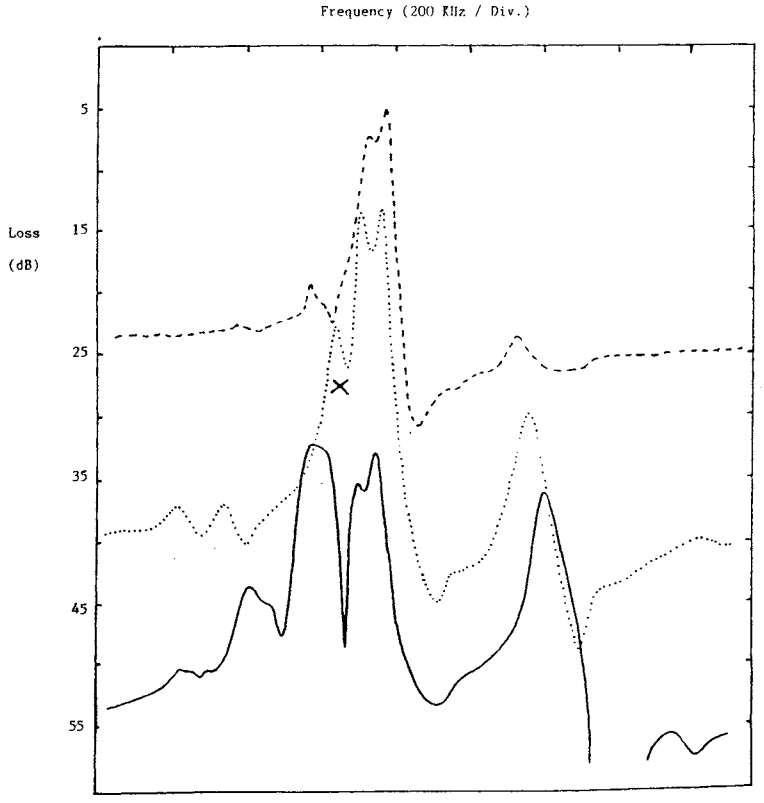


FIGURE 1

(a) In-line (b) Transversal

FIGURE 2 : Coupled pairs

Frequency (200 KHz / Division)



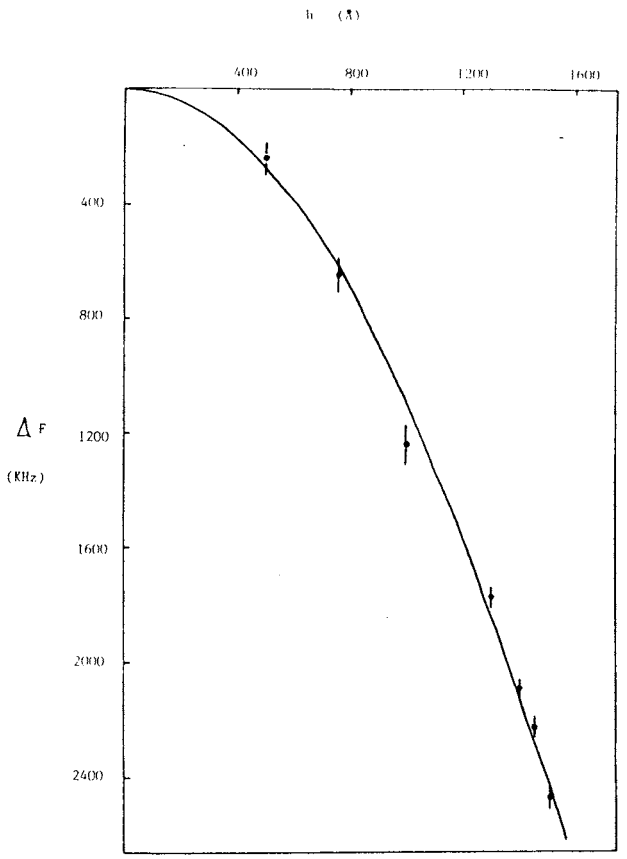


FIGURE 3: Centre frequency versus metal thickness.

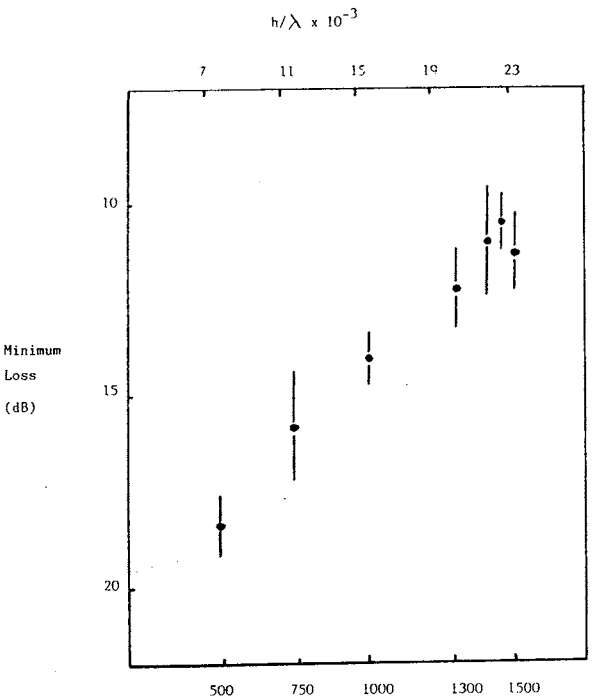


FIGURE 4: Minimum loss versus metal thickness for coupled devices at 466 MHz.

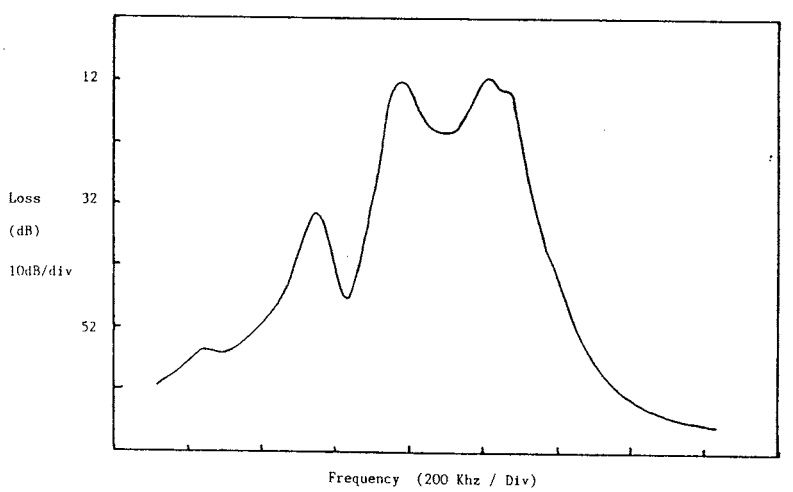
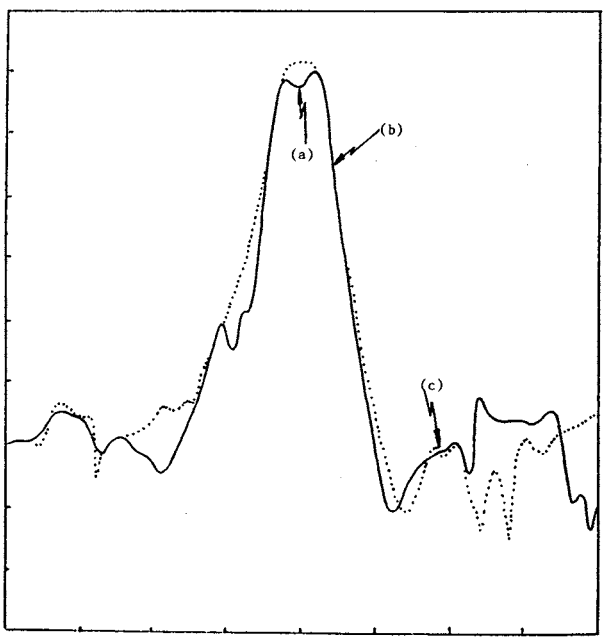


FIGURE 5: 4 - pole coupled resonator 50 measurement

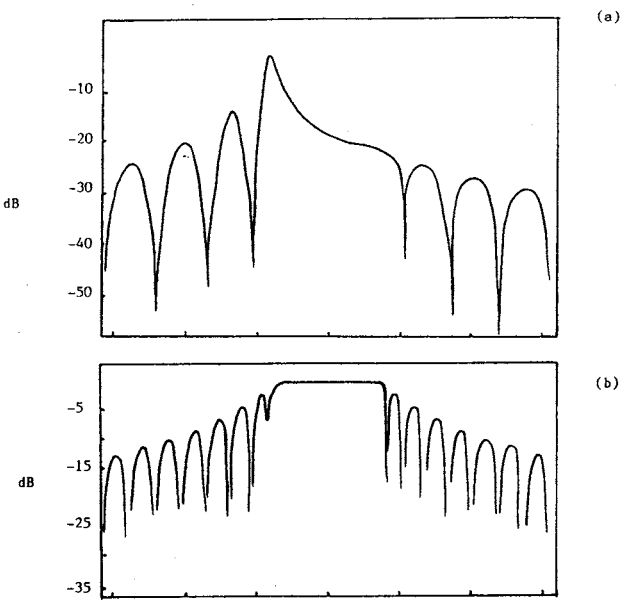
FIGURE 6: 4 - pole coupled filter matched performance

Dotted: Transducer at G_A max.
Solid: /4 constant spacing



Frequency (500 KHz/Div.)

(a) Centre frequency
(b) 227 KHz
(c) 910 KHz



Frequency MHz (2 MHz / Div)

FIGURE 7

(a) Conversion Loss
Matched at maximum of G_A

R_G = 84.02
L = 16 nH

(b) Reflection Coefficient.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-02 Number 4

CHARACTERISTICS OF THE VARIOUS MODES OF ACOUSTIC WAVES IN 36° YX-$LiTaO_3$ SUBSTRATE

Y. Zhang, J.J. Gagnepain, D. Hauden
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

Abstract

The measurements made in interdigital transducers (IDT) on 36° YX-$LiTaO_3$ substrate show that various acoustic waves exist, and that the nature of the dominant mode depends largely on the electrical boundary condition at the substrate surface, in contrast to the Rayleigh wave (SAW). Moreover the electromechanical coupling coefficient calculated by the relative velocity change ($\Delta V/V$) due to metallisation is different from that obtained by the admittance measurement. In order to achieve performances of the devices constructed on this substrate, it is necessary to have a better physical understanding of these phenomena.

In this paper we use the concept of the effective permittivity to analyse the complete spectrum of acoustic waves coupled to the IDT. Field profile of all surface modes are plotted as a function of the depth penetration into the substrate. The coupling coefficient of the SAWs is evaluated by fractional velocity change ($\Delta V/V$). These results are used to explain why the insertion loss is excessively important in the case of a free surface substrate and what is the limitation when the fractional velocity change ($\Delta V/V$) is used to measure the electromechanical coupling for a given wave in a given substrate.

I - Introduction

The measurements made in interdigital transducers (IDT) on 36° rotated Y-cut $LiTaO_3$ substrate with propagation in the X-direction show that a fascinating variety of acoustic waves are launched and propagated in the substrate, and that the electrical boundary condition (i.e. with or without a thin metallisation film) at the surface has a crucial influence on the nature of the dominant modes. Some other unusual behaviours were also observed. To understand the physical origins of these phenomena and further to see whether the characteristics of these waves are interesting for device performances or for new device designs, it is important to carry out a theoretical analysis of the acoustic waves excited by an IDT on this substrate.

In the following section, we study the complete spectrum of acoustic waves that are piezoelectrically active in this cut. We determine the propagation velocity of all possible acoustic modes by plotting the effective surface permittivity as a function of the slowness of the acoustic wave, and then evaluate the electromechanical coupling coefficient by the commenly used fractional velocity change due to surface metallisation. Field distribution of the corresponding modes so determined is plotted in section III as a function of depth into the substrate. These plots bring to light the characteristics of each mode and the role played by the surface metallisation. We give a discussion in section IV about the coupling coefficient defined by different methods or approaches, and at the end of the paper some conclusions of the present analysis.

II - Spectrum of acoustic waves determined by the effective permittivity

The charge distribution $q(x_1)$and the potential distribution $\phi_0(x_1)$ at the surface of a piezoelectric substrate are not independent, but related each other through the effective permittivity defined by $\varepsilon(k_1) = Q(k_1) / [|k_1| \Phi_0(k_1)]$, where k_1 is the propagation constant (wave vector) along the surface, $Q(k_1)$ and $\Phi_0(k_1)$ are the Fourier transform of $q(x_1)$ and $\phi_0(x_1)$, respectively. Since the distribution of the charge and potential gives the electrical boundary condition at

the surface, the function $\varepsilon(k_1)$ may be related to the boundary condition determinants Δ_f (for the free surface) and Δ_m (for the metallised surface). It was shown that [1,2] $\varepsilon(k_1)$ is homogenous in (k_1/ω), i.e. for a given substrate it depends on the slowness or the inverse velocity ($s=1/V$) alone. We have derived a relationship between the effective permittivity $\varepsilon(k_1)$ and the two boundary condition determinants Δ_f and Δ_m as follows [3]

$$\varepsilon(k_1/\omega)=\varepsilon_0(1+\alpha)\frac{j\,\Delta_f(s)}{|s|\,\Delta_m(s)} \qquad (1)$$

The plot of $\varepsilon(k_1/\omega)$ as a function of $s=1/V$ allows us to determine the surface wave velocities for both free (V_f) and metallised (V_m) surface cases. The zero corresponds to V_f and the pole to V_m (see Fig. 1), which are compatible with eq.(1). Possible existence of the pseudo- or leaky surface waves can equally be obtained from the curves, which correspond to the locations where ε or $1/\varepsilon$ is quasi-zero. The Surface Skimming Bulk Waves (SSBW) are located at the points where the slope of the ε curve is discontinuous. Moreover the bulk wave radiation in the substrate is directly related to the magnitude of the imaginary part of the function [2].

In Fig. 1.ε is plotted for 36° YX-$LiTaO_3$ as a function of the slowness, s. It is shown that two sets pole-zero exist; one is situated at the right of the cut-off velocity of the slow shear bulk wave, 1/Vss, and the other at its left. The first corresponds to the conventional Rayleigh wave, and the second is usually called leaky surface wave (LSAW). The SSBW velocities are indicated by V_L (longitudinal), V_{FS} (fast shear) and V_{SS} (slow shear). The fractional velocity change due to metallisation is generally considered as an estimation of the electromechanical coupling. The numerical values of the velocity and $(V_f - V_m)/V_f$ of the relevant modes in this cut are listed in Table I. We remark that the difference between V_{FS} and LSAW V_f is less than 0.1 m/s. The ratio of low and high frequency limite values of ε is about 1.175. The measured values of the LSAW velocities are also given for comparison [4].

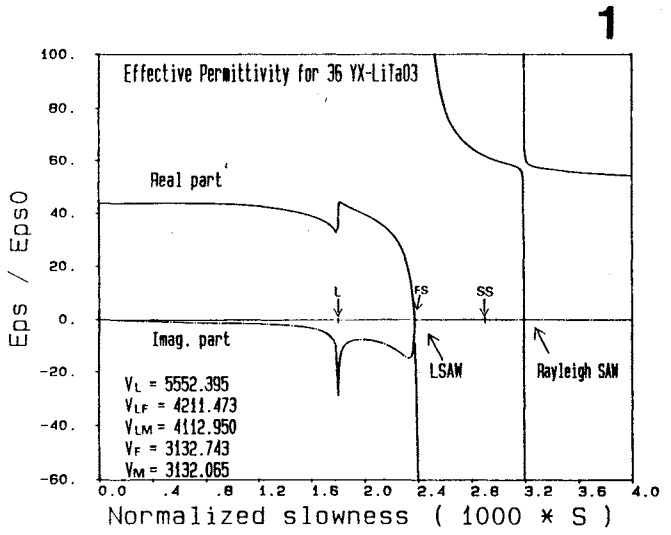


Fig. 1 : Effective permittivity vs. the slowness of the acoustic wave measured along the substrate surface. The positions at which the slope of the curves is discontinuous correspond to the cut-off values of bulk waves or SSBW velocities ; the zeros and poles correspond to the surface wave velocities.

Tableau I : Some characteristics of acoustic modes in 36° $LiTaO_3$

		V_L (m/s)	V_{FS} (m/s)	V_{SS} (m/s)	$\varepsilon_{HF}/\varepsilon_o$*	$\varepsilon_{LF}/\varepsilon_o$*		V_f (m/s)	V_m (m/s)	$2(V_f - V_m)/V_m$
	SSBW	5552,397	4211,570	3367,085	43,5	51,1				
LSAW	calculated							4211,473	4112,950	4,7 %
	measured [6]							4204	4092	5,3 %
Rayleigh								3132,743	3132,065	0,043 %

* ε_{HF} and ε_{LF} are the values of the effective permittivity at high and low frequency limit, respectively

III - Field profile of the surface modes in the substrate

The velocities of the surface waves coupled to the IDT being known, the study of their characteristics is done, in particular, the field profile in the substrate. The three components of the mechanical displacements u_k and the electrical potential ϕ associated with these waves are plotted in Fig. 2-9 as a function of depth into the substrate. Fig. 2 represents u_k and ϕ for the Rayleigh modes on a free surface. The u_1, u_2 and $u_4=\phi$ are purely imaginary, u_1 being real. Fig. 3 represents the metallized case. We see from these plots that the field profile remains almost the same without regard of the surface metallisation. These waves can be called generalized Rayleigh surface wave because u_1 and u_2 are the dominant components, u_3 being small; this means a quasi-sagittal-plane polarisation of the wave. The associated potential $u_4=\phi$ is also small, so a weak electromechanical coupling to these waves, as indicated by the relative velocity change (see Table I).

Figs. 4 to 7 are given for the LSAW modes. Now, all the mechanical components and electrical potential are complex; the part corresponding to the Rayleigh wave is plotted on Figs. 4 and 6 ; while the supplementary part is given in Figs. 5 and 7, respectively for metallized and free surface substrate. Now the waves are characterized by the component u_3 and the electrical potential ϕ, u_1 and u_2 are small, so the waves are quasi-horizontally polarized. We note that the depth penetration into the substrate of the waves is very dependent on the surface metallisation, as can be seen by comparison of the curves plotted on the Fig. 4 and Fig. 6. Also noticed is the fluctuation of the curves in the free surface case, in particular of the components u_1 and u_2. Though they are small with respect to u_3 and ϕ for $x_2 \leq 100\lambda$, they may become comparably large for deepper points x_2, as shown in Fig. 8 and 9 which extend the Fig. 6 and 7 respectively to 200λ in depth. All the u_k and ϕ oscillate with x_2 and do not tend to zero. This is due to the fact that one of the four partial waves,the slow shear one, has a bulk wave behaviour: it leaks energy into the substrate. However the coupling to this mode is so small that the energy losses caused by this leaky wave can be neglected when the surface is metallized. In other words, the attenuation along the propagation direction will not be significant in practical devices. When the surface is unmetallized, the power density at the surface is largely weakened by the great penetration of the wave which appears like a sagittal plane diffraction. The waves can be considered as a generalized Bleustein-Gulyaev type in spite of their leaky behaviour.

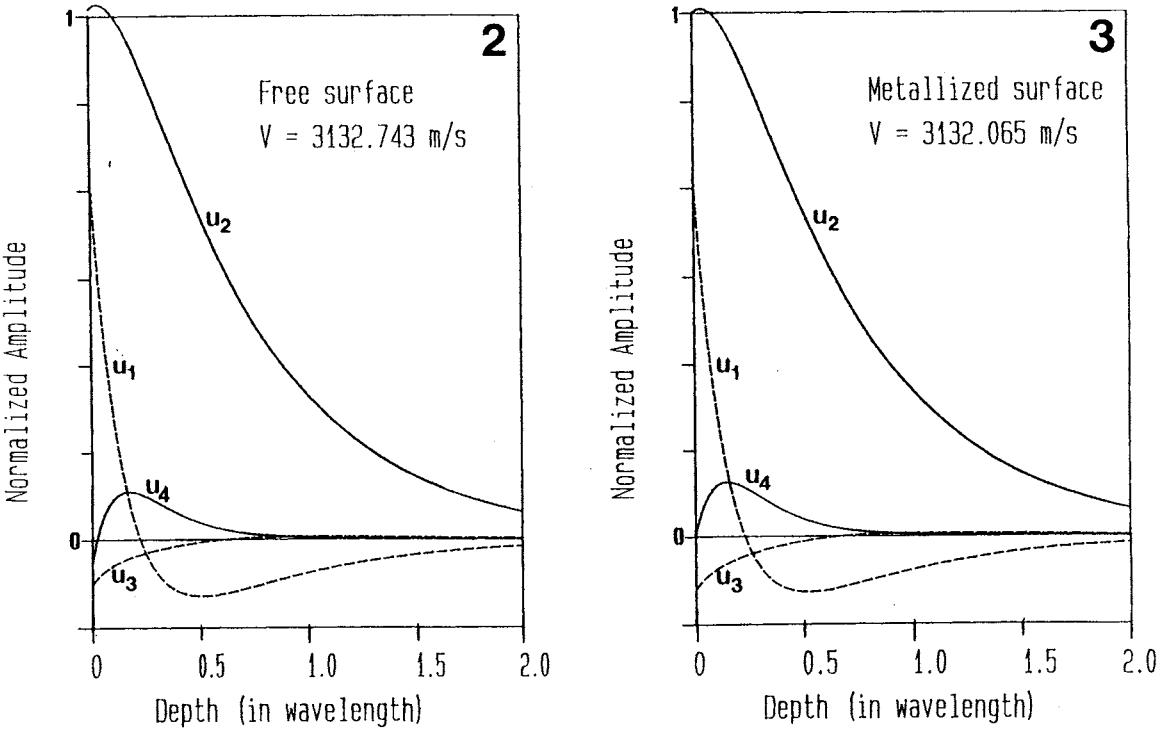


Fig. 2 : Field profile of the Rayleigh surface wave plotted as a function of depth into the substrate in the free surface case. u_1, u_2 and u_3 represent the mechanical displacement components, $u_4 = \Phi$ is the electrical potential

Fig. 3 : idem to Fig. 2 but for the metallized surface case

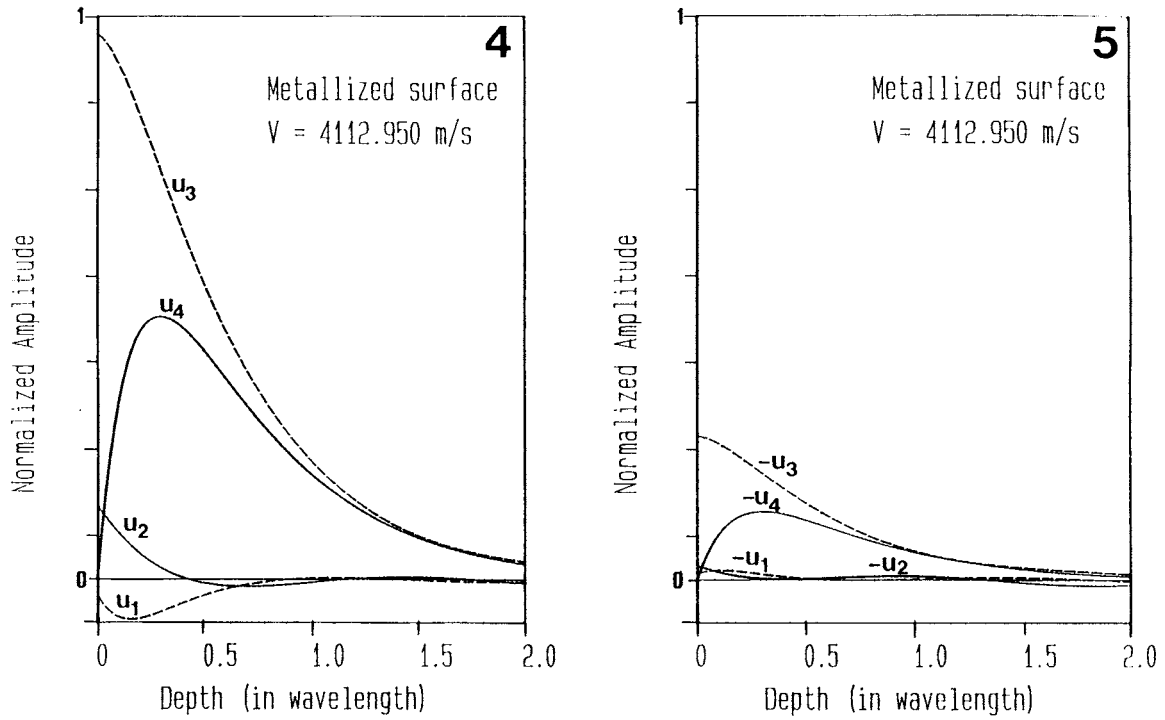


Fig. 4 : Field profile of the leaky surface wave on metallized substrate plotted as a function of depth into the substrate. The curves correspond to those in Fig. 2

Fig. 5 : The supplementary part of the fig. 4, which is in quadrature with it.

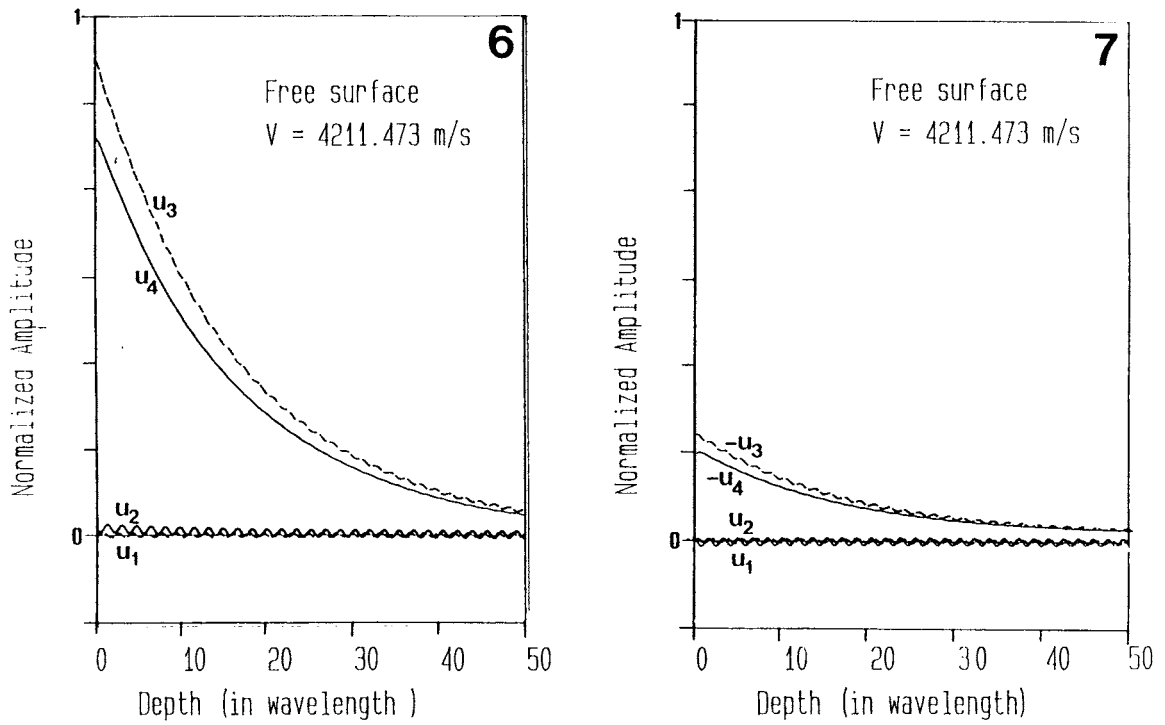
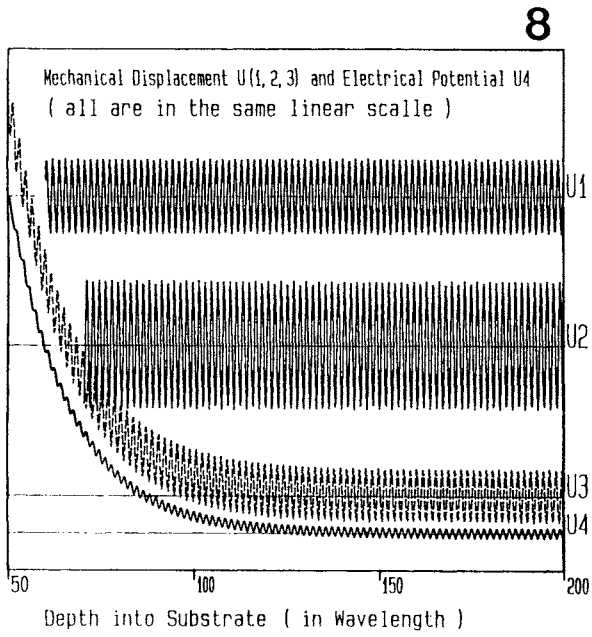


Fig. 6 : Field profile of the leaky surface wave on free substrate plotted as a function of depth into the substrate. The curves correspond to those in Fig. 2

Fig. 7 : The supplementary part of the Fig. 6, which is in quadrature with it



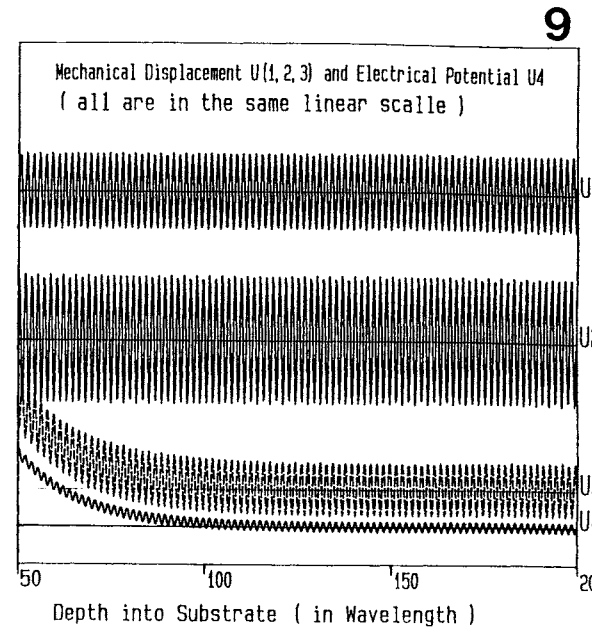


Fig. 8 : idem to Fig. 6, but for greater values of substrate depth (50 $\lambda \leq x_2 \leq 200\ \lambda$)
Fig. 9 : idem to Fig. 7, but for greater values of substrate depth (50 $\lambda \leq x_2 \leq 200\ \lambda$)

IV - Discussion on the electromechanical coupling coefficient

Disagreement exists between the results obtained by admittance measurement and by velocity measurement in what concerns the coupling coefficient. The measured admittance gives an effective coupling coefficient considerably larger than the value estimated from fractional velocity change ($\Delta V/V$) due to surface metallisation. This may be attributed to the fact that behaviours of the dominant waves are very different in the free surface case and in the metallised surface case.The structure of the Rayleigh waves remains nearly identical in both free substrate and metallized substrate. Only the velocity changes, and this allows one to define a coupling coefficient independant of the boundary condition at the surface. But the admittance measurement was made considering a free surface outside of the IDT region, and in this case a SSBW may be strongly launched simultaneously with the LSAW. Their interaction would take place and the resulting electrical admittance of the device be changed to a certain extent.

In fact it has been shown that[5] the surface potential of the excited wave is proportional to $\Delta V/V$, so the currently used criterion of selecting the optimum cut and optimum propagation direction of the electroacoustic surface waves is justified. Those cuts and directions, for which the relative difference in velocities at the free and metallized surface is maximum, are considered as the optimum. However, the parameter $\Delta V/V$ was derived by the "first order" Taylor series expansion of the boundary condition determinants Δ_f and Δ_m with respect to the propagation vector k_1 (or s) around the free and metallized surface wave vectors k_f and k_m, respectively. Physically speeking, the expansion is valid only if the two SAW modes are well isolated from the others at the frequencies considered. But for 36° YX-$LiTaO_3$ cut, the LSAW is very near the fast shear mode where the corresponding decay coefficient changes from complex to real (physically this means that the corresponding mode becomes a radiating mode in the form of bulk wave). Mathematically, the first order Taylor expansion is no longer valid when the two determinants are not analytic in this region because of the root behaviour of the penetration coefficients.

A simple mathematical model proposed by Hashimoto[6] which takes into account the interaction of the LSAW and the SSBW simultanously launched from the IDT predicts an effective value of $k^2 = 5.6$ to 6.6%, depending on the total

number of the IDT finger pairs. The simultanous excitation of the fast shear SSBW which has a relative large susceptance increases the LSAW conductance, so that the effective k^2 for LSAW is larger than the conventionally estimated $2(\Delta V/V)$.

V- Summary and Conclusions

Our theoretical analysis shows that in 36° YX-$LiTaO_3$ two types of surface waves exist : the generalized Rayleigh waves and the leaky surface waves. The first is not dependent on the electrical boundary condition and it presents a weak practical interest because of its extremely small piezoelectric coupling. In contrast, the LSAW with shear horizontal polarisation is the dominant propagating wave. When the substrate surface is metallized, the LSAW behaves like a Bleustein-Gulyaev wave but with a much smaller depth penetration into the substrate and two non zero mechanical displacement components in the sagittal plane; when the surface is free of metallisation, the main excited waves will be a mixture of the LSAW and the SSBW, originated from the fast shear bulk wave. Both are quasi-horizontally polarized. Because the LSAW and the SSBW modes have exceedingly close velocities and nearly the same energy distribution and polarisation, it would be difficult to identify them one from the other neither by calculations nor by experiments. As their energies are very widely spread into the substrate, important insertion losses occur due to this sagittal-plane diffraction. Consequently, they cannot be efficiently utilized in surface-transmission devices such as delay lines or filters unless some other measures of energy trapping are taken, for example, metal strips [7] or shallow gratings [8,9].

Acknowledgements

We wish to thank Dr. J. Débois of the Thomson-Sintra, DTAS, Valbonne, France for his interest and helpful discussions.

Bibliography

[1] R.F. Milson, N.H.C. Reilly and M. Redwood, "Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers", IEEE Trans. SU-24, 147 (1977).

[2] D.P. Morgan, "Surface-wave devices for signal processing", Studies in electrical and electronics engineering 19, Elsevier, 1985, pp. 45-46.

[3] Y.W. Zhang and M. Planat, Journal of Appl. Phys., (à paraître mi-février 1990).

[4] J.M. Hodé, Thomson-Sintra DTAS, Valbonne, France (private communication)

[5] L.L. Goryshnik and S.N. Kontratyev, "The excitation of electroacoustic surface waves by electrode transducers", Radio Eng. and Electron. Phys. 19 (8), 96 (1974).

[6] K.Y. Hashimoto and M. Yoronaguchi, "Effective electromechanical coupling to leaky SAW on 36° YX-$LiTaO_3$", Japanese J. Appl. Phys. 26 (1987), p. 159-161.

[7] K.H. Yen, K.F. Lau and R.S. Kagiwada, "Shallow bulk acoustic wave filters", IEEE 1978 Proc. Ultrasonics Symp., pp. 680-683.

[8] D.F. Thomson and B.A. Auld, "Trapping and temperature compensation of leaky surface waves on lithium niobate and lithium tantalate using shallow gratings", IEEE 1985 Proc. Ultrasonics Symp., pp. 203-206.

[9] K.Y. Hashimoto, M. Yamaguchi and H. Kogo, "Interaction of high-coupling leaky SAW with bulk waves under metallic-grating structure on 36° YX-LiTaO3", IEEE 1985 Proc. Ultras. Symp., pp. 16-21.

Applications

APP-03 Synchronized Systems

Chairman: S.J. Feltham

Philippe Monier

Philippe Monier was an extremely capable and intelligent research-oriented engineer, with a unique mixture of expertise in physics, chemistry and electronics. He possessed an insatiable curiosity and interest in all manner of physical phenomena ranging from the (so-called) mundane to the almost bizarre.

As a staff member of the European Space Agency working in the Technical Directorate he was involved in many projects.
Philippe was widely respected for his technical skills and wide ranging abilities, both in international industry and the Agency.
He was the recognised ESA authority on microwave devices, materials and technologies, for the special requirements of high reliability space hardware. He had a very good knowledge of semi-conductor physics and fabrication techniques and the related practical problems and applications limitations. This experience was gained in industry whilst working at the Thomson CSF Research Centre.

Although not his domain of competence when he joined the Agency, he evolved into the ESA expert and focal point for all matters regarding time and frequency standards together with the related advanced technologies, particularly for deep space, navigation and telecommunications applications.

The endless enthusiasm for the study, analysis and characterisation of the unusual together with a complete absence of arrogance or bombast made him many friends and close colleagues. He will continue to be remembered and missed, for a long time.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-03 Number 1

FREQUENCY SOURCES FOR SPACE APPLICATIONS

S. J. Feltham
RF Systems Division
European Space Agency
Keplerlaan
2200 AG Noordwijk
The Netherlands

Abstract

The current and planned areas of interest of the European Space Agency (ESA) in the field of time and frequency standards are presented. Subjects covered include Communications, Earth Observation and Scientific Applications. Requirements for future ESA programmes are discussed, and existing ESA hardware developments outlined.

1. Introduction

The European Space Agency (ESA) was formed in 1975 by agreement between 11 European governments to coordinate their efforts in a joint space endeavour. In 1987 Austria and Norway became full members and Finland an associate member. Additionally, Canada has an agreement for close cooperation with the Agency and participates in some programmes.

In order to ensure Europe remains at the forefront of progress, one of ESA's main tasks is to prepare for its member states, at regular intervals, long-term plans proposing the direction European Space Research should take. Examples of future programs will be briefly reviewed and their performance requirements for frequency stability discussed.
ESA is essentially an R and D organisation and does not develop or manufacture its own spacecraft. Development work is carried out by industry under the guidance of ESA engineers.

Precise knowledge of orbital location is important for all types of spacecraft. The measurement accuracy of range and range rate depends on the stability of the reference oscillators. Noise and environmental modulation are two types of oscillator instability affecting typical space systems performance. Spacecraft range is normally determined by the round trip travel time of an electromagnetic wave between the spacecraft and a ground station. This is achieved by measuring the phase of the wave returned to the ground station. Measurement of phase is, in principle, a time measurement, consequently the system master oscillator is the 'yard stick' by which range is determined.

Range rate or satellite velocity relative to the tracking site is determined by measuring Doppler shift.
A full description of Frequency Stability requirements for space communications and tracking systems is given in [1] and a discussion of the methods used for accurate range and range rate determination are outlined in [2].

2. Clocks For Space Applications

Clock types which are suitable for space use include quartz crystal oscillators, rubidium oscillators and cesium beam tubes. Hydrogen masers have not reached the stage of development where they can be considered for space applications. The final choice of device type, depends upon a number of parameters such as short and long term stability, temperature sensitivity, frequency retrace after warm-up, mass, volume, power consumption etc. A comparison of some important aspects of these devices is given in Table 1.

PARAMETER	Quartz Crystal	Rubidium	Cesium
Mass (Kg)	0.5 - 1.4	1.8 - 2.3	9 - 31
Relative Cost	1	2 - 4	10 - 20
Short Term Stability 1 sec.	1×10^{-11}	1×10^{-11}	$7X10^{-11}$ to $5X10^{-12}$
Drift per Year	1×10^{-7}	1×10^{-10}	1×10^{-13}
Retrace	1×10^{-9} in 2 hr.	1×10^{-11} in 120 sec.	1×10^{-11} in 1 hr.
Environmental Effects			
The numbers refer to the order of merit in which the devices perform.			
Warm-up	3	1	2
Acceleration	3	2	1
Temperature Range	2	1	3
Vibration/Shock	3	2	1
Magnetic Field	1	3	2
Radiation	3	2	1

Table 1
Comparison of Major Parameters for Space Clocks
Data from ref. [21]

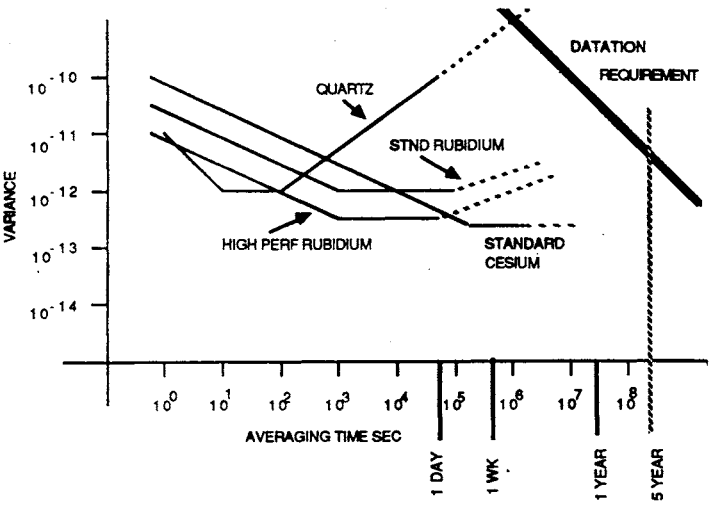


Fig.1
General Comparison of High Performance Oscillators Ref.[21]

Life time of the devices is dependant upon fundamental wear out mechanisms and tolerance to the space environment. Cost is shown relative to a quartz clock. The mass of the cesium tube and to a lesser extent the rubidium, can be expected to decrease considerably with the development of optical pumping using solid state laser diodes.
The hostile conditions of space can result in degradation of phase noise and produce frequency offsets. Environmental parameters which cause frequency or timing errors include magnetic field, pressure, acceleration, total dose radiation and temperature. Vibration effects degrade the short term stability of a quartz source, primarily due to impacts on the crystal. In atomic clocks, recovery is dependent upon the servo loop response time. From Table 1, it is clear that atomic clocks are in general more tolerant of environmental effects. For space applications, quartz and rubidium sources are preferred in terms of mass, volume, power consumption and life time. The final selection of device technology depends on the foreseen application and required performance accuracy.

For an application such an Advanced Terrain Tracking Altimeter (ATTA) the short term stability of either a quartz or rubidium source is satisfactory. The long term requirements could be met by a rubidium clock without any reference to an external reference. Quartz would need to be corrected at least 3 times per year.

The requirement to provide datation accuracy of $\pm$ 1msec. over a 5 year life time of the instrument, is a major problem if no reference external source is used. As can be seen from Fig.1 the quartz based system would require correction from an external source every week. A rubidium frequency source would require yearly corrections. Only the cesium clock could meet the system requirement without correction.

3. The Space Environment and Product Assurance

Commercially available quartz and atomic Ultra Stable Oscillators (USO) have the desired performance (phase noise, long term and short term stability) to meet the design requirements for most space applications. The main problem is to qualify the devices for operation in the space environment. Space qualification involves subjecting the device to a strict sequence of environmental tests supported by design analysis. The evaluation programme is under the control of Product Assurance (PA) engineers.

The qualification test programme includes:

- Electromagnetic Compatibility (EMC);
- Vibration and Shock;
- Thermal Vacuum Temperature Cycling;
- Electrical characterisation.

3.1 Technology Development

The first stage of any technology programme involves a feasibility study followed by the design and manufacture of a Bread Board Model (BB). At this stage emphasis is placed on miniaturisation (mass and volume) and efficiency optimisation (dc-power requirements). The BB must as a minimum meet the main technical specifications and demonstrate its ability to meet or exceed the foreseen technical and environmental requirement.
The next development phase calls for the manufacture of an Engineering Model (EM). The EM design must have the capability of satisfying the equipment specification over the full foreseen electrical and environmental range. The EM should ideally be manufactured to the Flight Model (FM) design but not necessarily using hi-rel components or formal PA control.

During the EM phase, components, materials and manufacturing process selection should be in accordance with ESA PA requirements [3]-[18]. Non qualified components, materials and processes should be identified, evaluated and qualified according to ESA PA requirements. Approval of the EM design for flight manufacture is based on two parts, test results and design analyses. The most important design analyses consists of:

- Design (electrical and interface requirements);
- Mechanical (survivability and reliability);
- Thermal (temperature distribution);
- Failure Modes, Effects and Criticality Analysis (FMECA);
- Parts Applications;
- Worst case analysis;
- Radiation Susceptibility;
- Reliability (probability of success).

ACCEPTANCE TEMP. RANGE (°C) REFERRED TO UNIT BASEPLATE				QUALIFICATION TEMP. RANGE (°C) REFERRED TO UNIT BASEPLATE				START-UP MINIMUM TEMPERATURE (C)
OPERATING		NON OPERATING		OPERATING		NON OPERATING		
MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	
0	+45	-40	+70	-10	+55	-50	+80	-25

ACCEPTANCE		QUALIFICATION	
20 HZ	0.025g^2/Hz	20Hz	0.053g^2/Hz
20-100Hz	+3dB/OCT	20-100Hz	+3dB/OCT
100-700Hz	.125g^2/Hz	100-700Hz	.263g^2/Hz
700-2000Hz	-3dB/OCT	700-2000Hz	-3dB/OCT
COMPOSITE VALUE: 13.2 grms		COMPOSITE VALUE: 19.14grms	

Table 2
Typical Acceptance and Qualification Thermal Vacuum and Random Vibration Ranges

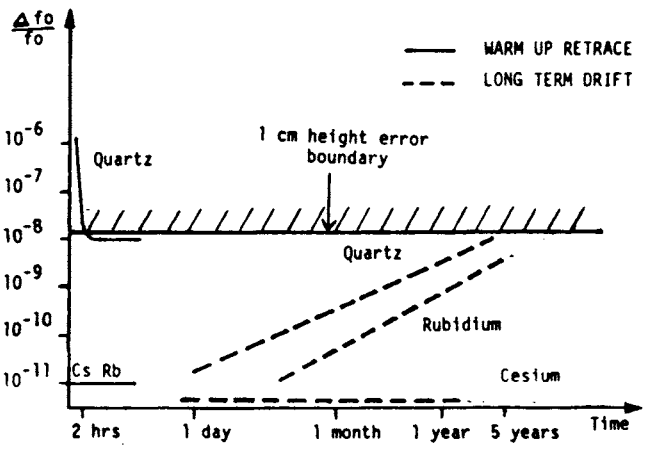


Fig.2
Long Term Stability Performance of Quartz and Atomic Clocks, ref.[20]

When non-compliance or failure occurs, the equipment supplier is obliged to identify design modifications and justify them by analyses and test. The finalised EM design is frozen at a Critical Design Review (CDR) where approval to start flight manufacture is given. The CDR data package forms the basis of the review and covers all aspects of the unit design.
Formal protocol flight equipment manufacture follows successful close-out of the CDR. The QM is manufactured on a PA approved line and is fully representative of the Flight Model (FM) design, in-corporating 'hi rel' components and PA control. Ideally the QM is chosen at random from the flight unit manufacturing run. The QM is tested to predefined qualification limits in order to validate the electrical, mechanical, thermal and environmental design.

FM units are tested to 'acceptance' levels which have a margin over the foreseen flight levels. The purpose of the acceptance test is to verify workmanship quality and electrical performance. Formal delivery of FM units (and the QM) is performed by a Delivery Review Board (DRB) which consists of representatives from all customer organisations. Typical qualification and acceptance levels are shown in table 2.

4. Oscillator Requirements for Communications Satellite

Communications satellites stationed in geostationary orbit, act as permanent relay 'stations' between different locations on the earth. Satellites make it possible to communicate quickly accross considerable distances whilst allowing multiple routing of traffic. They also lend themselves perfectly to mobile service communications with for instance ships at sea.

The latest ESA communications satellite, Olympus, was designed to demonstrate Europe's capability to develope and operate a multi-purpose spacecraft. Olympus used a standard rf link phase measurement technique to make range measurements for in-orbit station keeping.

The next Agency Communications Spacecraft will be the Data Relay System Satellite (DRSS). DRSS will support forward communications links from ground via satellite to user space terminals and vice versa. In addition to relaying data, the DRSS is designed such that accurate range and range rate measurements can be performed in order to support range and range rate measurements with user spacecraft.
To achieve the required measurement accuracy, it is necessary to phase lock all frequency sources to one common frequency reference.
This leads to a stringent requirement on the frequency stability of the master-reference, and an extremely good close to carrier phase noise performance.

The current design of the DRS transmission system and range rate measurement accuracies, assume that the frequency reference on board DRSS has a phase noise and one second frequency stability equal to those of existing high performance crystal oscillators. DRSS therefore does not require additional USO development but a demonstration that such an oscillator can be flight qualified. The Electrical Performance specification required for the reference clock is shown Table 3.

Output signal	: 5 MHz sine wave	
Long term stability	: $1X10^{-8}$/year	
Allan variance	: $1X10^{-12}$ for 1 sec.	
Phase noise	: $L(F_1)$ = 0.5 Hz,	-105 dBC/Hz
	$L(F_2)$ = 5.0 Hz,	-135 dBC/Hz
	$L(F_3)$ = 100.0 Hz,	-149 dBC/Hz
	$L(F_4)$ = 1000.0 Hz,	-149 dBC/Hz

Table 3 DRSS USO Specification

4.1 DRSS Frequency Generation Unit (USO Requirements)

The USO will form part of a frequency generation and distribution subsystem which provides a 5 MHz ultra stable, low phase noise, reference to 25 Tunable Frequency Converters (TFC) units. The subsystem includes the following functions:

- USO to provide 5 MHz reference;
- 5 MHz low noise distribution amplifier to overcome system insertion and division losses;
- Power Supply Unit (PSU) to provide low ripple supplies for the USO and distribution amplifier;
- 25 way power divider for 5 MHz reference distribution.

The proposed source device is a high precision, low-aging rate 5MHz Voltage Controlled Quartz Oscillator utilising a 5th overtone, AT cut electrodeless resonator.

The oscillator tuning voltage input is highly sensitive to pick-up. Consequently a main design driver for the PSU will be the suppression of output ripple to ensure that the phase noise performance of the USO is not degraded.

4.1.1 Mechanical/Thermal Description

The USO module is manufactured from solid aluminium alloy. The box is sealed using a aluminium alloy lid and a trapped rf gasget.

Temperature sensitive components are contained in a oven assembly, these include:

- BVA resonator;
- 5 MHz oscillator and automatic gain control circuit;
- Frequency pulling network;
- Oven control circuitry;
- 17 V to 7 V Voltage regulator for the tuning voltage reference;
- Tuning control circuit.

This arrangement provides a high degree of thermal isolation resulting in a low thermal susceptability. The mechanical integrity of the package is provided by using a high strength dewar which is embedded in rigid silicone foam. The dewar is sealed to ultra high vacuum.

4.1.2 Frequency stability: Drift and Calibration

The frequency stability requirement is driven by the tolerable range rate error. The specified Allan standard deviation of the USO is:

$$\sigma_y(\tau) = 1 \times 10^{-12} \text{ for } 0.1 \text{ sec.} < \tau < 10 \text{ sec.}$$

with a typical value of:

$$5 \times 10^{-13} \text{ for } \tau > 10 \text{ sec.}$$

The DRSS range rate error contribution due to short term frequency fluctuations is given by:

$$\Delta r = c \sqrt{2}\, \sigma_y(\tau)$$

or, 4.2×10^{-4} m/sec.

However during station keeping manoeuvres there may be a six fold degradation in performance due to vibration. Performance may also be degraded by the oscillator's magnetic susceptibility. The Earth's magnetic field at geo-stationery altitudes is approximately 2×10^{-3} gauss and can thus be neglected when compared to other sources of error.

5. Earth Observation

Viewing the Earth from space emphasises, more perhaps than any other perspective, the global nature of our environment and its sensitivity to change from both natural and man-made sources. Satellites have already contributed substantially in assisting man to make proper use of the environment.
ESA's earth observation programme covers two main fields, the observation of the seas, oceans and land masses and the observation of the atmosphere which provides meteorological information.

5.1 European Remote Sensing Satellite (ERS-1)

The first European Remote Sensing Satellite is currently under final testing ready for launch in December 1990.
The instrument payload consists of the following elements:

- Active Microwave Instrument (AMI)
  which provides all weather high resolution imagery over coastal zones, land and ice caps;

- Radar Altimeter (RA)
  provides accurate measurement of sea surface elevation, wave height, sea-surface wind speeds and various ice parameters;

- Along Track Scanning Radiometer and Microwave Sounder (ATSR)
  combines infrared and microwave sensors for the measurement of sea surface temperature, cloud top temperature, cloud cover and atmospheric water vapour;

- Precise Range and Range-Rate Equipment (PRARE)
  For the accurate determination of satellite position and orbit characteristics and for geodetic 'fixing' of ground stations;

- Laser Retro-Reflector (LRR)
  measurement of satellite position and orbit using laser ranging stations on the ground.

From the point of view of the satellite clock the RA and PRARE place the most demanding frequency stability requirements on the quartz oscillators.
The major contribution to the altimeter error budget is the frequency drift of the USO due to aging and frequency retrace after warm-up [19]. Frequency re-trace describes the ability to maintain the previous clock time on restart following an off period. The long term drift error is reduced by a dedicated in-flight calibration technique.

A number of frequencies are directly generated from the 5 MHz quartz USO. Amongst these is a 75 MHz signal used in the phase detector. This signal is also transferred to the signal processing sub-system (SPSS), where it is divided down to 15 MHz. It is then routed to the Instrument Data Handling and Transmission (IDHT) as a 1:1 square wave.

In the IDHT the 15 MHz clock can optionally be used as the bit clock during the 15 Mbps tape-recorder dump to ground. In the ground station the bit clock is recovered by the bit synchroniser and referenced to an Cesium standard. The long term stability performance for the USO is typically $< 10^{-8}$. The calibration technique improves the performance to 10^{-9} and, over a 5msec. period to 10^{-10}.
The frequency measured on ground is used to generate a correction factor which is applied during on-ground processing of the RA data.
Using this technique the sea surface elevation can be determined with an error of less than 10 cm.

5.2 Polar Platform

Evolving from ERS, a continuous series of missions are foreseen covering essentially all Earth Observation disciplines by using Columbus and the Free Flying polar platform.
ESA has initiated two parallel activities to study the design of an Advanced Terrain Tracking Altimeter (ATTA) for operation on the polar orbiting platform. [20], [21]. The goal for the uncertainly in height determination is 1 cm. The studies conclude that the main driver in the selection of the on board clock type is the long term stability and frequency retrace after warm-up. Typical long term and clock retrace performance data is given in Fig.2. Comparing these data with the maximum acceptance height error of 1 cm (or 1.2×10^{-8} frequency error) reveals almost no margin for the quartz frequency retrace parameter. The long term drift of quartz is about 10^{-8} in two years. Since the required life time of the instrument is five years, frequency monitoring or the use of an atomic clock is mandatory. A rubidium frequency source could provide sufficient long term stability over a period of five years. However the draw back of the rubidium source is its reliability.
The options for achieving the desired stability requirements include use of very high stability on-board clocks, synchronisation of the satellite clock with a more stable external clock using rf links (eg GPS), calibration of the drift of the on-board clock by reference to very accurate comparisons of the height over a calibration site using laser based measurements, calibration of the drift of the on-board clock using a rf down link to a ground based reference (as per ERS-1). A detailed study and trade-off analysis of performance, hardware, cost and risk of the various proposed techniques can be summarised as follows:

- Time transfer by means of laser pulses is too expensive;
- Frequency monitoring via the GPS provides a convenient solution to time correction of the on-board clock which could be a high accuracy quartz crystal source. This method requires a space qualified GPS receiver;
- On-board frequency monitoring with a rubidium based system using updates to the on-board clock, from either a ground based laboratory standard via PRARE for example, or alternatively by calibration of its drifts every few months. The main drawbacks are mass, power and reliability.

A final decision has yet to be made, but a method similar to that used on ERS-1 will probably be selected.

ESA has initiated two parallel activities to study the design of an Advanced Terrain Tracking Altimeter (ATTA) for operation on the polar orbiting platform. [20], [21]. The goal for the uncertainly in height determination is 1 cm. The studies conclude that the main driver in the selection of the on board clock type is the long term stability and frequency retrace after warm-up. Typical long term and clock retrace performance data is given in Fig.2. Comparing these data with the maximum acceptance height error of 1 cm (or 1.2×10^{-8} frequency error) reveals almost no margin for the quartz frequency retrace parameter. The long term drift of quartz is about 10^{-8} in two years. Since the required life time of the instrument is five years, frequency monitoring or the use of an atomic clock is mandatory. A rubidium frequency source could provide sufficient long term stability over a period of five years. However the draw back of the rubidium source is its reliability.
The options for achieving the desired stability requirements include use of very high stability on-board clocks, synchronisation of the satellite clock with a more stable external clock using rf links (eg GPS), calibration of the drift of the on-board clock by reference to very accurate comparisons of the height over a calibration site using laser based measurements, calibration of the drift of the on-board clock using a rf down link to a ground based reference (as per ERS-1). A detailed study and trade-off analysis of performance, hardware, cost and risk of the various proposed techniques can be summarised as follows:

- Time transfer by means of laser pulses is too expensive;
- Frequency monitoring via the GPS provides a convenient solution to time correction of the on-board clock which could be a high accuracy quartz crystal source. This method requires a space qualified GPS receiver;
- On-board frequency monitoring with a rubidium based system using updates to the on-board clock, from either a ground based laboratory standard via PRARE for example, or alternatively by calibration of its drifts every few months. The main drawbacks are mass, power and reliability.

A final decision has yet to be made, but a method similar to that used on ERS-1 will probably be selected.

6. Scientific Applications

6.1. Cassini

The Cassini mission is a joint undertaking by NASA and ESA to explore the Saturn system with a Saturn Orbiter and Titan Probe.
The mission gives special attention to Titan, Saturn's planet-sized moon, which is blanketed by an atmosphere thicker than out own and may be covered by a global hydrocarbon ocean. The mission will conduct detailed obsevations of several moons of saturn, Saturnian rings, atmosphere and magnetosphere. In addition the mission will pass through the asteroid belt and make observations of at least one asteroid during the cruise phase of the mission. Cassini will also include a fly by of Jupiter to allow observations of the planet, its environment and of the deep magnetotail.

The launch vehicle and orbiter are the responsibility of NASA while the Huygens Probe is furnished by ESA. The orbiter carries the probe and deploys it into the atmosphere of Titan where it will make detailed atmospheric measurements during its 3 hour parachute descent to the surface.

Instruments for the payload of the Huygens Probe and Saturn Orbiter will be supplied both by US and European scientific communities. Arrival time at Saturn is planned for October 2002, 6.5 years after launch. Proposals for mission opportunities are in fact, currently being evaluated. A description of the mission can be found in (22).

6.1.1 The Cruise Phase

During the cruise phase of the mission, it may be possible to determine the mass of a asteroid by Doppler tracking. For two way ranging measurements the carrier signal is generated from a hydrogen maser reference of the NASA Deep Space Network (DSN). For one way measurements the carrier is generated using an on-board reference oscillator.
Collection of 2-way data is achieved with the spacecraft transponder in the 2-way coherent mode. In this mode the downlink is derived from a USO, that is phase locked with the uplink signal producing a 2 way link with frequency stability characteristics determined by the ground maser.

6.1.2 Doppler Wind Experiment (DWE)

The Doppler Wind Experiment will measure the wind velocity in the Titan atmosphere as a function of altitude by Doppler tracking of the descending probe. With USO's on both the probe and the orbiter, their relative velocity can be accurately measured and the horizontal wind along the probe-orbiter line of sight calculated. The probe oscillator stability will limit the accuracy of the wind speed measurement.

A USO stability of 5×10^{-10} over a 3 hour period will permit relative wind velocity measurements over the same time interval, accurate to 2m sec^{-1}.

Minimum volume, mass, dc power consumption and sensitivity to 'g' force are of prime importance for the Huygens probe payload. The probe will only have battery power during its descent to the surface of Titan and will be subjected to strong de-acceleration forces following parachute deployment.

6.1.3 Radar Altimeter

The Radar Altimeter is a probe sub-system. Its main function is to provide, in real time, the altitude of the probe above the surface of Titan. Scientific investigations include the study of surface morphology and roughness, surface reflectivity and sub-surface sounding.

7. ESA Development Activities

7.1 Rubidium Ultra Stable Oscillator

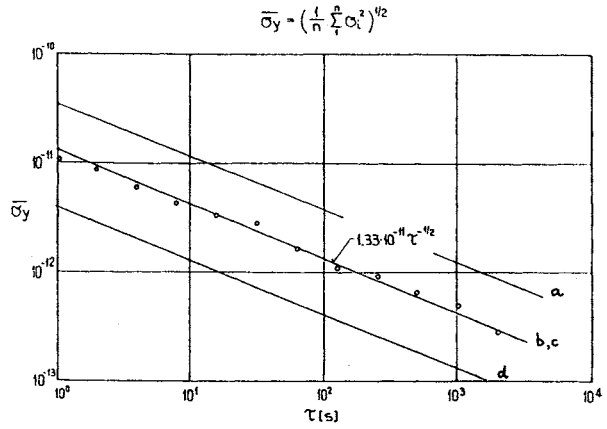
The rubidium USO offers a unique combination of very high short-term frequency stability (5×10^{-12} over 1 sec.), low power consumption and miniaturisation potential. This makes it extremely attractive for applications which require precise time / frequency generation on-board a spacecraft. The realisation of a miniature physics package, bread board electronics and the performance of an ageing test on rubidium bulbs has been successfully completed [23]. The replacement of the microwave cavity at 6.8 GHz by a helical resonator has been demonstrated. The size of the physics package was reduced to 68 cm^3, giving a predicted overall volume of 200 cm^3 for the complete unit including miniaturised electronics. Power consumption was reduced to 6 W and a short term stability of 3×10^{-11} achieved.

7.2 Optically Pumped Cesium Clock

The major advantage of a cesium clock is its long term stability of typically 10^{-13}/year. The stability arises because cesium tubes use collision-less free atomic beams rather than a confined vapour / buffer gas, as for example rubidium. The introduction of optical pumping has the potential to halve the mass of a conventional cesium tube by the elimination of the state sorting magnets. The main barrier to progress is currently the frequency stability of the laser diodes used to pump the cesium atoms. The diode frequency characteristics show distinct jumps which could be subject to unpredictable ageing. The cause of this ageing is not fully understood, but could be due to ageing of the thermal contacts inside the chip. The Frequency stability achieved to date [24], is shown in Fig 3.

8. Conclusions

- ESA is taking an active role in developing the technology for the next generation of communications, observation and scientific satellites.
- DRSS will space qualify a voltage controlled quartz USO.
- The Earth Observation Directorate is currently evaluating the merits of a rubidium clock, compared to a GPS and a phase detection calibration technique for the Polar Platform ATTA.
- Proposals for Cassini experiments are presently being discussed. The advantages of atomic oscillators have been recognised, the main disadvantage being device reliability. The situation should improve as results from the development work feeds back into device technology.
- In addition, ESA is also developing a super hydrogen maser, with improved stability and life for ground applications. The possibility of funding a small mobile earth station for precise time and frequency dissemination and co-ordination via satellite is being evaluated. A plan to set up a small time and frequency laboratory at the European Space Research and Technology Centre (ESTEC) is also under discussion.



List of Acronyms

ESA	European Space Agency
ATTA	Advanced Terrain Tracking Altimeter
USO	Ultra Stable Oscilator
PA	Product Assurance
BB	Bread Board
FM	Flight Model
QM	Qualification Model
EMC	Electromagnetic Compatability
CDR	Critical Dersign Review
DRB	Delivery Review Board
DRSS	Data Relay System Satellite
DRS	Data Relay System
PSU	Power Supply Unit
TFC	Tunable Frequency Converter
ERS-1	European Remote Sensing Satellite
AMI	Active Microwave Instrument
RA	Radar Altimeter
ATSR	Along Track Scanning Radiometer
PRARE	Precise Range & Range Rate Equipment
LRR	Laser Retro Reflector
SPSS	Signal Processing Sub-System
IDHT	Instrument Data Handling & Transmission
GPS	Global Positioning System
NASA	National Aeronautics & Space Administration
VCO	Voltage Controled Oscillator
DWE	Doppler Wind Experiment

Fig.3.
Short-Term Stability of a small Cs-Standard
a. Typical commercial magnetic selection resonator
b. Prototype of optically pumped standard calculated from signal, noise and line Q
c. Prototype of optically pumped standard measurement (average of nine 20 sample runs)
d. goal of this study, ref.(24)

(1) R.Sydnor, J.J.Caldwell.B.E.Rose. ' Frequency Stability Requirements for Space Communications & Tracking Systems '. Proc. IEEE Vol. 54 no.2 Feb.1966 p.231

(2) M.Lopriore ' ESA application for Frequency & Time ' 4th European Frequency and Time Forum. Neuchatel Switzerland March 13-15 1990.

(3) Basic Requirements for Product Assurance of ESA Spacecraft & Associated Equipment. ESA Document ESA PSS-01-00

(4) Software Product Assurance for ESA Space Systems. ESA Document ESA PSS-01-21

(5) Reliability Assurance for ESA Spacecraft & Associated Equipment.ESA Document ESA PSS-01-30

(6) Derating Equirements & Application Rules for Electronic Components.ESA Document ESA PSS-01-301

(7) Failure Rates for ESA Space Systems. ESA Document ESA PSS-01-302

(8) ESA Fracture Control Requirements. ESA Document ESA PSS-01-401

(9) Component Selection, Procurement & Control for ESA Space Systems. ESA Document ESA PSS-01-60

(10) ESA Preferred Parts List. ESA Document ESA PSS-01-603

(11) The Technical Reporting & Approval Procedure for Materials & Processes. ESA Document ESA PSS-01-700

(12) Data for the Selection of Space Materials. Esa Document ESA PSS-01-701

(13) A Thermal Vacuum Test for the Screening of Space Materials. ESA Document ESA PSS-01-702

(14) ESA Process and Materials Test Application Specifications (Subsidiary to ESA PSS-01-700) ESA Document ESA PSS-01-7xx

(15) ESA Software Engineering Standards. ESA Document ESA PSS-05-0

(16) ESA SCC Qualified Parts List ESA Document ESA SCC QPL

(17) Outgassing & Thermo-optical Data for Spacecraft Materials. ESA Document ESA RDxx

(18) Reliability Prediction of Electronic Equipment. Mil-HDBK-217

(19) RA Measurement Error Analysis and Budget. ESA ESTEC Noordwijk (NL) DOC ER-TN-SES-RA-0020

(20) Feasibility Study of an Advanced Terrain Tracking Alitimeter, Selenia Spazio May 1988 Final Report ESA Contract 7199/87/NL/JS

(21) Feasibility Study of an Advanced Terrain Tracking Alitimeter, British Aerospace Sept. 1988 Final Report ESA Contract 7200/87/NL/JS

(22) Cassini Mission: Huygens Probe ESA ESTEC Noordwijk (NL) ESA SCI(89)2 Oct. 89

(23) Rubidium USO Feasibility Study. Final Report Dec.89 ESTEC Contract 7720/88/NL/JG(SC)

(24) Study & Development of a Miniaturised Cesium Clock for Space Applications. Final Report March 1988 ESTEC Contract 5967/84/NL/JG/(SC)

(25) Study & Development of a Miniaturised Cesium Clock for Space Applications. Final Report Rider 1 July 1989 ESTEC Contract 5967/84/NL/JG/(SC)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-03 Number 2

SYNCHRONIZATION OF TELECOMMUNICATIONS NETWORKS

H.J.Bosshard, P.A.Probst
Research and Development, Swiss PTT, Berne

Abstract

In this paper an overview is given on synchronization aspects of the existing plesiochronous network and the relevant CCITT recommendations. Further the specific jitter and wander behaviour of equipment of the Synchronous Digital Hierarchy (SDH) is outlined. Simulation results of a desynchronizer design illustrate the problems encountered in the interworking with the existing Plesiochronous Digital Hierarchy (PDH).

1. INTRODUCTION

The integrated services digital network (ISDN) defined in the I-series of CCITT-Recommendations is based on the transmission and switching of 64 kbit/s channels carrying the information of the various services. The frame structure of the primary hierarchical level at 1544 and 2048 kbit/s supports 24 respectively 30 synchronous multiplexed 64 kbit/s channels. Synchronous multiplexing TDM technique was already adopted about 20 years ago with the first CCITT-Recommendations dealing with PCM equipment for point-to-point applications on existing symmetric pair cables. For these systems no external timing was required, since they were operating in an analogue environment. In the ISDN, where 64 kbit/s are switched, timing signal must be supplied to the digital equipment operating at the primary rate in order to avoid or to limit the occurence of "slips" (loss or repetition of 8 bits word).

A short description of the slip mechanism is given in section 2 below.

Reference to the CCITT Recommendations relevant for the synchronization of digital a network based on the existing plesiochronous hierarchy is given in section 3.

In section 4 the synchronization aspects of the introduction of the new synchronous digital hierarchy specified in CCITT are discussed.

2. SLIP MECHANISM

From the functional viewpoint, a switch is equipped with so called "frame aligners" at the primary rate inputs of the switching network. They perform a function similar to an intermediate buffer, converting the timing of the incoming primary rate signal to the local clock, supplied by the synchronization unit, before TDM processing in the switching network.

The frequency of slips depend both on the capacity of the buffer and on the relationship between the local clock and the timing of the incoming signal. The frame aligners are usually designed to produce 8-bits word (octet) slip in the switched 64 kbit/s channels and to tolerate at the digital interfaces the amount of jitter and wander specified for the primary rate (see section 3).

Concerning the phase relationship between the clock of the incoming signal and the local clock, basically three classes of influences have been identified:

-**Characteristics of network clock** (frequency accuracy, aging, short term stability,...)

-**Propagation delay of the digital transmission and switching equipments** (phase jitter, wander, phase hits,...)

-**Interferences** (external influences like EMC, maintenance and operational failure, system failure, excessive amount of jitter and/or wander,...)

As indicated in table 1, slips due to the sources described above can be reduced or avoided by corresponding means in the design and planning of the network.

3. CCITT-RECOMMENDATIONS

At present the CCITT Recommendations relevant for network synchronization are part of the Blue Book. They are based on studies valid for digital networks using the so called plesiochronous hierarchy as specified in Rec. G.702 (multiplexing from the primary level to the other hierarchical levels using positive/negative stuffing techniques).

When designing a network synchronization system, the following Recommendations are important:

-Network Performance Objectives

* G.822: Controllled slip rate objectives on an international digital connection

-General considerations on network synchronization:

* G.813: Considerations on timing and synchronization issues

-Synchronization requirements to network nodes

* G.811 Timing requirements at the outputs of primary reference clock suitable for plesiochronous operation of international digital links

* G.812: Timing requirements at the output of slave clocks suitable for plesiochronous operation of international digital links

-Specification of digital interfaces based on the plesiochronous hierarchy

* G.823: The control of jitter and wander within digital networks which are based on the 2048 kbit/s hierarchy
* G.824: The control of jitter and wander within digital networks which are based on the 1544 kbit/s hierarchy

In [6] the network timing concept designed for the swiss network based on CCITT Recommendations is described.

4. Synchronization in the Synchronous Digital Hierarchy (SDH)

4.1 Introduction

The introduction of the synchronous hierarchy in the transit network requires the investigation from the synchronization viewpoint of the following items:

-characteristics of the digital path (usually at primary rate) over which timing signal are transmitted between network nodes

-requirements for the external synchronization of SDH multiplexing equipments

-SDH equipment show specific jitter and wander behaviour compared to the plesiochronous system.

4.2 Concept of the Synchronous Digital Hierarchy

BELLCORE proposed in 1984 a synchronous optical network (**SONET**). Bitstreams of 1.5, 6 and 45 Mbit/s of the existing plesiochronous network are packed into the Synchronous Transport Signal (**STS-1**) at 51.84 Mbit/s. Several STS-1 can be synchronously multiplexed by the method of byte interleaving to form the STS-N signal. Optical Carriers of the order N (**OC-N**) transport these STS-N. Bitrates as high as 2.48 Gbit/s are foreseen. But in principle higher capacities could be formed the byte interleaving method.
The recommendations G.707, 708 and 709 of the CCITT define the Synchronous Digital Hierarchy (**SDH**) which is based on a Synchronous Transport Module (**STM-1**) at a bitrate of 155.52 Mbit/s. The plesiochronous bitrates of 2, (8), 34 and 140 Mbit/s are mapped into a STM-1. By a byte interleaving process synchronous transport modules of the order N (STM-N) are formed. Today STM-4 and STM-16 (2.48 Gbit/s) with defined optical interfaces are envisaged for use in a future SDH- Network.

The following advantages led to the proposal of the introduction of the synchronous digital hierarchy (SDH):

- simple multiplexing/demultiplexing

- enhanced drop/insert and cross- connect functions

- introducing broadband services

- high transmission capacities

- performance monitoring on a section and path basis
  due to the built in Section and Path Overhead (**SOH, POH**)

- integrated Data Communication Channels (DCC)

- integrated control functions for enhanced Transmission Network Management (TMN)

4.3 Frame Structure of a Synchronous Transport Module STM-1

In annex 1 the frame structure of a STM-1 is shown. The section and path overhead occupies about 5.8 Mbit/s and provides the above mentioned capabilities. A sytem of pointers indicate the location of the various channels which are mapped into the payload. By this mean multiplexing, drop/insert and cross- connect functions can be realised. The other function of the pointers is the adjustment of the channel position in the case of frequency errors, jitter and wander in the network. The payload of a synchronous transport module STM-1 contains plesiochronous channels which are mapped by a positiv stuffing method. In annex 2 the mapping of a 140 Mbit/s channel into the synchronous hierarchy is shown.

4.4 Jitter and Wander in the Synchronous Digital Hierarchy

In annex 7 a simplified reference model for a synchronous network topology with the necessary timing is given. The plesiochronous channels according to the CCITT recommendation G.703 are mapped into a STM-1 with a terminal multiplexer. The STM-1 signal is routed through a Synchronous Digital Cross- Connect (SDXC) (which allows different routing inside a network node) to a SDH- terminal. After transmitting over STM-N line equipments, regenerators and other network nodes the plesichronous channels are demultiplexed in network node 3. Here the plesiochronous bit- streams have to be fed to a desynchronizer circuit, due to the jitter and wander generated in the SDH- portion of the network.

The network node 1 is timed from a clock through the Multiplex Timing Physical Interface (MTPI). A G.703 interface allows timing with a 2048 kHz clock. The Multiplex Timing Source (MTS) unit gives the possibility to either external timing, internal timing (in case of failure) or extraction of the timing information from the STM-N signals. In the given configuration the other network nodes 2 and 3 and regenerators extract their timing information from the incoming STM-N streams. The other option is to distribute the clock over seperate channels to each network node. It has to be noted that a high multiplication ratio is needed to generate STM-N clock signals (i.e. 2.4 Gbit/s, 622 Mbit/s and 155 Mbit/s) with a basis of 2048 kHz.

Here the main interests shall lie in the jitter which is generated as a plesiochronous channel at a bitrate of 140 Mbit/s is passed through a synchronous network. For other bitrates the results are of a similar nature but less severe. A desynchronizer at the end of a SDH- path has to copé with the following jitter components.

4.4.1 Mapping Jitter

According to the CCITT recommendation G.709 for a 140 Mbit/s channel a row of 1935 data bits contains one justification bit (see annex 2). This justification or stuffing bit normally carries data. The stuffing ratio can be calculated as follows:

The time interval of a row is 125 us/9 in which for a 140 Mbit/s signal 1934.22 bits have to be transmitted in the mean. Therefore 0.77 justification bit per row will contain in the mean no information. For cases with no frequency errors we have a stuffing ratio of 0.77.

Synchronous equipment have a tolerance in frequency of 20 ppm according to CCITT draft recommendations when the external synchronization is lost (holdover mode). The tolerance for the plesiochronous 140 Mbit/s stream is 15 ppm. With these values the stuffing ratio varies between 0.71 and 0.85 . Investigations [1] show that the resulting jitter is easily attenuated by a phase locked loop.

4.4.2 Fixed Jitter

This jitter arises as the Section and Path Overheads (SOH,POH) of the SDH- signals have to be removed. As the frame is 8 kHz there are jitter components at 8 kHz and multiples.

4.4.3 Pointer Adjustments in the SDH

If there exist differences in frequency or phase between the incoming STM- signals and the local node a pointer adjustment mechanism takes place. In a synchronous transport module the pointer adjustment allows a 3 byte positive-/negative- justification. By this

mean we have no loss of data. The pointer (H1,H2 byte) which indicates the position of the Virtual Container (VC-4) will be incremented or decremented by one. The VC-4 is composed of the mapped 140 Mbit/s- channel and the associated 9 path overhead bytes.

4.4.3.1 Negative Justification

If the frequency of the incoming STM-1 stream is higher than the local clock frequency the 3 H3 bytes of the outgoing STM-1 signal are used for data transmission. Consequently the pointer value is decremented by one. This corresponds to an advancement of the virtual container (i.e. 140 Mbit/s- signal) by 3 bytes. Only every forth frame this process can occur. In annex 3 the scheme of this negative justification is shown.

With a stability in frequency of 20 ppm (degraded or holdover mode) every 7 ms a pointer adjustment, which correspond to a 24 bit phase step for 140 Mbit/s- signals, can occur. Lower bitrates which are mapped into the SDH undergo only 8 bit phase steps. Therefore the clock distribution at a bitrate of 2048 kHz over the SDH- network has take in account these 8 bit- time steps.

With a relative stability of 10^{-11} the time interval would be about 4 hours.

4.4.3.2 Positive Justification

The case of the positive justification is shown in annex 4 for in-formation. Here the frequency of the incoming STM-1 is lower than the local clock. The 3 bytes after the last H3- byte are used for the positive justification (i.e. not used for data transmission). Normally these bytes carry information. Consequently the pointer value is incremented by one.

4.4.3.3 Effects of a large wander on pointer adjustments

For illustration purposes in annex 5 the behaviour of the pointer adjustment mechanism on a large saw tooth incoming wander is shown [7]. For the pointer buffer, which decides on pointer adjustments, a threshold hysteresis of 12 bytes has been defined in a draft CCITT recommendation for 140 Mbit/s- signals. The corresponding values for 2 and 34 Mbit/s- signals should be at least 2 and 4 bytes. As one can see in annex 5 this threshold attenuates the incoming wander but on the other hand the phase steps produce large jitter spectra.

By the above procedure the dependency of the characteristics of a synchronous network with several nodes on clock parameters, synchronization strategy and wander can be simulated.

4.4.4 Control of the Jitter in the SDH

At the border between the synchronous hierarchy and the existing plesiochronous hierarchy a desynchronizer has to attenuate the high phase steps of 24 bits at 140 Mbit/s and 8 bits for lower bitrates. Different schemes have been presented. There exist two categories of possible desynchronizers.

4.4.4.1 Non- Phase Spreading Desynchronizers

In this case a phase- lock loop of a very low bandwidth filters directly the incoming phase step. A simulation of an analog first order phase- lock loop (PLL) ([2],[3]) with the method of Runge- Kutta ([4],[5]) with a bandwidth of 0.5 Hz (should be realized with a digital PLL) has been made.A measurement filter (high pass) with a cutoff frequency of 200 Hz (first order analog PLL) is in cascade to the first PLL. The following conclusions can be drawn from the time response of the resulting output jitter to a 24 bit phase step as input (see annex 6). The jitter- amplitude of 24 [UI] (bit unit intervals) can be reduced substantially. In this example to 0.12 [UI]. On the other hand it needs a large amount of decay time due to the low bandwidth of the first PLL (0.5 Hz). Therefore following pointer adjustements in the same direction, which arrives before the transients have died out, can produce an accumulation in the resulting jitter- amplitude.

4.4.4.2 Phase Spreading Desynchronizers

The fixed and adaptive bit- leaking desynchronizers spread out the incoming phase steps over a certain time interval with reduced phase steps. After each reduced phase there will be a waiting time before the next step is applied. By this way the transients of each step can be reduced by filtering. The so called fixed-bit leaking method uses a fixed waiting time interval. In the adaptive bit- leaking mode the waiting time interval is adjusted depending on the difference in frequency (i.e. time interval between pointer adjustments). In the limit the phase- spreading can be achieved by a digital PLL with a

low bandwidth which is followed by a PLL of moderate bandwidth (order of 100 Hz). These methods have to be analyzed further.

5. Conclusions

This presentation shows that the interworking between the existing plesiochronous digital network and a future synchronous network poses stringent requirements upon the design of desynchronizers. Further the synchronization strategy, clock parameters, clock distribution over a SDH network, jitter and wander planning have to be developed carefully.

References

[1] M.Robledo, R.Urbansky :Jitter bei der Uebertragung plesiochroner Signale in der SDH
PKI Technische Mitteilungen, 3/1989

[2] A.Blanchard :Phase-Locked Loops: Application to Coherent Receiver Design
Wiley-Interscience, 1976

[3] W.C.Lindsey :Synchronization Sytems in Communication and Control
Prentice Hall, 1972

[4] W.Gander :Computer Mathematik
Birkhäuser Verlag, 1985

[5] D.G.Schultz J.L.Melsa :State Functions and Linear Control Systems
McGraw Hill, 1967

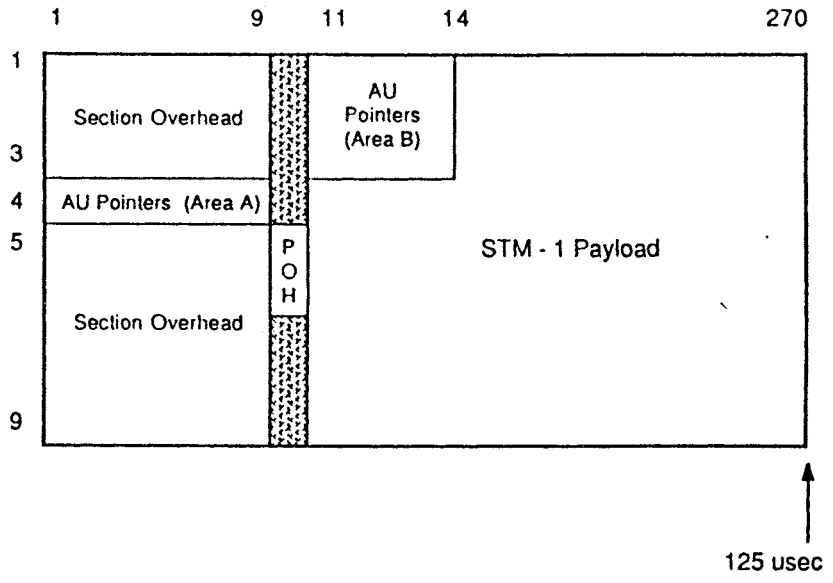
[6] P.Kartaschoff P.A.Probst P.Vörös :A network timing concept for Switzerland
Congrès SSC-ASMT 1985, Interlaken

[7] CCITT SG XVIII:Information on the Specification of SDH Clock Requirements and Pointer Arrival Statistics.
Contribution D-395, Nov.1989

TABLE 1:

SOURCES OF SLIPS AND MEANS TO REDUCE THEIR OCCURENCE

Parameters	possible means
NETWORK CLOCKS	. synchronization of the network node clocks . plesiochronous operation with high stability clocks
PROPAGATION DELAY OF SWITCHING AND TRANSMISSION EQUIPMENTS	. frame aligner (see Fig. 1) . specifications of tolerable limits for Jitter and wander at the equipment interfaces
INTERFERENCES	. allocation of a part of the overall network performance objectives (slip rate) to the effects of interferences



	1	2	3	4	5	6	7	8	9	
1	A1	A1	A1	A2	A2	A2	C1	X	X	J1
2	B1	X	X	E1	X	X	F1	X	X	B3
3	D1	X	X	D2	X	X	D3	X	X	C2
4	AU Pointers									G1
5	B2	B2	B2	K1	X	X	K2	X	X	F2
6	D4	X	X	D5	X	X	D6	X	X	H4
7	D7	X	X	D8	X	X	D9	X	X	Z3
8	D10	X	X	D11	X	X	D12	X	X	Z4
9	Z1	Z1	Z1	Z2	Z2	Z2	E2	X	X	Z5

A1, A2	Framing bytes
B1, B2, B3	Byte Interleaved Parity code for bit error monitoring
C1, C2	STM Identifier, Signal label
D1-D12	Datacommunication
E1, E2	Orderwire
F1, F2	User channels
G1	Path status
H4	Multiframe Indicator
J1	Path trace
K1, K2	Automatic Protection Switching
Z1-Z5	Spare
X	Bytes reserved for national use

STM - 1 Frame Structure of the SDH

Annex 2

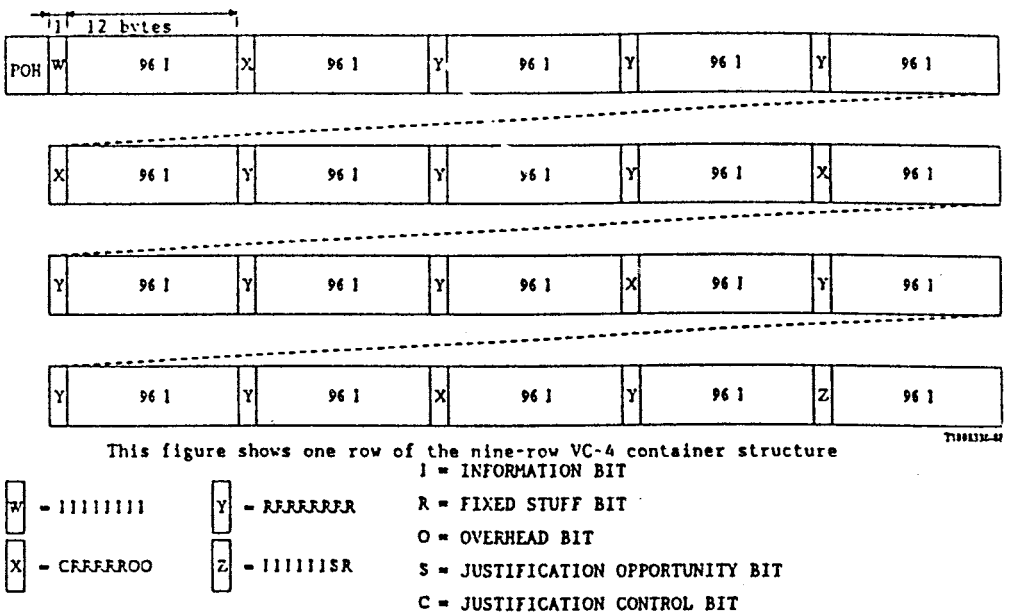
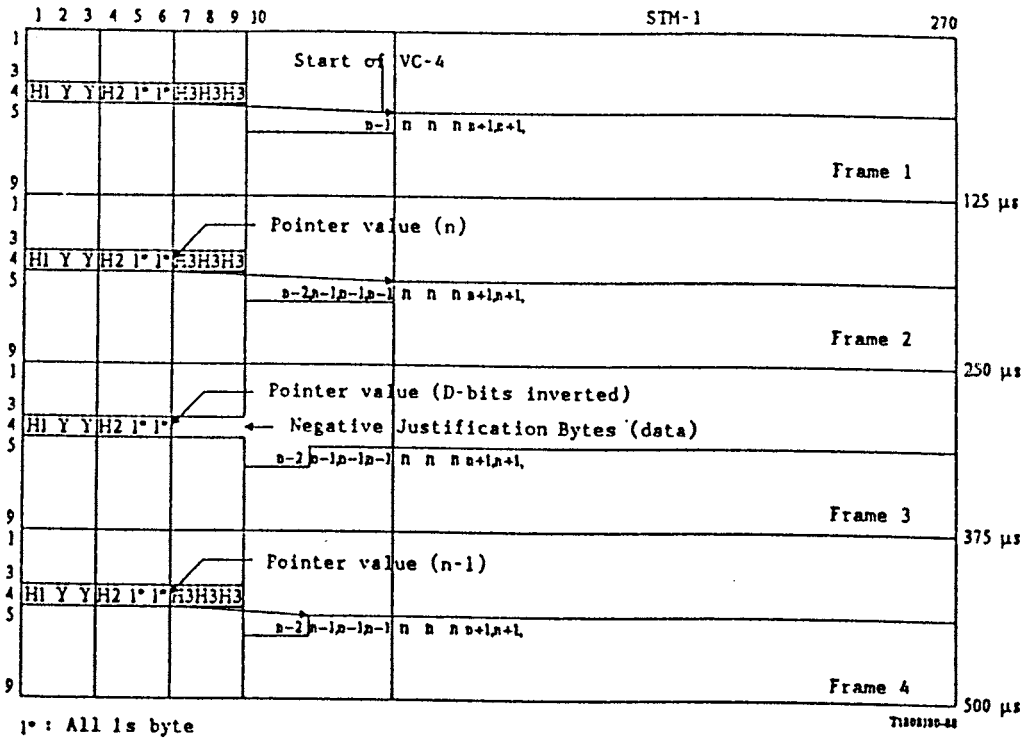


FIGURE 5.3/G.709

Mapping of asynchronous 139 264 kbit/s tributary into VC-4

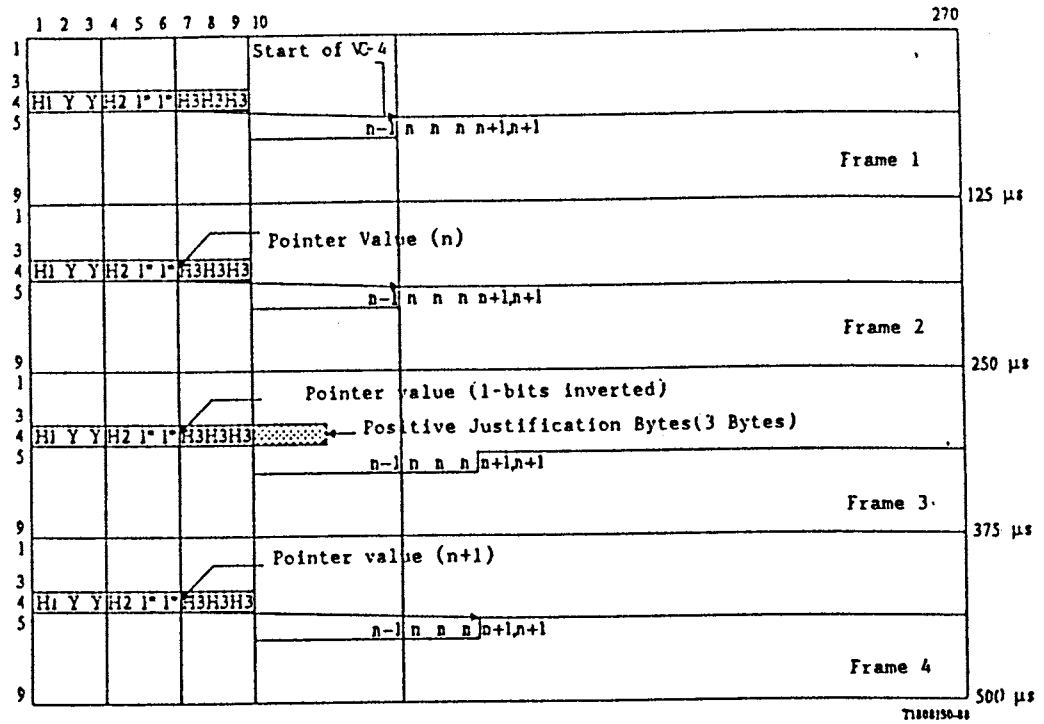


1* : All 1s byte

Y : 1001SS11 (S bits are unspecified)

FIGURE 3.8/G.709

AU-4 pointer adjustment operation - negative justification

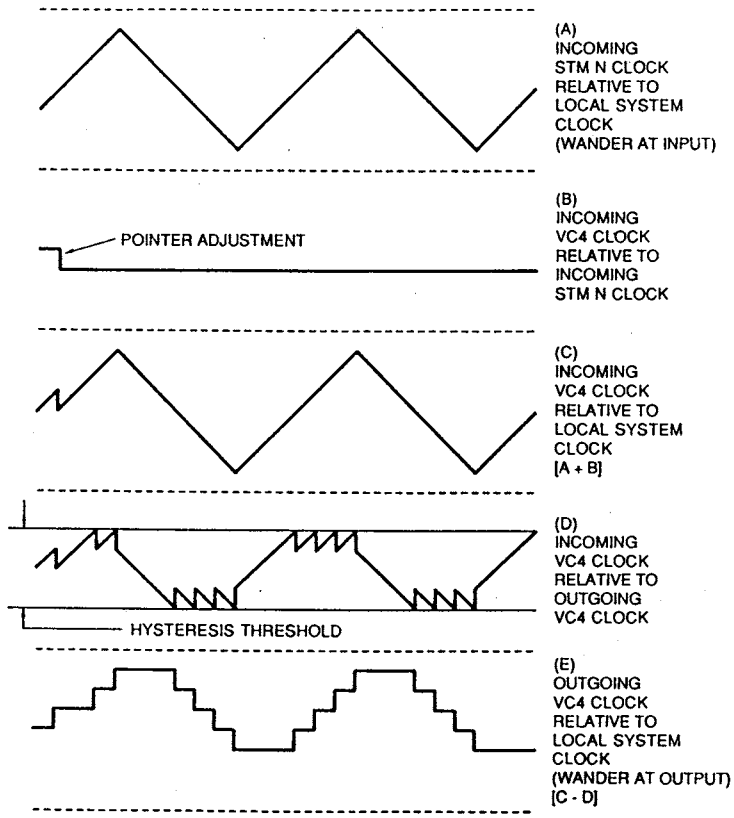


1* : All 1s byte

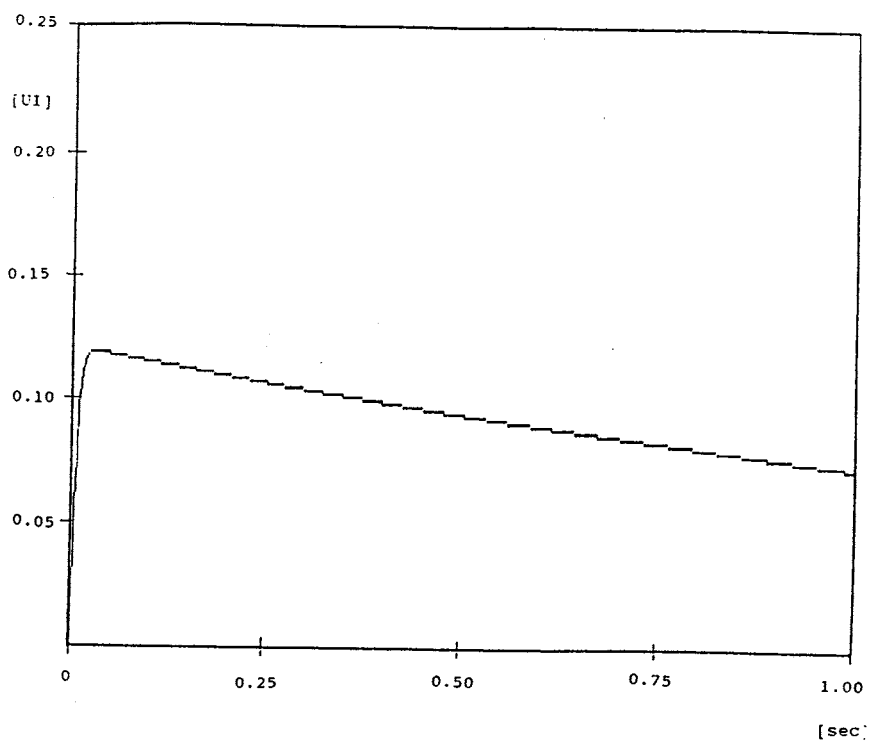
Y : 1U01SS11 (S bits are unspecified)

FIGURE 3.5/G.709

AU-4 pointer adjustment operation - positive justification



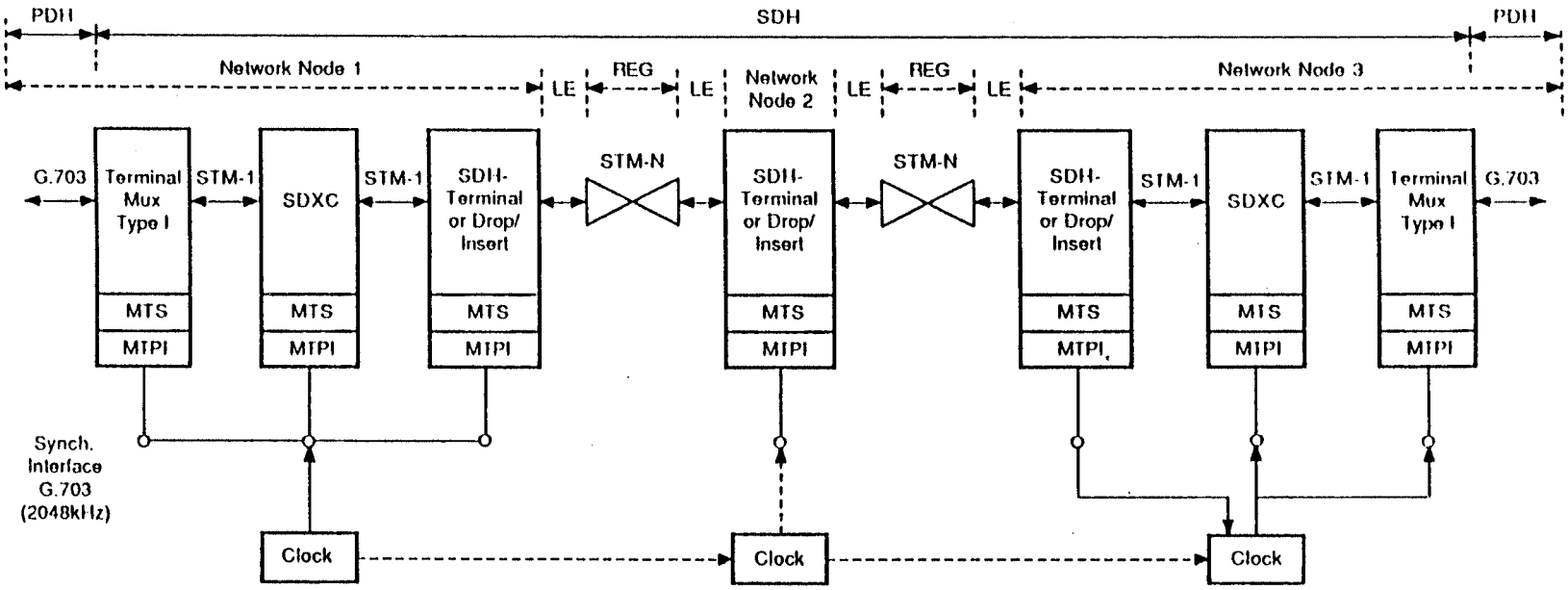
Effect of Wander on Pointer Adjustment in a Network Node



Output Jitteramplitude [UI] vs. Time [sec]

Synchronous Network Topology

Reference model



Legend

LE	-	Line Equipment (optical, radio)
PDH	-	Plesiochronous Digital Hierarchy
MTPI	-	Multiplex Timing Physical Interface
MTS	-	Multiplex Timing Source
REG	-	Regenerator
SDH	-	Synchronous Digital Hierarchy
SDXC	-	Synchronous Digital Cross-Connect
STM-1	-	Synchronous Transport Module of Order 1 (i. e. 155Mbit/s)
STM-N	-	Synchronous Transport Module of Order N (i. e. N•155Mbit/s)

H.J. Bosshard, VL 26
13 March 1990

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-03 Number 3

Distributed Master Clock System for Synchronization of a 2048 kHz Telecomm Network

Steven M. Bass, Ralph E. Simons, David W. Martin

Frequency and Time Systems, Inc.
34 Tozer Road
Beverly, MA 01915

Abstract

A synchronization system has been developed to distribute timing in a multi-national digital communications network. Three reference clocks in separate locations are used to enhance the survivability of the timing system. An inter-clock communciation protocol is used to automatically select one reference clock to provide the master timing for the network. The communciations channel also enables the lower level clocks to switch away from a timing source which has experienced a failure.

Introduction

A novel synchronization system has been developed to provide the timing for a digital telecommunications network in the eastern Caribbean. The network is shown in Figure 1. Using a digital microwave backbone, this network provides digital communication facilities among the islands and to the USA and Europe. Providing common timing to a digital network which spans many nations presented some unique challenges.

The network required its own source of reference timing which meets the CCITT guidelines in recommendation G.811. The primary requirement for the reference clock is a frequency accuracy of 1e-11 or better.

Multiple reference clocks are required, to ensure that single failures do not cripple the network. The spatially distributed nature of the network and the linear structure meant that separate reference clocks at each end would provide the most reliable timing.

The system design mandated that the timing be derived from a single master clock during normal operation. Other reference clock sites, and all the other clocks derive their timing from the master reference clock. When failures occur the reference clocks must automatically choose a new reference clock to provide the master timing. Some faults can be detected by monitoring the incoming data streams. Frequency errors and signal faults (loss of signal (LOS), alarm indication signal (AIS) and loss of frame (LOF)) are indications that the incoming signal is unacceptable to use as a reference. However, monitoring the line will only detect gross failures. Additional information must be passed between the clock sites to allow each clock to know the status of the remote clocks so it can react appropriately.

A key feature of the timing system design was to provide the best timing possible under all fault scenarios. The system has to choose the proper timing source in the presence of failures of the local clock hardware, the clock hardware at other sites, or in the transmission path between the sites.

The network provides timing for all the clocks in the system in a hierarchical fashion. To ensure that the lower level clocks receive valid timing, the signals from the co-located exchange are retimed to the system clock before leaving the site. Frequency monitoring of the signals from the exchange provide indications of potential problems in the central office.

System Design and Clock Hierarchy

The timing network is hierarchical, and contains clocks at Levels 1, 2, and 3. The Level 1 clocks have high quality quartz oscillators and can be equipped with cesium standards to provide the master reference clock. The Level 2 clocks use the same high quality quartz oscillators, but are simpler in design and cannot accept a cesium standard. The Level 3 clocks utilize lower quality quartz oscillators, with lesser thermal stability and higher aging. The Level 3 clocks were not part of this development program.

The timing network contains three Level 1 clock systems, two of which are equipped with cesium standards. There are two Level 2 clock systems and several Level 3 clock systems. The interconnections between these clocks is shown in Figure 2. The cesium equipped Level 1 reference clocks are at Tortola and Barbados. Antigua is presently not equipped with a cesium standard. Addition of a cesium standard at Antigua would bring it to the same condition as the other Level 1 clocks, further enhancing the network reliability. The Level 2 clocks are at St. Lucia and Trinidad, with Level 3 clocks at the other sites.

Signalling using bits embedded within the 2048 kbps data stream is used to communicate status among the clock systems. This signalling is essential to allow the clocks to coordinate their operation properly.

The reference clock site with the highest priority will provide the master clock. The priority for the Level 1 clock sites is Tortola, Barbados, Antigua. The other sites will derive their timing from a data stream originating at the higher priority site.

Normally, Tortola will derive its timing from its local cesium standard. Barbados and Antigua will extract timing from the data streams originating at Tortola. A failure at Tortola which prevents Tortola from acting as the master clock will cause the clocks to reconfigure. Barbados will utilize its cesium standard to generate the reference timing. Tortola and Antigua will synchronize to Barbados. While Antigua is currently not equipped with its own cesium standard, the logic has been implemented to allow Antigua to serve as the master clock, if Tortola and Barbados fail. The addition of a cesium standard would provide the timing system with three independent sources of reference timing.

To prevent the system from unnecessarily changing to a different master clock, a delay procedure has been implemented. When signalling from a remote site indicates that a Level 1 clock should operate from its local cesium standard, it first runs off its local quartz oscillator in holdover. It will stay in this state for 0 - 6 hours (operator selectable). At the end of this delay, if the condition still exists which requires switching to a new master clock, it will do so. If the original problem has been fixed in the interim, the clock will return to its original state.

Each clock system can be operated manually to support maintenance and trouble-shooting. The operator can select any of the possible sources to provide the master timing. Manual control of the outgoing signalling information is also supported. This allows the operator to operate one system manually, perhaps to diagnose a failure, without affecting operation at the other sites. Manual operation must be used with care. Improper operator actions can potentially cause the whole digital network to derive its timing from the wrong source.

The Level 2 clocks accept inputs from two different Level 1 clocks. At these sites priority is given to the signal which is designated as the Primary input. A Synch Fail message (defined below) or unacceptable quality on the incoming data stream will cause the clock to use the other line as the reference.

Signalling

Since each clock can derive its timing from more than one source, the status of the upstream clocks needs to be communicated between sites. This information is embedded in the 2048 kbps data streams, using bits reserved for national use. Figure 3 shows the structure of 2048 kbps data stream indicating the bits allocated for national use. Three bits have been designated for communications between the Level 1 clock sites. These bits are designated the Synch Fail (SF) bit and the Global State determination bits (GS1 and GS2). The SF bit is set when the source clock timing is not traceable to a cesium standard. The GS bits are used with the SF bit to more fully describe the state of the source clock timing. Table 1 describes the messages used for communication between the Level 1 clocks.

The Level 2 and 3 clocks use only the SF bit for communication.

SF	GS1	GS2	Name	Meaning
0	0	0	MASTER	using local cesium as source
0	0	1	SLAVE	locked to clock which is MASTER
0	1	0	SEC SLAVE	locked to clock which is SLAVE
0	1	1	FR CD	free-running during reconfiguration delay
1	0	0	FAIL	undefined message
1	0	1	FAIL	undefined message
1	1	0	FREERUN	using freerunning oscillator as source
1	1	1	FAIL	clock unable to provide valid timing

Table 1. Synchronization Message Definitions

The MASTER message is sent by a clock which is running off its local cesium standard. Normally only one clock should send this message. During reconfiguration changes or when faults occur, multiple clocks may run as MASTER simultaneously.

The SLAVE message is used by clocks which are deriving their timing from a source which is sending the MASTER message.

The SEC SLAVE message is used to prevent timing loops. A clock which is locked to another clock which is sending SLAVE will send the SEC SLAVE message. The other clocks will not lock to a data stream which has the SEC SLAVE message embedded in it. This

prevents the system from entering a state where two or more clocks are locked in a circular fashion to each other, without any master.

FREE RUN is sent when a clock is in holdover. Since the source clock is not traceable to a reference clock, the SF bit is set.

The FR CD message is a special message used when the Level 1 clocks are in holdover during a reconfiguration delay. This message is separate from the normal FREE RUN message, specifically so the SF bit won't be set. While the Level 1 clock is not traceable to a cesium standard during the delay, the highly stable oscillators used in the clock will not drift significantly during this period. Setting the Synch Fail bit would cause the lower level clocks to switch inputs or enter holdover. This would be undesirable, as their oscillators might be of lower quality then those in the Level 1 clocks.

FAIL is sent whenever a clock is unable to provide acceptable timing or is unable to determine if the timing it produces is valid.

Reference Selection

Each clock site autonomously selects the appropriate source (local cesium, one of the two remote clocks, or holdover) to provide the timing at that site. The criteria used to make this determination are the key to making the whole timing network operate properly.

The process of choosing the source to use follows several stages. The local hardware is interrogated to determine which sources can be used at the moment. With this information and the signalling which is being sent from the other clocks, a Level 1 clock selects a timing source by following the rules listed below.

> Lock to a MASTER site, if there is one of higher priority. If there is no MASTER site of higher priority, use the local cesium if possible.
>
> If unable to use a MASTER as source (remote or local), lock to highest priority remote site which is generating SLAVE.
>
> Do not lock to a signal which is sending SEC SLAVE, FREE RUN, FR CD, or FAIL.
>
> If no acceptable source is available, enter holdover.

These rules decide the final state of the clock. The exact sequence of events as a clock enters or leaves a reconfiguration delay can differ.

The Level 2 clocks use a simpler rule. If the Primary signal is acceptable (no frequency offset and no LOS, AIS, or LOF) and the SF bit isn't set, the clock locks to it. If it isn't acceptable, the clock locks to the input from the other Level 1 site. If neither is acceptable, the clock enters holdover.

Alarm Philosophy

The clock system supports two types of alarms, Prompt and Deferred. Prompt alarms require immediate attention, while Deferred alarms can be handled when time is available. There are two rules which determine if an alarm is Prompt.

> Any fault which affects data streams which pass through the clock system is Prompt.
>
> Any fault or combination of faults which brings the system to within one failure of losing the ability to provide cesium traceable timing is Prompt.

Any faults which do not meet either of these requirements is Deferred.

Normally, only failures at the local site are considered when determining if the alarm is Prompt. The clocks are on different islands and failures elsewhere cannot be fixed by the local personnel. The system architecture was used to determine that some conditions should always cause Prompt alarms. Since the clock network only has two possible master clocks (Tortola and Barbados), a failure of a cesium standard causes a Prompt alarm.

Clock Architecture

All the Level 1 and Level 2 sites utilize the same basic hardware. Figure 4 shows the units that comprise a Level 1 clock system. There are two Timing Signal Generators (TSG) at each site. The TSGs receive the data stream from the other sites, extract the embedded signalling, and lock their oscillator to one of the streams. They also monitor the frequency of the cesium standard against the incoming data stream to detect failures of the cesium standard or the incoming stream.

The cesium standard generates the reference frequency used by the reference clock.

The Timing Signal Monitor accepts signals from the local exchange, retimes them, and transmits the data to the network. It embeds the appropriate signalling information to inform other clocks about the state of the local clock system. The TSM is present in all the Level 1 clocks and is optional for the Level 2 clock sites.

The Signal Processor (SP) performs two functions. It accepts three 10 MHz signals from the two TSGs and the cesium standard. These are converted to 2048 kHz signals, combined, and distributed as 20 timing outputs. The SP also monitors and controls the TSGs, cesium standard, and the TSM. It implements the global logic to enable the whole timing system to automatically select the appropriate master clock.

Alarm information is gathered and condensed by the Alarm Panel (AP), providing a summary point which details the state of the clock. The AP is also used to manually select a master source and set the outgoing signalling bits.

The Level 2 clock systems are similar to the Level 1 clocks, but do not contain a cesium standard. Only one of the Level 2 clock systems contains a TSM.

Timing Signal Generator

The main function of the FTS 3800 Timing Signal Generator is to lock a high quality quartz oscillator to one of the incoming data streams. The TSG looks at the quality of the signal, both in frequency and for signal faults such as LOS, AIS, or LOF. If the incoming signal does not meet the requirements, it is rejected as a synchronization source. The TSG also extracts the signalling information from the incoming data stream. This data is used by the Signal Processor to determine which source to use to generate the timing.

The TSG monitors the frequency of the cesium standard, as an additional verification that the cesium standard is operating properly.

Figure 5 shows a block diagram of the FTS 3800 TSG used in this system. The TSG is modular and 4 different modules are used in this system.

The CEPT Input modules accept a G.703/6 data stream at 2048 kbps. The signal is checked for LOS, AIS, and LOF. The module extracts a clock from the incoming stream. The phase of this clock is then compared to the phase of the oscillator.

In addition, the CEPT Input module extracts the national bits from the incoming stream. The bit state is passed on to the Processor module. The CEPT module transmits the data stream back out after setting the signalling bits to 1.

The Sine Input module accepts a 1, 5, or 10 MHz signal. The phase of this signal is compared against the oscillator phase. The Sine Input module also accepts a TTL signal which can be used to indicate a fault in the source of the input signal.

The Processor module controls the operation of the TSG. It collects the data gathered by the input modules (phase, line faults, signalling bits, etc.), and uses this data to decide how to update the oscillator. The Processor implements a phase lock loop in software tó control the oscillator. Time constants in the range 20-10000 seconds (operator adjustable) are supported.

The Processor generates alarms when faults occur, either line faults, frequency errors, or hardware failures. The TSG also communicates with the Signal Processor (SP) and Alarm Panel (AP), which normally control the operation of the TSG. If the TSG cannot communicate with the SP and AP, it will autonomously decide how to operate. The TSG passes information to the SP and AP, indicating the state of its hardware, the received signalling information, and its current mode of operation (e.g. locked to the Primary channel). It also receives commands from these units. These commands determine which channel to use as a reference and when to enter holdover.

The TSG has a RS-232 port for operator monitoring and control. The operator can use the serial port to gather data on the frequency of the inputs, determine the cause of alarms, and change the values of the configuration parameters. The operator can also control the operation of the TSG, choosing the input to use as the reference and the oscillator operational mode (e.g. put the oscillator into holdover). Both the AP and SP have higher priority than the RS-232 port, so commands entered by the operator only have effect if these connections have failed.

The Oscillator module contains a high quality quartz oscillator. The aging rate is less than 5e-11/day and the thermal stability is about 2e-12/°C. The oscillator generates 18.528 MHz and 1 pps signals which are used by the input modules to perform the phase comparison. A 16 bit DAC is controlled by the Processor module to adjust the oscillator frequency. The output from the module is a 10 MHz sine wave, which goes to the SP.

Timing Signal Monitor

The FTS 3850 TSM is a variation on the TSG. It supports the same CEPT Input module and Processor module. Its main function is to retime the output of the local exchange, to guarantee that the local clock timing will be distributed to the other sites. It also embeds the signalling information into the data stream. The frequency of the signals from the local exchange is monitored; errors can be quickly detected and reported to maintenance personnel.

Figure 6 shows a block diagram of the TSM. The TSM contains a Processor module, a Reference module, and up to 14 CEPT Input modules.

The CEPT input modules are identical to those used in the TSG. A jumper on the module changes the module's operation so the signal is passed through an elastic buffer and is retimed to the clock timing. The buffer is monitored for slips, which are reported to the Processor module.

The CEPT input module updates the signalling bits, as commanded by the Processor module. Some modules are assigned to communicate with the Level 1 sites and the others are set for the Level 2 or 3 sites. Only the Level 1 paths are updated with the 3 bit signalling information. The other channels send just the SF bit.

The Processor module is also identical to that used in the TSG, both in hardware and software. A jumper on the module selects whether it should operate as a TSG or TSM.

The TSM Processor serves the same basic function as in the TSG. It collects data from the input modules, generates alarms, controls the hardware, and communicates with the SP and AP. The communications link is used to indicate problems, and to receive the signalling bits to send onward.

The TSM also has an RS-232 port for maintenance and troubleshooting. Parameters can be set and CEPT input modules assigned to communicate with Level 1 or Level 2 or 3 sites.

The Reference module serves the same function as the Oscillator module in the TSG, i.e. it provides the reference against which the input signal frequency is measured. It does not contain an oscillator. Instead, it accepts a 2048 kHz input from the Signal Processor and generates the 18.528 MHz and 1 pps signals from this. The 2048 kHz signal is also used as the clock to which the CEPT Input modules retime the data streams.

Signal Processor

The FTS 3905 Signal Processor provides the central control and distribution for the clock system. Figure 7 shows a block diagram of the SP. There are three types of modules in the Signal Processor: Phase Synchronized Synthesizers (PSS), Processor, and CEPT Distribution modules.

The PSS modules convert the 10 MHz signals from the cesium standard and TSGs into 2048 kHz signals. These are then fed to the Distribution modules to provide timing to the local equipment. The system specification required that phase steps be smaller than 10 ns whenever the master source changed or modules were added or removed. The PSS modules were designed to implement this feature. The SP Processor module chooses one master from among the three PSS modules. The other PSS modules align the phase of their 2048 kHz output in 15 ns steps to maintain phase synchronization with the master. The PSS modules all drive their output onto a common line on the SP backplane, which acts to passively combine the signals. This provides a high degree of redundancy and security, as PSS modules can be added or removed at any time without disrupting the timing signal.

The SP Processor module utilizes the same hardware as the TSG and TSM hardware. Different software is used to provide the functionality needed. The Processor module has two main functions. One function is to control the PSS modules and monitor the SP hardware for failures. The other main function is to collect information from the other units in the clock (TSG, TSM, cesium standard, and AP) and to direct these units to achieve the desired configuration.

The SP Processor determines how to configure the clock system. It gathers data from the TSGs, TSM, cesium standard, and AP. From this it determines what the allowed choices are for the local master: cesium, TSG1 or TSG2. Using this information and the signalling received from the remote clocks the Processor then determines which source it should use to provide timing. The TSGs and PSS modules are appropriately commanded to create this state. The Processor also determines the signalling information which the TSM sends out. If the AP is set to manual operation, the Processor uses this to determine the proper operation, overriding the normal process.

The Processor also determines when faults are Prompt or Deferred. The SP Processor communicates with all the other hardware. It is the only place where all the information is available to determine if a set of failures should generate a Prompt alarm. The Prompt/Deferred status of an alarm is sent to the AP, to trigger the appropriate alarm relays and indicators.

The CEPT Distribution modules accept the combined 2048 kHz signal generated by the PSS modules. The signal is shaped to meet CCITT G.703/10 requirements and is output on 20 channels. The distribution modules are fully redundant. Adjacent modules' outputs are combined to drive the 20 channels redundantly. If a module fails, relays isolate the defective module from the output lines. The level of the second card in the pair is boosted slightly, to compensate for the lost module. There is no loss of output at any time, simply a brief decrease in amplitude. The defective module can be removed and replaced with no loss in signal availability.

Cesium Standard

The FTS 4060 provides an independent source of timing accurate to better than 1e-11. It has no interaction with any of the other units, except to indicate a failure if one occurs.

Conclusion

The timing system implemented for the digital network has been designed to automatically provide an independent source of timing. The system was designed to be highly redundant, to use separated reference clocks to ensure survivability, and to continuously monitor the operation of the network to reliably deliver high quality timing.

The use of separated reference clocks which automatically can select one member of the ensemble to provide the master timing for a hierarchical timing network is new. This is made possible by use of the signalling messages carried within the 2048 kbps data streams.

Carrying synchronization messages within a telecommunications data path is a subject which is of interest to European, American, and other telephone network planners. This application shows how embedded messaging can enhance the reliability and improve the operation of a telecommunication timing system.

Acknowledgement

The development of this system was carried out in collaboration with Redifon, Ltd. and we would like to acknowledge the contributions of Steve Chilver, Malcolm Blair, and Louise Scott whose work was an integral part of the program.

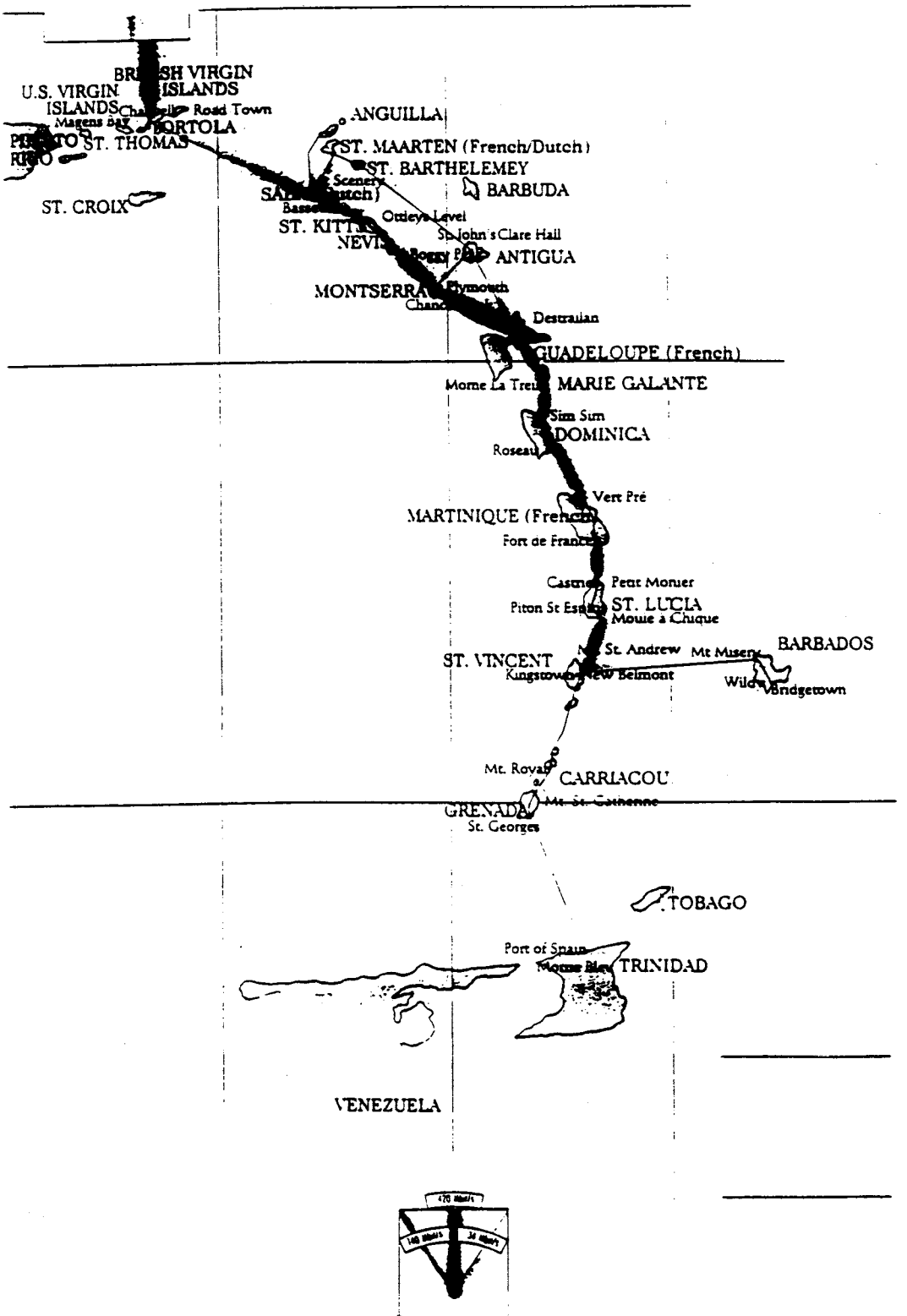


Figure 1. Caribbean Digital Telecommunication Network

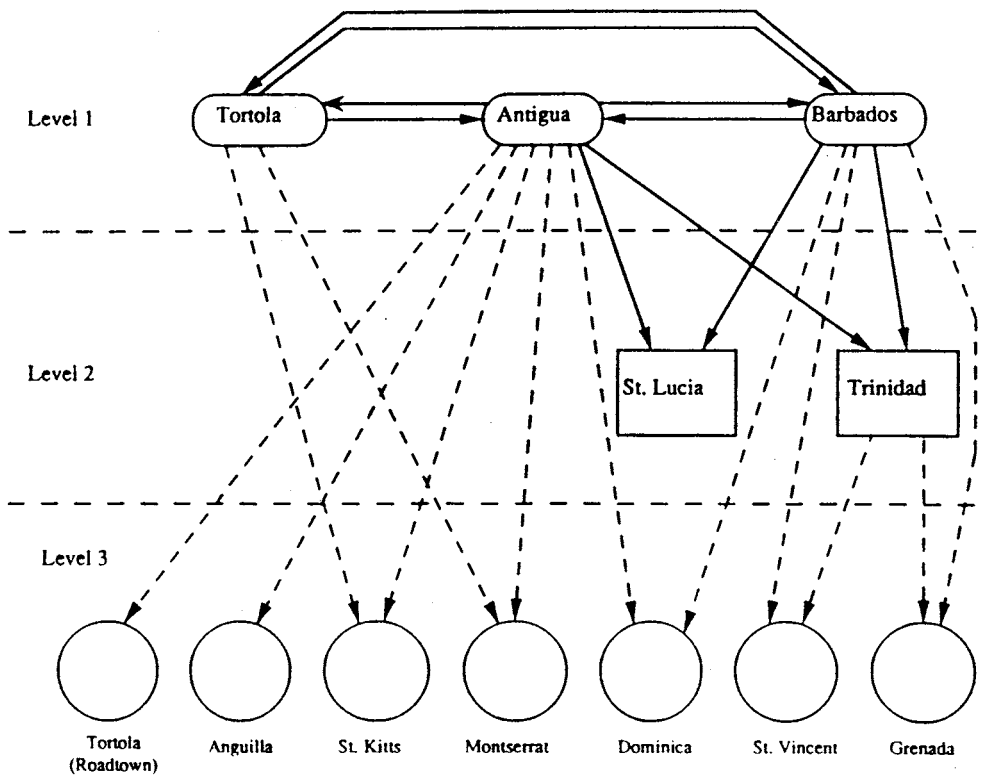


Figure 2. Network Timing Connections

Alternate frames \ Bit number	1	2	3	4	5	6	7	8
Frame containing the frame alignment signal	S_i Note 1	0	0	1	1	0	1	1
		Frame alignment signal						
Frame not containing the frame alignment signal	S_i Note 1	1 Note 2	A Note 3	S_n	S_n	S_n	S_n	S_n
				Note 4				

Note 1 — S_i — Bits reserved for international use. One specific use is described in § 2.3.3. Other possible use may be defined at a later stage. If no use is realized, these bits should be fixed at 1 on digital paths crossing an international border. However, they may be used nationally if the digital path does not cross a border.

Note 2 — This bit is fixed at 1 to assist in avoiding simulations of the frame alignment signal.

Note 3 — A — Remote alarm indication. In undisturbed operation, 0; in an alarm condition, 1.

Note 4 — S_n — Spare bits reserved for national use. The bits allocated for national use may not be used internationally. On a digital path crossing an international border or when not being used, they should be fixed at 1.

Figure 3. Allocation of Bits in Time Slot 0

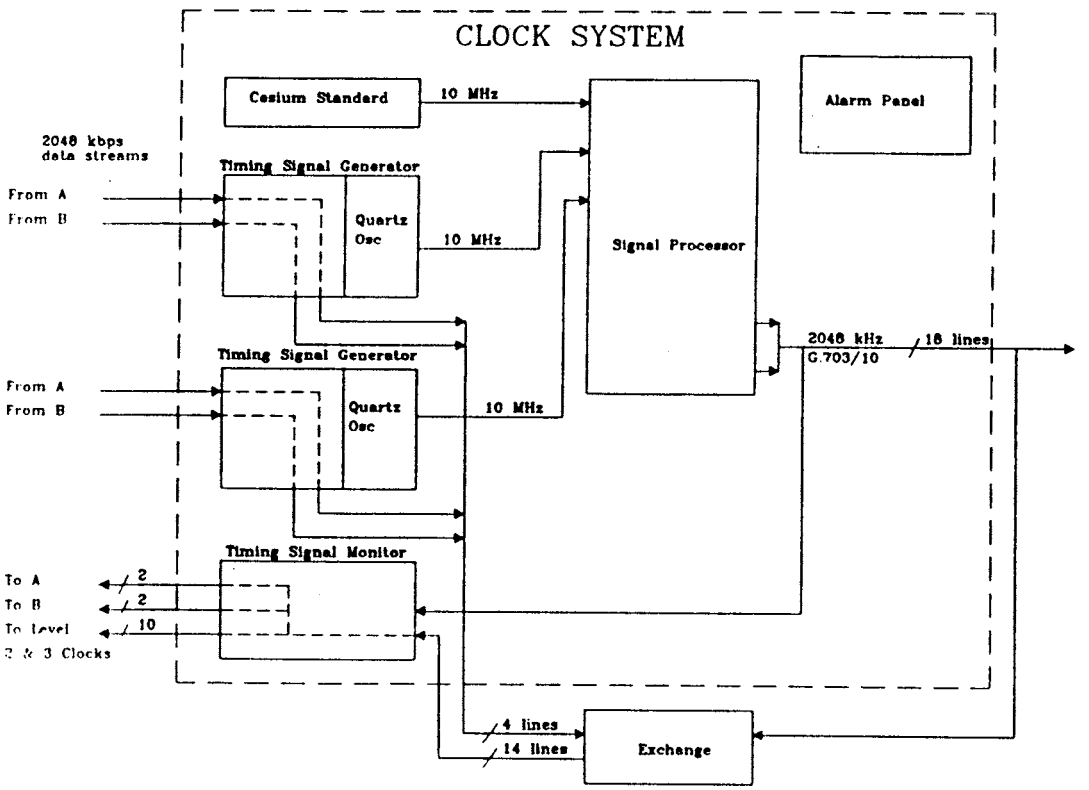


Figure 4. Level 1 Clock System

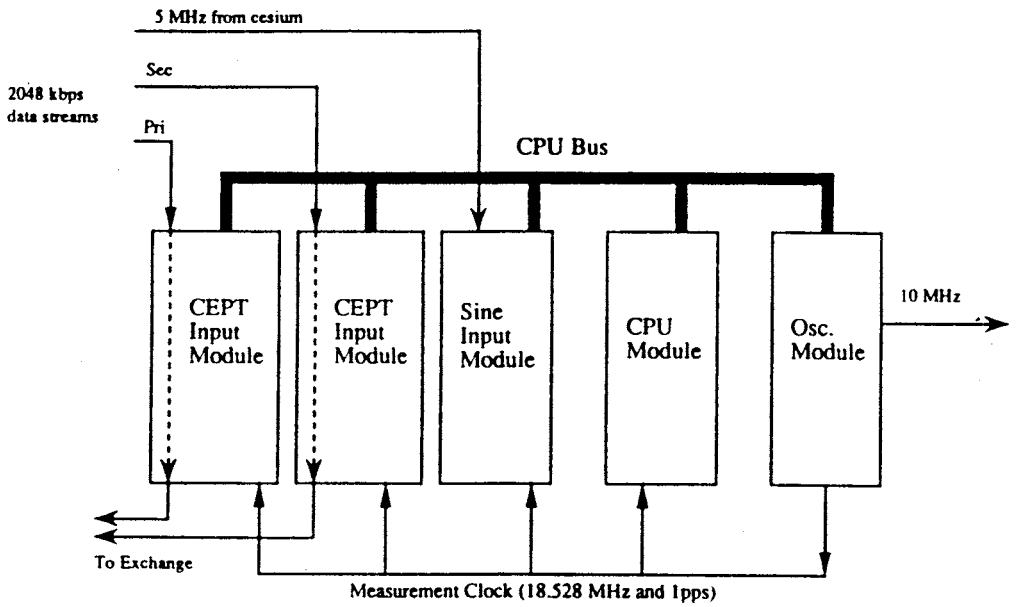


Figure 5. Timing Signal Generator

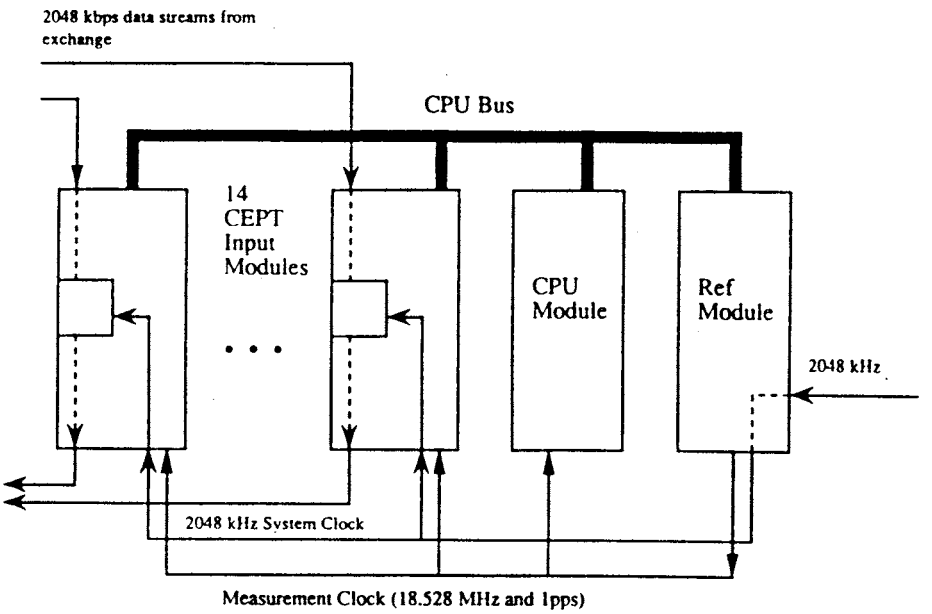


Figure 6. Timing Signal Monitor

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-03 Number 4

HYBRID SYSTEM OF COMMUNICATION AND RADIO DETERMINATION USING TWO GEOSTATIONARY SATELLITES

Shingo Ohmori, Yasushi Matsumoto and Eihisa Morikawa

Kashima Space Research Center
Communications Research Laboratory
Ministry of Posts and Telecommunications
Kashima, Ibaraki 314, Japan

Masayoshi Wakao

R & D Center for Radio Systems
Hirakawa, Chiyoda, Tokyo, Japan

Abstract

Communications Research Laboratory (CRL) of the Ministry of Posts and Telecommunications has developed a new hybrid satellite system which can provide both communications and positioning services in one system using two geostationary satellites. The distinctive feature is that location information can be provided by transmitting and receiving ranging signals over the same channel as communications through two geostationary satellites.

1. INTRODUCTION

Mobile satellite services are classified into two categories, one being "communications", providing voice, message, data transmission and so on, and the other being radio determination which gives information about location. In the past, these two satellite systems have been developed and implemented independently. However, in recent years, new satellite systems, which function both for communication and radio determination, have been studied by various organizations around the world.

Communications Research Laboratory (CRL) of the Ministry of Posts and Telecommunications has developed a new hybrid satellite system which can provide both communications and positioning services in one system using two geostationary satellites. The distinctive feature is that location information can be provided by transmitting and receiving ranging signals over the same channel as communications through two geostationary satellites.

The two types of terminals for the hybrid system developed by CRL are as follows;

Type A: in which one channel is used for both communication and positioning. A PN code (1023 chip rate) is used as a ranging signal and is transmitted over a single channel by OQPSK modulation.

Type B: in which the entire frequency bandwidth of SCPC channels is used. Communication and ranging information at 5 kbps, is spread by a PN code (255 chip rate) into a 1.2 MHz bandwidth and is transmitted by BPSK modulation.

This paper gives an the outline of system and the preliminary results.

2. EXPERIMENTAL SYSTEM

The experimental system consists of two geostationary satellites, ETS-V (150 E) and INMARSAT (180 E), a base earth station (Kashima Space Research Center) and mobile earth stations [3]. The frequencies between the satellites and mobiles are 1.6/1.5 GHz. In order to determine the position of a mobile precisely, ranging accuracy is an essential factor, which depends on the performance of the test equipment and the prediction accuracy of satellite orbits. In order to evaluate system performance, the following two ranging systems will be evaluated in the experiments.

(1) Two-way Ranging

As shown in Fig.1, a ranging signal is transmitted from the base earth station to a mobile station through the ETS-V satellite. The mobile station transmits the signal, which is repeated on IF or base band signal levels, to the base earth station through the ETS-V and INMARSAT satellites. The base station measures the time delay of reception between the respective signals from each satellite. In this system, the base station can determine the position of mobiles, so the mobiles are not required to have any complex functions for position determination. However, with this method satellite capacity is not efficiently used, because the mobile has to access to two satellites simultaneously.

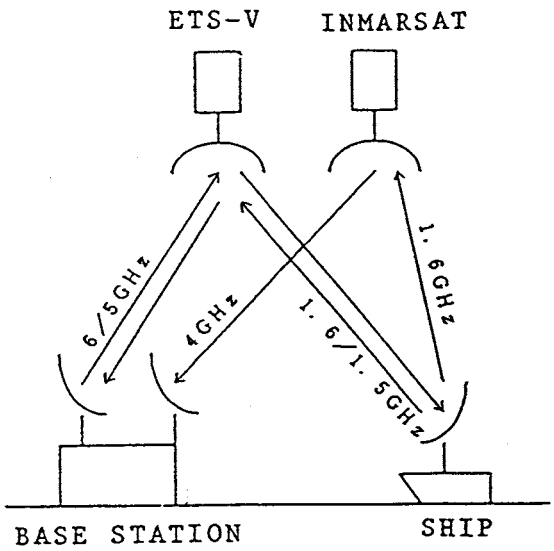


Fig.1 Two-way ranging.

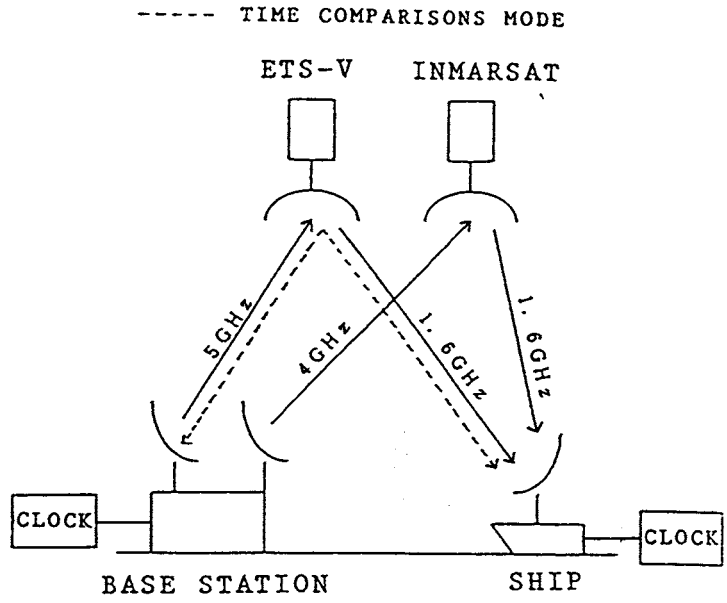


Fig.2 One-way ranging.

(2)One-way Ranging

As shown in Fig.2, the base station transmits the ranging signal, which is synchronized with a highly accurate clock, to a mobile station through the ETS-V and INMARSAT satellites. The mobile station measures the time delay with its clock, which is calibrated at appropriate intervals by the accurate clock through the ETS-V link. In this system, a mobile station can find its own position using the information of satellite orbits, which is provided by the base station through the satellite link.

Although, in this system, the mobile station has to have an additional function, whereby it can refer to the accurate time transmitted by the base station's master clock and adjust its own clock. Accordingly, this system can use a transponder more efficiently than the previous Two-way system.

3. EQUIPMENT CONFIGURATION

Two types of test equipment have been developed for the experiments with the hybrid satellite communication system. The first one is called the SCPC method, which uses one communication channel in a Single Channel Per Carrier (SCPC) access system for both communication and radio determination. The second one is called the SS method, which uses a Spread Spectrum technique to transmit and receive both the communication and ranging signals .

Both methods use digital modulation and demodulation techniques, and have ranging capabilities using a PN (pseudo-random noise) code. The Main characteristics of both methods are shown in Tablel.

Table 1. Method of modulation and parameters of ranging signals

	SCPC	SS
METHOD OF MODULATION	24kbps OQPSK	5kbps BPSK
SYMBOL RATE OF PN SIGNALS	12kbps	5kbps * 1.275Mbps
PN SIGNAL LENGTH	1,000 chips, 88.33msec	500 chips,100msec * 255 chips,0.2msec **

* PN SIGNALS FOR ELMINATING AMBIGUITIES
** PN SIGNALS FOR SPREADING SIGNALS

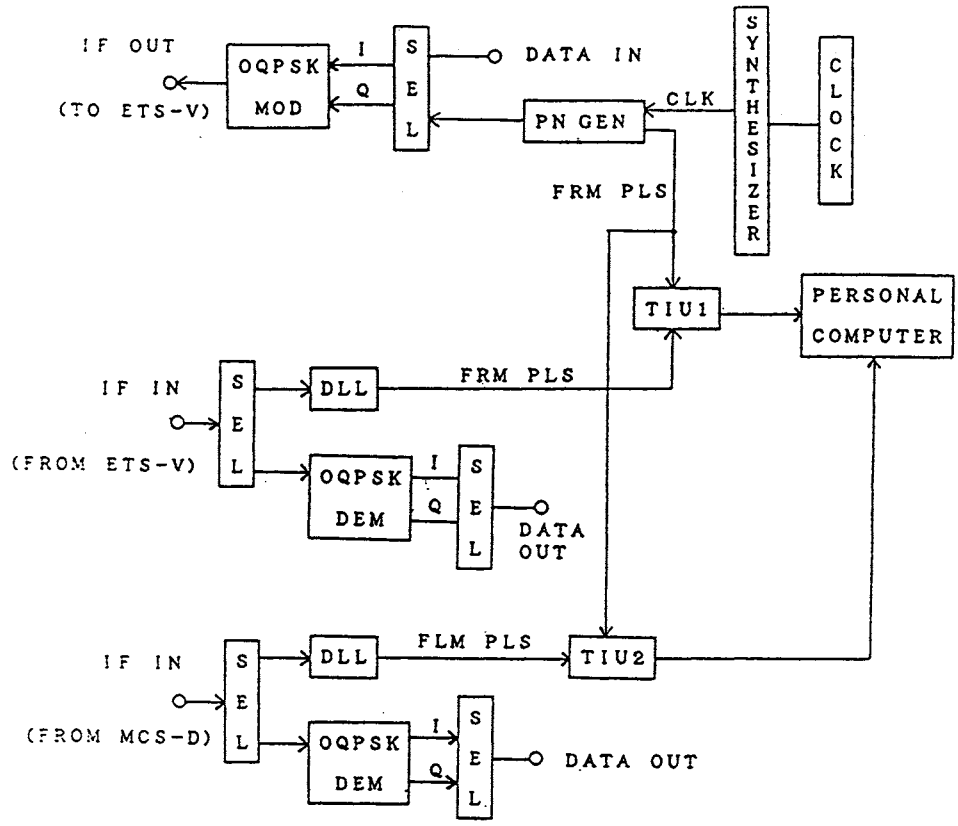


Fig.3 Base station block diagram.
(SCPC ranging system)

(1)SCPC method

Fig.3 is a block diagram of an SCPC terminal, which is connected to the RF section of the base station on IF signal levels. The PN code with 1023 chips is used for the ranging signal, which modulates and demodulates the carrier through a 24kbps-offset QPSK modulator (MOD) and demodulator(DEM). The I and Q channels of MOD/DEM are used for the communication signal and the ranging signal respectively. Synchronization of the received signal is locked by a Delay Lock Loop (DLL) with a Matched Pulse Detector (MPD), which consists of a digital correlation detector. The received PN signal is correlated with only 31-chip length of MPD, which allowing very fast acquisition of the initial carrier. The theoretical probabilities of mis-detection and false-detection of received PN signals are shown in Fig.4. If an appropriate threshold level of detection is chosen, for example 25-bit, the probabilities of mis- and false-detection are expected to be as negligible as 10-7 even in the case where Bit Error Rate (BER) is 10-2. A Time Interval Unit (TIU) measures the time delay of the received ranging signal, which is transmitted by the PN generator unit of the base station. The data are processed by computer to decide the position of a mobile station, taking account of the data of satellite orbits.

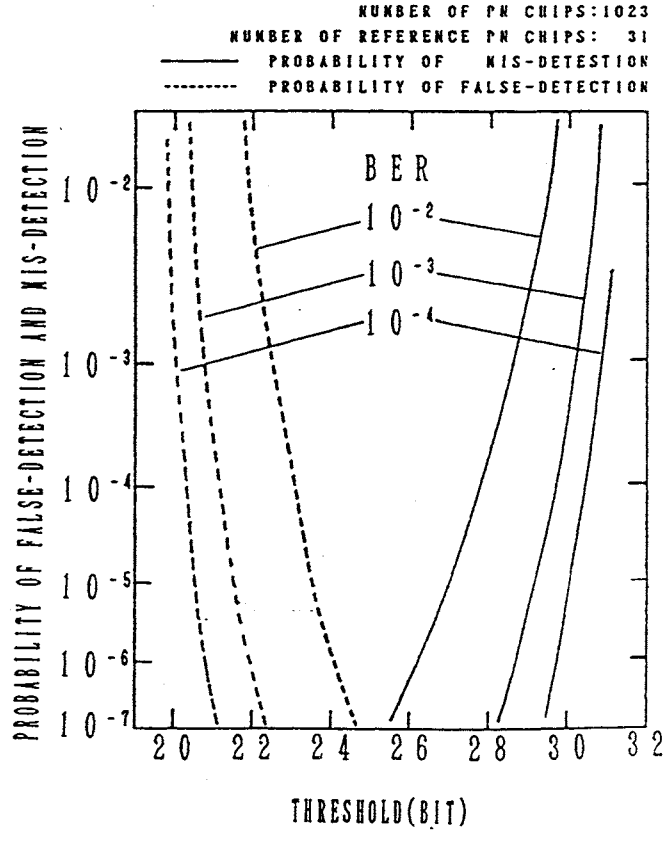


Fig.4 Probability of false- and mis-detection of matched pulse.

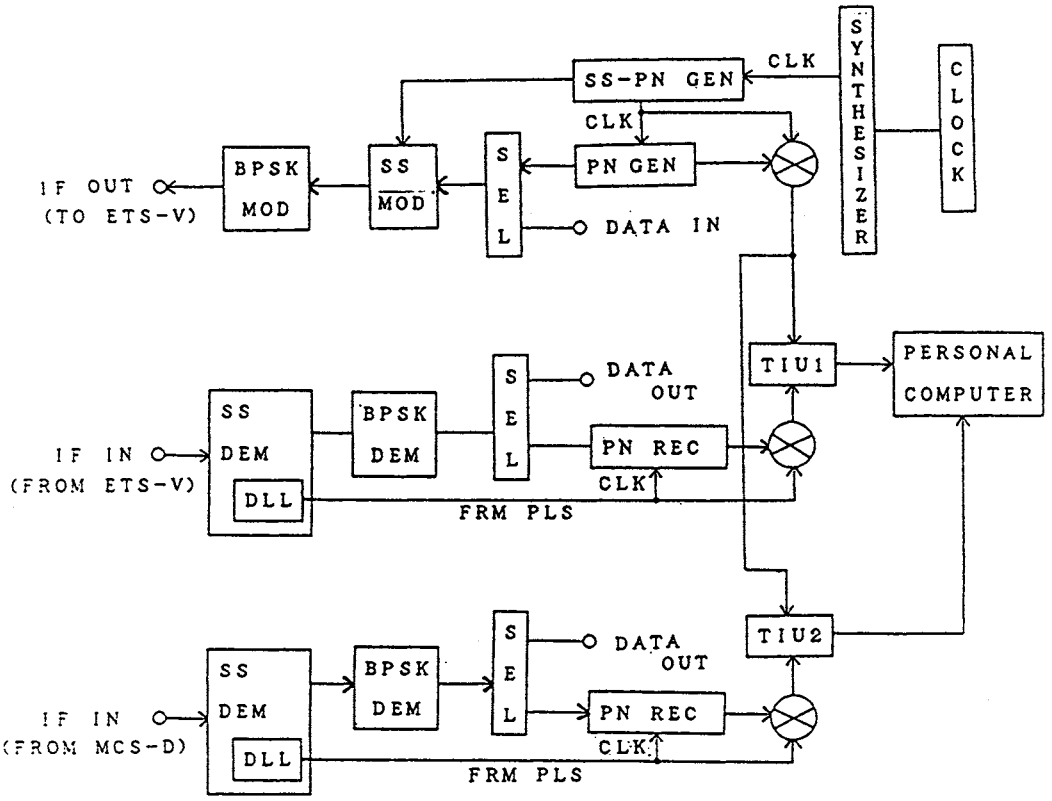


Fig.5 Base station block diagram.
(SS ranging system)

(2) SS method

Fig.5 is a block diagram of the SS test terminal, which is connected to the base station on IF signal levels. A base band signal of 5kbps is spread over a 1.2MHz frequency bandwidth by an SS modulator, and is transmitted through a BPSK modulator. The number of chips and length of frame of the PN code for the SS are 255 and 0.2msec, respectively. This method has some ambiguity in deciding the time delay, because the time delay through a satellite link is greater than that of a frame length of the PN code. In order to eliminate the ambiguity, a 5kbps PN code is transmitted first, and when the initial acquisition is completed, voice and/or data can be transmitted instead of the 5kbps PN code through the same channel. At a receiving section, the DLL is used to synchronize with the received PN code of 1.275Mbps, which is used for spreading signals. The Time Interval Unit (TIU) measures the time delay of both PN signals, which are used for signal spreading and ambiguity elimination respectively, and a computer calculates the position of a mobile station with taking account of the information of satellites.

4. Positioning Accuracy

(a) Ranging Accuracy

The major ranging error is caused by the tracking error of the DLL in a receiving unit. The standard deviation of phase detection error is given by the following equation [5].

$$\frac{\delta}{f_g}=\sqrt{\frac{B_L N_g}{2P}}$$

δ : standard deviation of phase detection error

f_g : clock frequency of ranging signal

B_L : equivalent noise bandwidth of DLL

Fig.6 shows the theoretical results of phase detection error of the DLL in the SCPC and SS systems, calculated by the above equation.

The SCPC system has the advantage of a simple configuration, but the ranging accuracy is worse than that of the SS system because of its ranging signal's lower clock frequency. On the other hand, the SS system is affected by other communication carrier signals in the SS band which reduce its channel quality C/No, the ratio of carrier to noise density. Fig.7

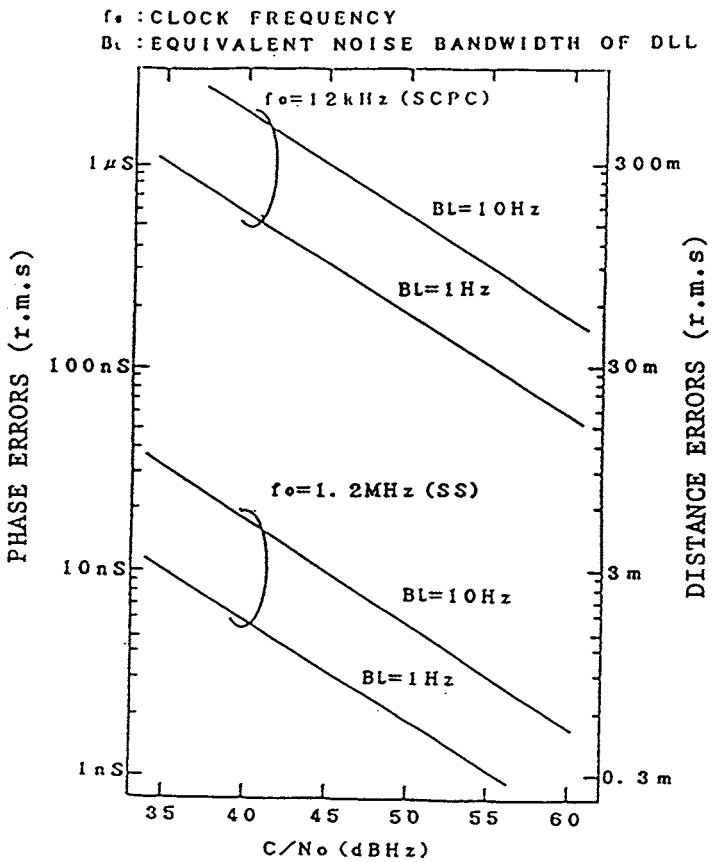


Fig.6 Received C/No versus phase errors of DLL.

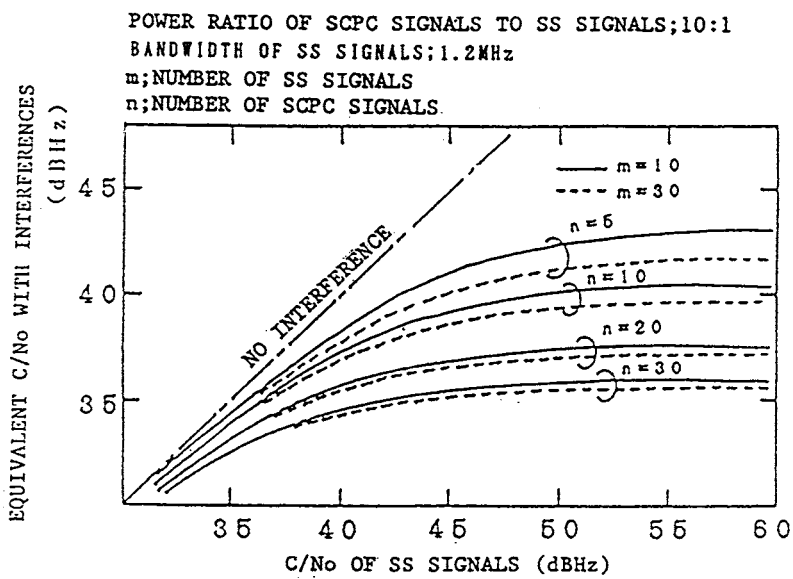


Fig.7 C/No for SS ranging system with interferences.
(Levels of SS and SCPC signals are equivalent.)

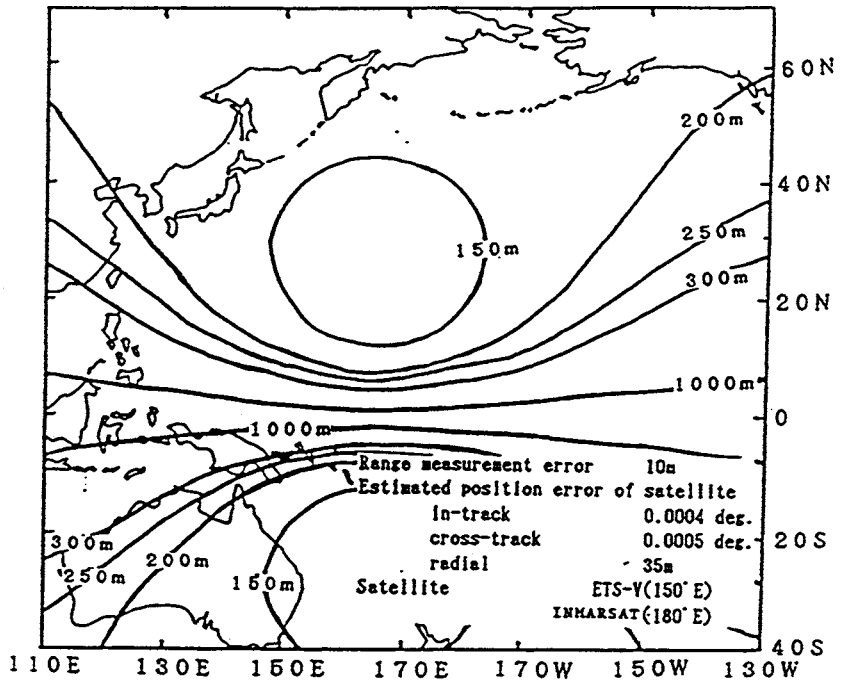


Fig.8 Accuracy of radio determination of this system.

shows the relation between the C/No and an equivalent C/No with interference signals affecting the SS channel.

The link budget of the SCPC system predicts the channel quality of C/No to be about 47-58 dB, which will give a ranging accuracy of several tens of meters as shown in Fig.6. In the SS system, the expected C/No with interference signals is about 35-45 dB, which gives a ranging accuracy of about several meters.

(b) Positioning Accuracy

Positioning accuracy depends on both accuracy of ranging and satellite location information. Fig.8 shows an example of the positioning accuracy using ETS-V and INMARSAT satellites. The theoretical prediction is calculated under the assumption that ranging accuracy is 10m, the ambiguity of satellite location is 0.004 degrees in latitude, 0.0005 degrees in longitude and 35m in radius. In low and middle latitude areas, a positioning accuracy of about several hundred meters is expected as shown in Fig.8. These figures are obtained without using any reference earth stations, but it is expected that greater accuracy can be obtained using reference stations which can eliminate many ranging bias errors.

5. CONCLUSION

A hybrid satellite system, which functions both for communication and radio determination, is proposed and the system configurations of two different methods are described. The proposed system can provide not only ordinary communications such as voice and data, but a radio determination service using a single communication channel. Radio determination and communication experiments using ETS-V and INMARSAT satellites are planned to start from the end of 1989. The experimental results are expected to contribute to establishing new satellite services, which have dual missions of communications and radio determination.

REFERENCES

[1] G.K. O'Naill;"The Geoster Radio Determination Satellite System", Telecommunication Journal, Vol.53, 1987.

[2] Y.Hase et al.;"Experimental Mobile satellite System using ETS-V", IEEE Denshi Tokyo, No25, 1986.

[3] Miura et al.;"Code ranging system using ETS-V and INMARSAT system", Technical Report of IEICE of Japan, SAT88-10,1988.

[4] W.Gill;"A Comparison of Binary Delay-Lock Tracking Implementations", IEEE Trans., Vol.AES-2, No.4, 1986.

Session APP-03 Number 5

HIGH RESOLUTION MICROWAVE TECHNIQUES IN ATMOSPHERIC POLLUTION DETERMINATION

Erwin Schanda
Institute of Applied Physics
University of Bern
Switzerland

Abstract:

The molecular transitions of atmospheric gases cause line radiation in the millimeter wave spectrum. The widths of these lines are dominated by pressure broadening up to about 80 km. The integrated radiation as emitted by a given species at any atmospheric level along the line of sight can be measured against the cold sky background over spectral ranges where the atmosphere is otherwise sufficiently transparent. It contains the information on the height distribution of the density of this species in its line shape. For the inversion procedure of line shape into height profile the accurate high resolution (down to 10^{-7}) measurement of the shape of the line in emission is required. The line amplitude to be measured corresponds only to a few or even fractions of a degree Kelvin. Therefore accurate laboratory determinations of the line broadening parameters and the development of highly sensitive and frequency stabilized spectro-radiometers are prerequisited for the successful measurement of pollutant or other minor constituent gases in the strato- and mesosphere.

On a few experiments of line broadening of SO_2 and NO_2 and several remote determinations of strato- and mesospheric O_3, H_2O and CO as well as synthetic spectra of ClO, all in the frequency range between 90 GHz and 205 GHz, will be reported.

1 THE RADIATIVE TRANSFER OF MOLECULAR LINES IN THE ATMOSPHERE

In connection with the stratospheric ozone depletion problem many activities have been started by research groups all over the world to measure the change of the column density and of the height distribution of ozone as well as of many gaseous species related to the ozone photochemistry. Among the various methods applied, the millimeter wave spectro-radiometry has several advantages: The spectral lines can be measured in emission, i.e. no artificial or natural source is needed; the dominant atmospheric lines (mainly oxygen and water vapour) are sparsely distributed over the millimeter waves and between the lines the atmosphere is relatively well transparent and lines of minor constituents show up clearly over a modest and smooth background; the scattering of millimeter wave at molecules is negligible and scattering at cloud droplets is sufficiently weak in order to permit measurements through cloud cover. Therefore the remote sensing of atmospheric parameters by millimeterwave spectroradiometry is feasible independent of the time of the day and almost independent of weather.

For the computation of radiation received from a specific gas distributed throughout the atmosphere, the equation of radiative transfer of the spectral brightness I ($Wm^{-2}Hz^{-1}sr^{-1}$) in a non-scattering medium

$$\frac{dI(f,s)}{ds} = -\alpha(f,s)\, I(f,s) + j(f,s) \qquad (1)$$

has to be solved. Here f is the frequency, s the coordinate along the path, α the absorption coefficient (m^{-1}), and j the emission coefficient ($Wm^{-3}Hz^{-1}sr^{-1}$). Assuming frequencies below 1 THz and temperatures of the atmosphere higher than 150K it is well-known that the Planck-radiation law may be approximated by the Rayleigh-Jeans law and hence the source function $S = j/\alpha$ is proportional to temperature T.

The gases to be studied are only part in the composition of the atmosphere with a volume mixing ratio V(s) dependent on height. Therefore the absorption coefficient can be written as $\alpha(f,s) = \alpha_0(f,s) \cdot V(s)$ where the normalised absorption coefficient α_0 is

assumed to be dependent on pressure, hence on height, and the frequency dependence at a given height can be expressed by a line shape function, e.g. van Vleck-Weisskopf up to about 50 km where the pressure broadening dominates and a Voigt-profile above that height where the Doppler broadening starts to have a sensible share on the line shape.

With these assumptions the integrated equation of radiative transfer for the brightness temperature received by a radiometer at s_0 can be given under simplifying assumptions by

$$T_r(f,s_0) = T_b \exp[-\int_{\infty}^{s_0} \alpha \, ds] + \int_{\infty}^{s_0} \{V(s)\ \alpha_0(f,s)\ T(s) \exp[-\int_{\infty}^{s} \alpha_0(f,s)\ V(s)\ ds]\}\ ds \quad (2)$$

A single line is assumed well-separated from any strongly frequency-dependent wings of lines by other consitutents. T_b represents a general background radiation which is partly absorbed by the observed molecules.

One can now distinguish between two limiting situations:

In the optically thick case where $\exp(-\int \alpha\ ds) \approx 0$ (e.g. when measuring an abundant species like oxygen O_2) equation (2) can be simplified to

$$T_r(f,s_0) \approx \int_{\infty}^{s_0} T(s)\ V(s)\ \alpha_0(f,s) \exp[-\int_{\infty}^{s} \alpha_0(f,s)\ V(s)\ ds]\ ds \quad (3)$$

If as in the case of oxygen $\alpha_0(f,s)$ and $V(s)$ are well-known, (3) gives a possibility to measure the temperature along the paths, i.e. as a function of height. For this, however, another condition has still to be fulfilled. The weighting function

$$W_T(f,s) = V(s)\ \alpha_0(f,s) \exp[-\int_{\infty}^{s} \alpha_0(f,s)\ V(s)\ ds] \quad (4)$$

must have a peak along s and vanish for extreme values of s. Typically this peak is approximately at s for which $\int \alpha_0(f,s)\ V(s)\ ds \approx 1$. For each frequency along the line spectrum a weighting function (as a function of height) results. Because of the strong dependence of $\alpha_0(f,s)$ on the frequency f, one needs many narrow-band filters around the center frequency of a spectral line in order to measure accurately the shape of the line. At low pressure (e.g. 1 millibar corresponding to about 50 km height), the line is only a few MHz wide while at higher pressures the line is much wider (e.g. at 15 km height more than a hundred times). Each filter position corresponds to a weighting function with a peak at its proper height.

The optically thin case occurs for minor constituent species like ozone for which the volume mixing ratio (VMR) varies between about 1 O_3 per million air molecules (1 ppm) and 10 ppm in the stratosphere. An important molecule in the ozone chemistry ClO is present in the stratosphere at a relative density of about or less than 1 part per billion (ppb). For these cases $\exp[-\int \alpha\ ds] \approx 1$. If the temperature distribution T(s) and (from laboratory measurements) $\alpha_0(f,s)$ are known where s represents pressure as it appears along a path, the integrated equation of radiative transfer

$$T_r(f,s_0) \approx T_b + \int_{\infty}^{s_0} V(s)\ T(s)\ \alpha_0(f,s)\ ds \quad (5)$$

contains the weighting function for measuring the volume mixing ratio V(s) as a function of height

$$W_V(f,s) = T(s)\ \alpha_0(f,s) \quad (6)$$

Now T_b can also include a frequency-independent contribution by the troposphere. At a given frequency, $W_V(f,s)$ will in general have no maximum value along the path. Therefore, the construction of difference weighting functions, i.e. differences between weighting functions of neighbouring frequencies, have been derived [1]. Table 1 and Figure 1 taken from the Ph.D. thesis by Lobsiger [2] illustrate the weighting functions and their relation to the filter positions and bandwidth of a 142 GHz ozone spectro-radiometer. The widths of the weighting functions at half power points are typically between 7 and 10 km.

Table 1. Center frequencies (f) of the 31 filter channels with respect to the center of the resonance line under investigation and bandwidths (b) of the channels showing the increasing frequency resolution toward the center of the line [2]

Channel number	f (MHz)	b (MHz)	W. F. January (km)		W. F. July (km)	
1/31	−/+85.0	30.0	25.5	5.0	25.5	5.5
2/30	−/+60.0	20.0	27.5	5.5	27.5	5.5
3/29	−/+40.0	20.0	30.5	5.5	30.5	5.5
4/28	−/+25.0	10.0	33.0	5.5	33.5	5.5
5/27	−/+17.5	5.0	35.5	5.5	36.0	6.0
6/26	−/+12.5	5.0	37.5	5.5	38.5	6.0
7/25	−/+ 9.0	2.0	39.5	5.5	40.5	6.5
8/24	−/+ 7.0	2.0	41.5	6.0	43.0	6.5
9/23	−/+ 5.0	2.0	43.5	6.0	46.0	7.0
10/22	−/+ 3.5	1.0	46.0	6.0	49.0	7.0
11/21	−/+ 2.5	1.0	49.0	6.5	52.0	7.5
12/20	−/+ 1.6	0.5	53.5	7.0	57.5	7.5
13/19	−/+ 0.8	0.2	59.5	7.5	63.5	7.5
14/18	−/+ 0.4	0.2	66.0	8.0	70.5	7.5
15/17	−/+ 0.2	0.2	73.5	9.5	77.5	8.5
16	0.0	0.2				

Altitudes of maximum and half width at half power for the respective weighting functions are given for a 47°N CIRA (1972) January and July atmosphere.

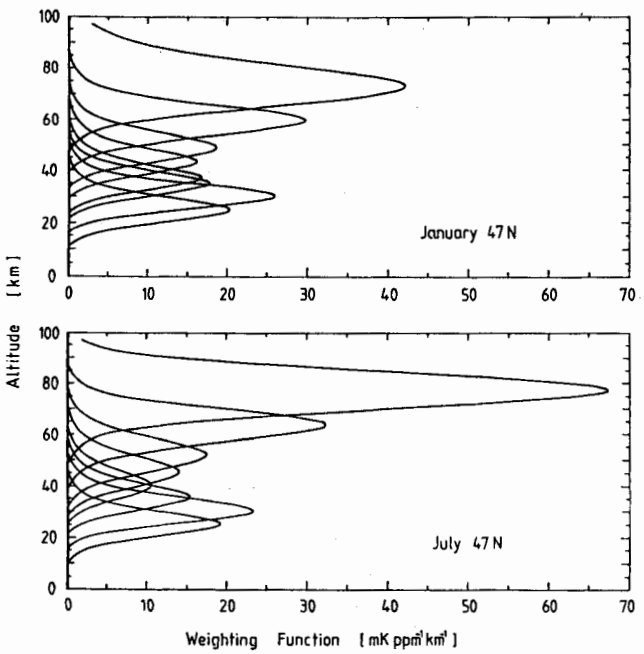


Fig. 1: Weighting functions for the change of brightness temperature differences with respect to unit (1 ppm VMR) ozone changes are displayed. The functions were calculated by integration of the radiative transfer equation (RTE) using the CIRA (1972) January (upper half) and July (lower half) reference atmospheres. For clarity, from the 15 brightness temperature differences used in our inversion procedure (see Table 1) only every second respective function is shown. [2]

The resulting line shape (in brightness temperature) contains the contributions from all atmospheric levels and can be determined by the utilisation of a large number of filters with bandwidths matched to the steepness of the line shape as indicated in Table 1. In order to determine the volume mixing ratio, Equation (5) has to be solved for V(s) which means solving an integral equation. Various methods for this inversion procedure have been developed which will not be discussed here. However, it is evident that the accurate and high-resolution determination of the line shape is needed as input to the inversion in order to achieve reliable volume mixing ratios of minor constituents as a function of height.

2 LINE-BROADENING DETERMINED IN A BRIDGE SPECTROMETER AT 94 GHZ

The importance of the line shape as a function of pressure for the remote determination of the distribution of minor constituents in the middle atmosphere motivated our group to develop a bridge spectrometer [3]. Figure 2 shows a special version as it was used for the measurement of molecular interaction in the system SO_2-H_2O [4]. The pair of cells is 8 meters long and the second focus of each elliptic mirror is at the center of the respective cell. All other components are explained in figure caption. In the following a short discussion of two investigations by D. Meier taken from his Ph.D. thesis will be presented.

In the frequency range between 91.5 GHz and 95.5 GHz, three rotational lines of the $^{32}S^{16}O_2$ and two rotational lines of the $^{34}S^{16}O_2$ molecules in the fundamental vibrational state and also two lines of the $^{32}S^{16}O_2$ molecule in the ν_2 vibrational state have been investigated. In particular the self-broadening and the broadening by H_2O and by N_2 of the transition 23(6,18)-24(5,19) with the line center at 94.064 GHz have been studied and the following linewidth parameters were found: SO_2-SO_2 18.2 ± 0.3 MHz/torr, SO_2-N_2 3.8 ± 0.1 MHz/torr and SO_2-H_2O 15.2 ± 0.2 MHz/torr. The measured absorption coefficients in the pressure range 0.139 torr to 0.464 torr were fitted by a least squares method to the Lorentzian line shape

$$\alpha = C\,N\,\frac{\nu^2\,\Delta\nu}{(\nu-\nu_0)^2 + \Delta\nu^2} \qquad (7)$$

with a coefficient C and the number N of absorber molecules per cm^3. Figure 3 shows the linewidth $\Delta\nu$ and the coefficient C as a function of pressure as they were used to fit (7) with the measured data points. These values yielded also good agreement between measured and theoretical maximum absorption coefficients according to [6].

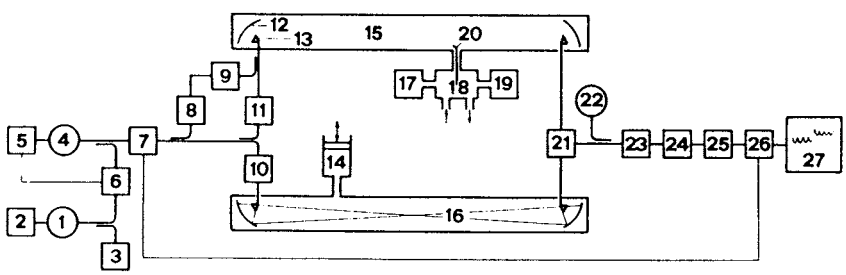


Fig. 2: Block diagram of the mm-wave bridge spectrometer. (1) 8-12 GHz klystron, (2) Frequency stabilisation 8-12 GHz, (3) Frequency counter, (4) 91.5-95.5 GHz klystron, (5) Frequency stabilisation 91.5-95.5 GHz, (6) Harmonic mixer, (7) Modulator, (8) Phase shifter, (9) Attenuator, (10) Phase shifter, (11) Attenuator, (12) Elliptical reflector, (13) Horn antenna, (14) Motor driven piston, (15) Absorption cell, (16) Reference cell, (17) Water vaporizer, (18) Pressure chamber, (19) Pressure Gauge, (2) Sonde, (21) Hybrid T, (22) Local oscillator BWO 75-110 GHz, (23) Mixer IF 1.5 GHz, (24) Detection diode, (26) Lock-in amplifier, (27) X-Y plotter [4]

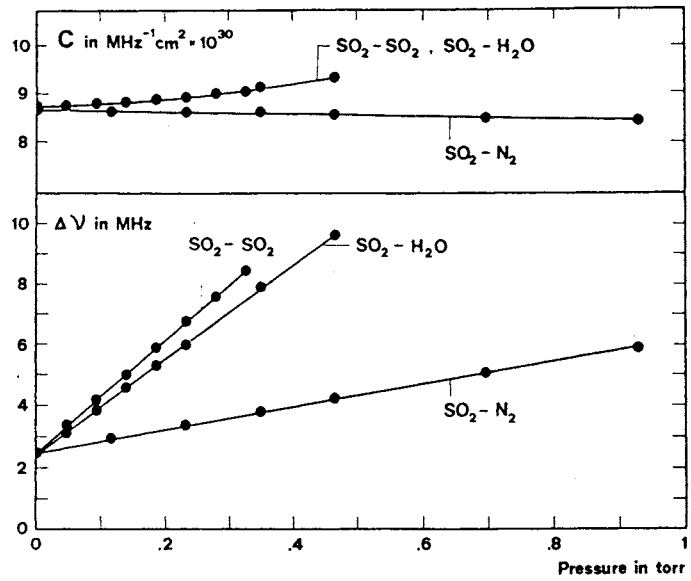


Fig.3: Line broadening ($\Delta\nu$) and the coefficient C as functions of pressure for SO_2-N_2, SO_2-H_2O and SO_2-SO_2 collisions. T = 295°K. Points derived from the Lorentzian fits. [5]

The second example to be presented here is the pressure broadening of the NO_2 hyperfine multiplet around 93 446 MHz of the 26(1,25)-25(2,24) rotational transition [7]. The multiplet consists of six lines between 93 430.37 MHz and 93 458.55 MHz with $\Delta F = \Delta J = \Delta N = 1$. The same line-broadening parameter of 3.7 MHz/torr and the same maximum absorption of 1.8 10^{-5} cm^{-1} have been found for all 6 lines. At 10 torr the maximum absorption of the strongly overlapped multiplet was measured to be 1.06 10^{-4} cm^{-1} in good agreement with the theoretical value. The shape factor consists of a sum of Lorentzian shapes plus an interference term for each line. The measured points were fitted according to

$$\alpha = \sum_{i=1}^{6} a \frac{\Delta\nu + (\nu - \nu_i)\, g_i}{(\nu - \nu_i)^2 + \Delta\nu^2} \qquad (8)$$

where a is an amplitude factor, g_i are the overlap parameters, ν_i the respective line frequencies.

Figure 4 shows the line broadening and the variation of the amplitude factor as well as the overlap parameters as a function of the pressure. The interference effect turns out to be less than 10 % at all pressures used in this investigation.

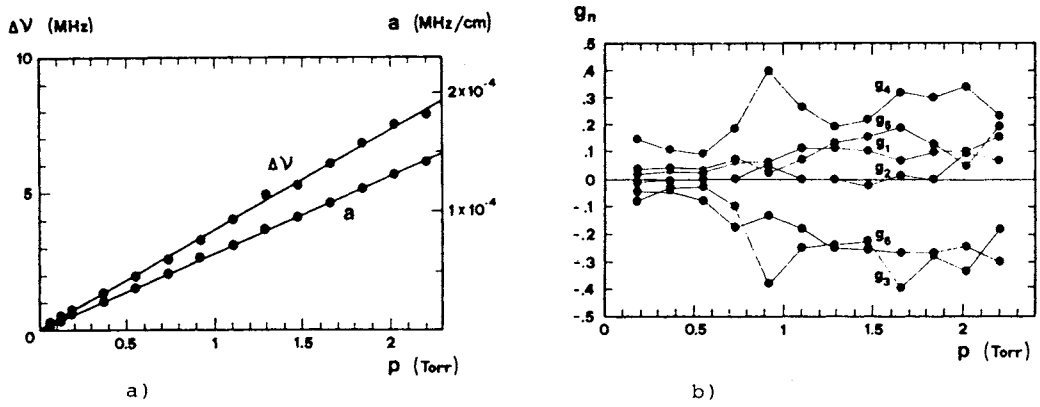


Fig. 4 The linewidth $\Delta\nu$ and the coefficient a in eqn (2) as functions of the gas pressure. [7]

The line shapes, line widths and overlap of lines in the case of multiplets have to be studied for each individual rotational transition which is to be used for the remote sensing of an atmospheric constituent. For the determination of the center frequencies, line strengths, partition functions, there are comprehensive compilations of spectroscopic data available, e.g. [8].

3 REMOTE SENSING OF MOLECULAR ABUNDANCES

There are a few research groups who developed spectro-radiometers for the remote detection of various important species of the stratosphere to be used from ground-based, balloon-borne, airborne and space-borne platforms, e.g. [9, 10, 11] and they measured very successfully e.g. water vapour, ozone, chlorine monoxyde and other gases.

Figure 5a is a simplified diagram of the ground-based cooled ozone sensor [2] which has been developed at our institute in Bern for long-term observation of the height profile of the stratospheric and mesospheric (15 to 75 km height) ozone. Via a rotating mirror the signal (spectral line at 142 GHz) enters a cryogenically cooled (20 K) mixer-preamplifier stage where at 3.7 GHz the signal is either fed directly to broadband filters (centered along the far wings of the line up to ± 550 MHz off line center) or to a second and third mixer stage where finally at about 115 MHz narrow band filters are arranged (approximately as given in Table 1) close to the line center. The entrance mirror can be turned alternately to a hot (≈ 320 K) and a cold (≈ 80 K) load for periodic calibration. Figure 5b indicates the frequency stabilisation needed to keep the local oscillator (at 73 GHz) stable to less than 10 KHz over many days. The front-end mixer operates according to the subharmonically-pumped principle, i.e. the local oscillator needs to deliver only half of the nominal l.o. frequency.

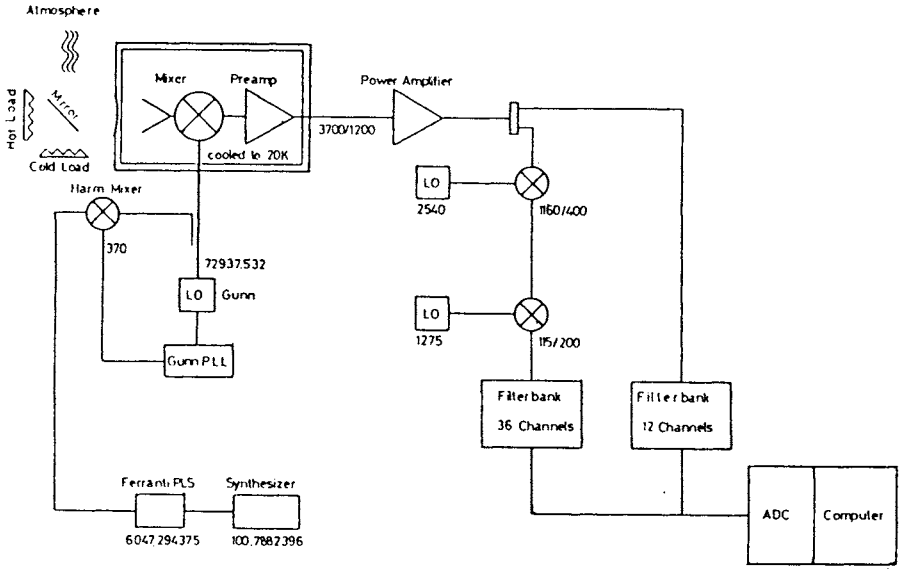


Fig. 5 a):
Block diagram of the 142 GHz ozone radiometer (all frequencies in MHz)

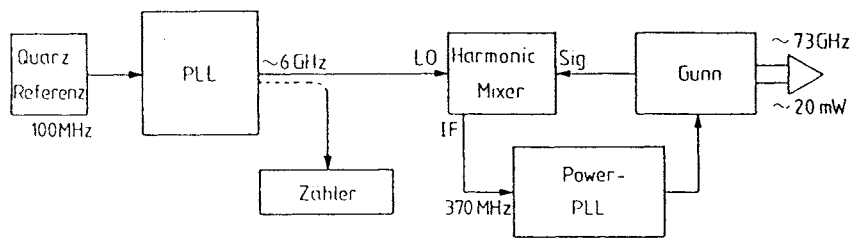


Fig. 5 b):
Block diagram of the frequency stabilizing network [12].

Figure 6 shows the output of the filter channels over the inner part a) of the spectal line and the wing part b), expressed in brightness temperature [13]. The radiation enters under a zenith angle of approximately 35° and contributes a line amplitude of about 10 K over a continuum radiation by the atmosphere and the sky of about 44 K. This kind of spectral lines are the input to the inversion procedure from which the height profiles of ozone (or other minor constituents) can be determined. In this way it was possible to determine the day-and-night differences of the mesospheric ozone distribution. Figure 7 shows the result of this kind of observations [14]. The observed (full lines) distribution demonstrates the required sensitivity of the measuring system. At 70 km the change between noon-values (about 2 O_3-molecules in 10^7 atmospheric molecules) and midnight (approx. 10 to 10^7) is almost a factor 5; the error bars in VMR are approximately 10^{-7} (0.1 ppm).

Another species of high environmental significance is carbon monoxyde. At mesospheric levels (>50 km) CO is produced by photo dissociation of CO_2 due to the intense UV-light up there. Burri [16] measured an increase of CO with height from ≈ 1 ppm at 55 km to ≈ 10 ppm at 80 km observing the line shape of the 115 GHz rotational line with a ground-based spectro-radiometer.

Water vapour is an important trace molecule in the stratosphere. R. Peter has measured H_2O during his diploma thesis work with an air-borne instrument in December 1986, receiving the 183.3 GHz-line [17]. The height profiles along the flight path between 45° and 75° northern latitudes shows a maximum of approximately 5 to 5.5 ppm at 40 to 60 km height over mid latitudes shifting continuously down to 25 to 35 km at latitudes closer to the pole.

Space-borne spectro-radiometers for use on board the UARS-satellite [11] and on board the ATLAS-missions of the Space Shuttle [18] have been developed for flights in 1991 and later. This allows for a different observing geometry. In the so-called limb-sounding mode the radiation from different heights can be received by viewing tangentially through the atmospheric layers and changing elevations of a narrow antenna beam. The much larger path length of radiation from a given tangential height and the much lower

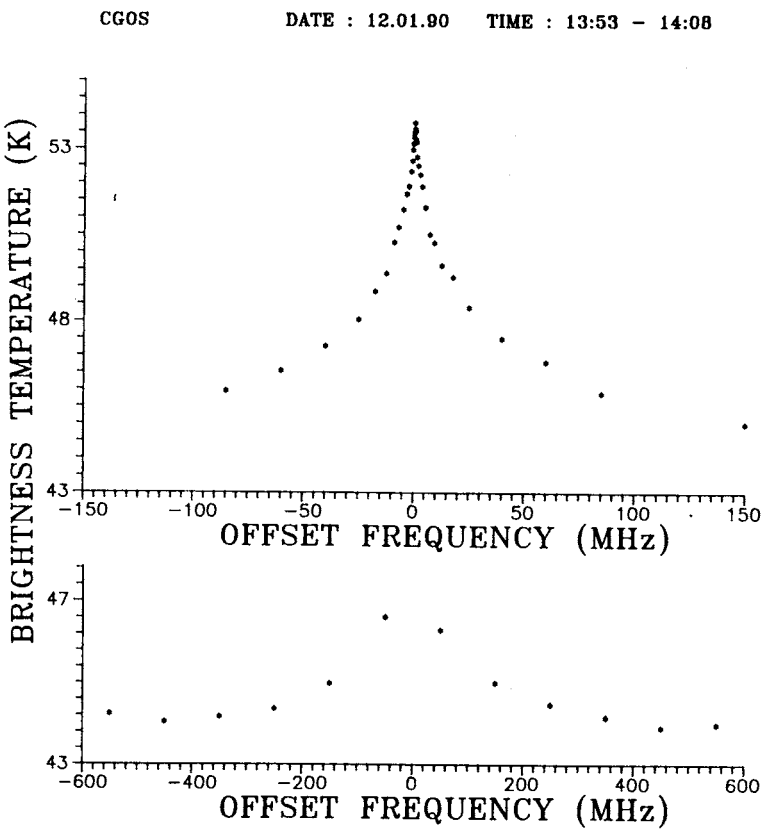


Fig. 6:

Brightness temperature of the 142 GHz ozon line,

a) ± 150 MHz about line center,

b) ± 600 MHz about line center.

Each dot represents the output of one filter [13]

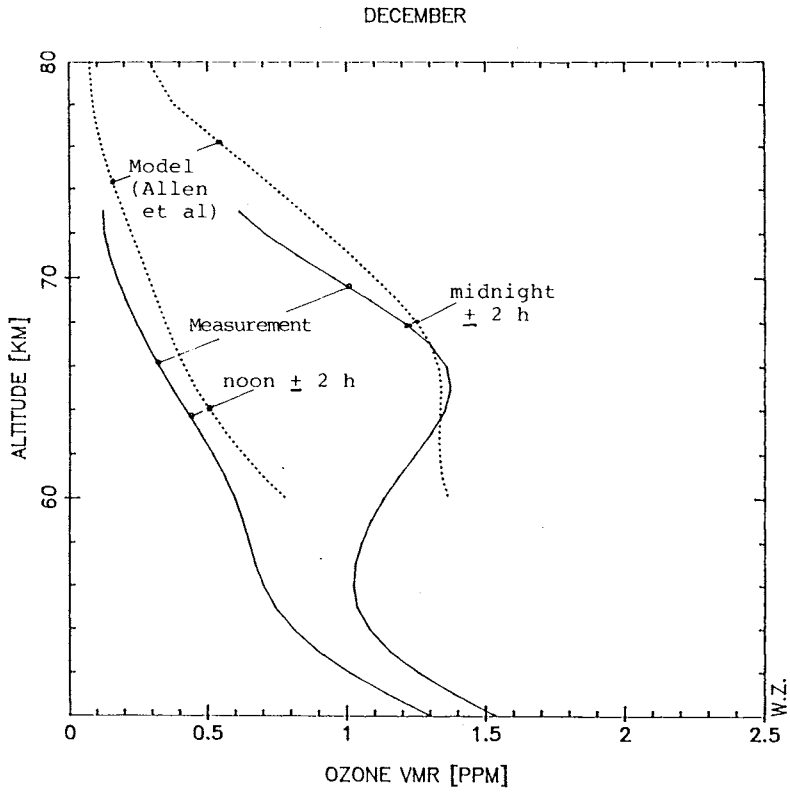


Fig. 7:
Mesospheric height profile of ozone at noon and midnight for mid-latitude (Bern, 47°N) [14]

continuum radiation due to avoiding the deeper atmospheric layers yield much higher sensitivity. Therefore one can hope to measure the global height distribution of species with abundancies several orders of magnitude less than e.g. ozone.

Chlorine monoxyde is estimated by ground based millimeter wave measurements [9] to appear in the stratosphere at volume mixing ratios of less than 10^{-9}, more than 3 orders of magnitude less than O_3.

Figures 8a and 8b are synthetic spectra of the ClO 204 GHz lines due to the J = 5 to J = 6 transition [19], computed for an expected height profile. Only the four most prominent lines of the multiplet of $^{35}Cl^{16}O$ appear on this frequency range of 40 MHz wide. The geometry assumed was tangential sounding by a radiometer placed at the respective tangent points at 20 km and 50 km. As background only the cosmic radiation (2.7 K) was taken into account. For a limb sounding from a spacecraft a twice as large line amplitude may be expected on top of the same background. The pressure broadening at 20 km causes an almost complete overlap of the lines while at 50 km each line appears separately, however, with an intensity of only a few hundredths of a degree. Not only the instrument sensitivity presents a problem but also the inversion of the measured brightness temperatures of this multiplet into a height profile is a delicate task.

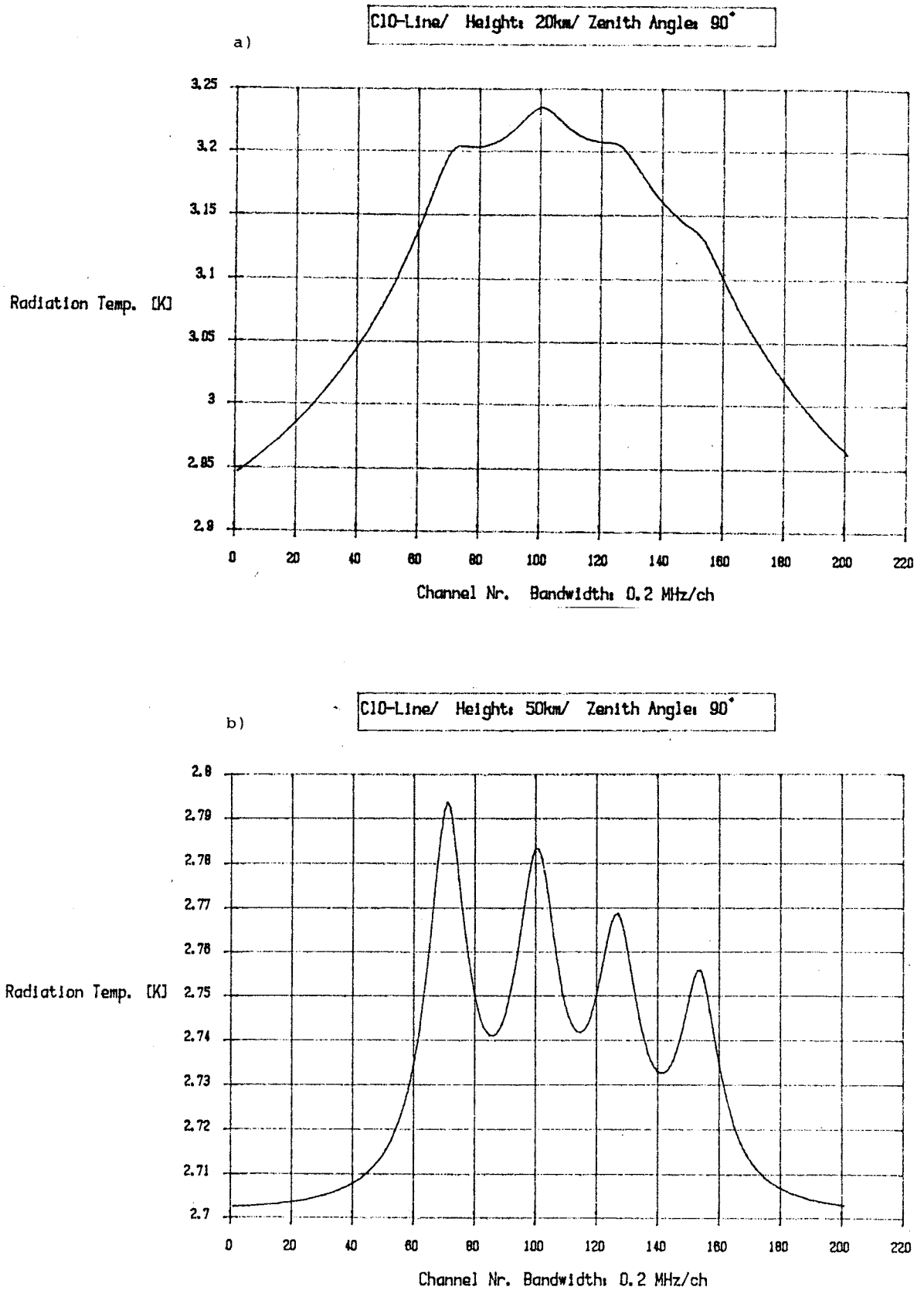


Fig. 8 : Synthetic spectra of the ClO 204 GHz multiplet for tangential observing geometry [19], a) at 20 km altitude, b) at 50 km altitude

4 REFERENCES

[1] A.K. Randegger (1980) Pure appl. Geophys, 118, pp. 1052-1065.

[2] E. Lobsiger (1987), J. Atmospheric and Terrestrial Physics. 49, No. 5, pp. 493-501.

[3] G. Schaerer, D. Meier, E. Schanda (1975), Journal of Physics E: Scientific Instruments, 8, pp. 561-562.

[4] D. Meier (1979), Adv. Molec. Relax. and Interact. Processes, 14, pp. 91-98.

[5] D. Meier (1978), J. Quant. Spectrosc. Radiat. Transfer, 19, pp. 323-329.

[6] C.H. Townes, A.L. Schawlow (1955), Microwave Spectroscopy, McGraw Hill, New York.

[7] D. Meier (1978), J. Quant. Spectrosc. Radiat. Transfer, 20, pp. 129-133.

[8] R.L. Poynter, H.M. Pickett (1984), Submillimeter, Millimeter and Microwave Spectral Line Catalogue. JPL Public. 80-23, Revision 2, Jet Propulsion Laboratory, Pasadena, Calif.

[9] A. Parrish, R.L. de Zafra, P.M. Solomon, J.W. Barrett (1988), Radio Science, 23, No. 2, pp. 106-118.

[10] R.M. Bevilacqua, J.J. Olivero, C.L. Croskey (1989), J. Geophys. Res., 94, No. D10, pp. 12807-12818.

[11] J.W. Waters (1989), Atmosph. Res., 23, pp. 391-410.

[12] G. Rubin (1987), Diploma Thesis, Institute of Applied Physics, Bern.

[13] P. Bodenmann (1990), personal communication, Institute of Applied Physics, Bern

[14] W.C. Zommerfelds (1988), Ph.D. Thesis, University of Bern, Institute of Applied Physics, Bern.

[15] W.C. Zommerfelds, K.F. Kunzi, M.E. Summers, R.M. Bevilacqua, D.F. Strobel, M. Allen, W.J. Sawchuck (1989), J. Geophys, Res., 94, No. D10, pp. 12819-12832.

[16] J. Burri (1987), Diploma Thesis, Institute of Applied Physics, Bern.

[17] R. Peter, K. Künzi (1988), Geophys. Res. Lett., 15, No. 11, pp. 1173-1176.

[18] E. Schanda, K. Künzi, N. Kämpfer, G. Hartmann, W. Degenhart, E. Keppler, A. Loidl, G. Umlauft, V. Vasyliunas, R. Zwick, P.R. Schwartz, R.M. Bevilacqua (1986), Acta Astronautica, 13, No. 9, pp. 553-563.

[19] L. Gerber, C. Aellig (1990), personal communication, Institute of Applied Physics, Bern.

Resonators, Oscillators and Materials

ROM-03 Resonators, Oscillators and Measurements

Chairman: J. Dowsett

Session ROM-03 Number 1

PI-Network Measurement of the Load Resonance Frequency without Load Capacitor

Dipl.-Phys. Bernd Neubig
TELE QUARZ GmbH, Landstrasse, D-6924 Neckarbischofsheim 2
Fed.Rep. of Germany

Introduction

The measurement of quartz crystal units with load capacitance is probably the subjects that still causes most of the troubles in measurement correlation between manufacturers and users.
While the PI-Network Method according to IEC 444 is generally accepted, the measurement of load resonance parameters remains to be unsatisfactory, and therefore the relevant IEC-Report 444-4 did not find general acknowledgement. Previously published methods which calculate fL from the measured motional capacitance, show high errors due to their high sensitivity to the error of the measured C_1 and C_0 values.
The new procedure described below is based on the PI-network method of IEC 444. Instead of using physical load capacitors, this software-assisted approach determines the load resonance frequency f_L and load resonance resistance R_L from a measurement at two frequencies - one at the resonance frquency f_r and one off resonance. Once per measurement batch a calibration with open and short circuit has to be performed. (11)
The applicable frequency range is up to 125 MHz, i.e. the same as that of IEC 444. Basically this method can also be adopted by other measurement techniques such as the s-parameter method of EIA-512.

1. Principle of measurement

1.1 Fundamentals

If a crystal unit with a series resonance frequency f_s and the parameters R_1, C_1 resp.L_1 and C_0 is connected with a load capacitor C_L in series to it, the resulting two-pole network behaves in the same way as a crystal unit with a higher frequency, the so-called load resonance frequency f_L

$$f_L = f_s \sqrt{1 + \frac{C_1}{C_0 + C_L}} \approx f_s \left(1 + \frac{C_1}{2(C_0 + C_L)}\right) \qquad [1]$$

a higher resonance resistance R_L

$$R_L = R_1 \left(1 + \frac{C_0}{C_L}\right)^2 \qquad [2]$$

as well as modified motional and static parameters L_1', C_1', C_0'

$$L_1' = L_1 \left(1 + \frac{C_0}{C_L}\right)^2$$

$$C_1' = \frac{C_1}{\left(1 + \frac{C_0}{C_L}\right)^2 \cdot \left(1 + \frac{C_1}{C_0 + C_L}\right)} \approx \frac{C_1}{\left(1 + \frac{C_0}{C_L}\right)^2} \qquad [3],[4],[5]$$

$$C_0' = \frac{C_0 \cdot C_L}{C_0 + C_L}$$

Therefore[1], at the load resonance frequency f_L the crystal unit shows an impedance

$$Z_L = R_L + j * X_L \qquad [6a]$$

with

$$X_L = 1/(w_L * C_L) \qquad [6b]$$

whose magnitude and phase are:

$$|Z_L| = \sqrt{(R_L^2 + X_L^2)} \qquad [7]$$

$$arc(Z_L) = arc\ tan(X_L/R_L) \qquad [8]$$

If the crystal is inserted in a PI-network according to IEC 444, the effective impedance Z_{eff} consists of Z_L plus the PI-network source and load impedance "seen" by the crystal, which is nominally R_T = 25 Ω :

$$Z_{eff} = Z_L + R_T = R_L + R_T + j * X_L \qquad [9]$$

Magnitude and phase of Z_{eff} are

$$|Z_{eff}| = \sqrt{(R_L+R_T)^2 + X_L^2)} \qquad [10]$$

$$arc(Z_L) = arc\ tan(X_L/(R_L+R_T)) \qquad [11]$$

The output voltage U_b of the PI-network and the impedance Z_x of the inserted crystal are generally related by the equation

$$Z_x = R_T * (U_{bk}/U_b - 1) \qquad [12]$$

where U_{bk} is the PI-network output voltage with a short circuit instead of the crystal. The complex voltage ratio U_{bk}/U_b can be broken down into magnitude and phase ϕ_π as measured with the network analyzer or the vector voltmeter :

$$|U_{bk}/U_b| = \sqrt{((R_L+R_T)^2 + X_L^2)} / R_T \qquad [13]$$

$$\phi_\pi = - arc\ tan(X_L/(R_L+R_T)) \qquad [14]$$

1.2 Evaluation of the new method

In the following text we have considered typical quartz crystal units in the fundamental mode at 1 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz, 20 MHz, 30 MHz, and 40 MHz as well as third overtone crystals at 30 MHz, 50 MHz, 70 MHz, and 90 MHz. The assumed motional and static parameters are given in Table 1. They are typical values for AT-cut crystals in the standard enclosure families HC-51/U, HC-49/U, and HC-52/U.

fL [MHz]	ovt	Rr [Ohm]	C1 [fF]	C0 [pF]
1	1	100	10	3.5
3	1	50	8	4
5	1	30	12	4.5
10	1	15	25	5.5
15	1	10	25	6
20	1	8	25	6
25	1	5	25	6
30	1	5	25	6
40	1	5	20	5.5
30	3	30	1.8	4
50	3	25	2.2	5
70	3	20	2.2	5
90	3	20	2	5

Table 1: Parameters of evaluated quartz crystal units

1.2.1 PI-network output amplitude and phase

Figure 1a shows the relative output amplitude $|U_b/U_{bk}|$ of the PI-network as a function of the load capacitance C_L for typical fundamental mode quartz crystal units from 1 MHz to 40 MHz. Because of the high crystal impedance at f_L - especially for small load capacitances and low frequencies - the amplitude drops to a rather low value for frequencies below 5 MHz. For a 1 MHz crystal the output amplitude for C_L = 10 pF...100 pF is 36 dB...56 dB below the short-circuit voltage U_{bk}. This critical range is depicted in magnified form in Fig.1b. Above 10 MHz the attenuation of U_b with respect to U_{bk} is always less then 20 dB.
Fig.1c displays the amplitude ratio for the 3rd overtone crystals 30 MHz ... 90 MHz. In this range, the amplitude is always sufficiently high.
Figure 2 displays the PI-network phase shift (referred to short-circuit) at f_L for fundamental mode and 3rd overtone crystals respectively. For the crystals below 20 MHz the phase is mainly in the range between -80°...-90° for C_L = 5 pF...50 pF. For the overtone crystals the phase varies within a range of -40°...-80°.

1.2.2. Crystal drive level (power and current)

At the load resonance frequency the crystal possesses an impedance, which is essentially a reactance with a small resistive part. Therefore the voltage drop across the crystal is mainly reactive, and consists only of a small dissipative portion. Thus a higher input level at the PI-network is necessary than for the measurement of f_r to achieve the same crystal drive level (i.e. dissipated power).
From the short-circuit output voltage U_{bk} of the PI-network the crystal drive level P_q and the crystal current I_q can be computed as follows:

$$P_q = 4*U_{bk}^2*R_L/(k_\pi^2*((R_L+R_T)^2 + X_L^2)) \qquad [15]$$

$$I_q = 2*U_{bk}/(k_\pi * \sqrt{((R_L+R_T)^2 + X_L^2)}) \qquad [16]$$

where

k_π = 0.3649166
$X_L = 1/(w_L * C_L)$
R_T = 25 Ω

For a specified drive level P_q resp. I_q the voltage U_{bk} has to be adjusted to

$$U_{bk} = k_\pi/2 * \sqrt{(P_q * ((R_L+R_T)^2 + X_L^2)/R_L)} \quad [17]$$

respectively

$$U_{bk} = k_\pi/2 * I_q * \sqrt{((R_L+R_T)^2 + X_L^2))} \quad [18]$$

Figs. 3a and 3b show the crystal drive level for a PI-network input power level of +30 dBm computed from equation [15]. A crystal drive level of 100 µW can be achieved for crystals above 10 MHz with a load capacitance of 25 pF. At lower frequencies and smaller C_L-values the maximum drive level drops down to some 10 µW. As can be seen in Fig.3b, the 3rd overtone crystals can be easily driven to 100 µW and above.
Applying equation [16], Fig.4 depicts the crystal current for +30 dBm PI-network input power level. A crystal current of 1 mA can be realized for crystals above 5 MHz with a load capacitance of 25 pF and more. Again, no problem exists for overtone crystals to be operated above 1 mA.

1.2.3 Measurement sequence

If the resonance resistance R_r and the static capacitance C_0 are known, the crystal impedance Z_L at the load resonance frequency f_L can be calculated from equation [6]. With the equations [13] and [14] the appropriate voltage amplitude and phase readings on the network analyzer can be computed.
Thus the load resonance can be determined by tuning the generator to that frequency above the resonance frequency f_r, at which the PI-network output voltage has the required values of amplitude and phase. As will be shown in paragraph 2, the best accuracy is achieved, if the search is performed for the voltage condition only and the phase condition is neglected (or used only for a plausibility check).

For the sequence of measurements two different generator levels are used:

- For the measurement of f_r and R_r : generator output level adjusted such, that the specified crystal drive level at phase zero is realized. Subsequently this generator level will be called "LO"-level. For a crystal drive level of P_q(mW) resp. a crystal current I_q (mA), the PI-network output voltage U_{bkLO} with short circuit inserted is (for a 25 Ω crystal):

$$U_{bkLO} = 57.8 \text{ mV} * \sqrt{(P_q/1\text{mW})} \quad [19]$$

respectively

$$U_{bkLO} = 9.138 \text{ mV} * I_q/1\text{mA} \quad [20]$$

With the conventional IEC 444 set-up the generator level is about 35.6 dB higher than U_{bk}.

- For all other measurement : the generator level is set to the "HI"-level, at which the nominal drive level is reached at the load resonance frequency. The applicable formulas [17] and [18] are given in paragraph 1.2.2. The PI-network output voltage with short circuit is denoted as U_{bkHI}.

Based on the measurement set-up according to IEC 444-1, the complete measurement sequence is as follows:

Step 1: calibration
a) PI network without crystal ("open circuit"): The generator output is set to the "HI"-level and is tuned to f_{C0}, the frequency at which C_0 will be measured (see step 2). The cross-talk capacitance C_c is computed from the PI-network output voltage U_{open}:

$$C_c = 1/(w_{C0}*R_T*\sqrt{(U_{bkHI}/U_{open}-1)}) \quad [21]$$

b) PI-network with short circuit: The generator is tuned to the nominal frequency f_{nom} of the load resonance. The phase reading is set to zero (reference value). To compensate possible amplitude errors of the vector voltmeter, this calibration is performed at the "HI"-level (PI-network output voltage U_{bkHI}) and at the "LO"-level (voltage U_{bkLO}). While U_{bkHI} is used as reference for the f_L measurement, U_{bkLO} is the reference for the zero phase measurement of f_r and R_r.

Step 2: Measurement of C0
PI-network with crystal: The generator is adjusted to the "HI"-level and tuned to a frequency f_{C0} approximately 10% below f_{nom}. For nominal frequencies below 5 MHz, f_{C0} should be selected between 5 and 15 MHz to avoid noise problems at too low output voltages. Alternatively the recent suggestion of IEC TC49/WG 6 may be followed[2].
The static capacitance C_0 is determined from the output voltage reading U_{C0}. From the result the cross-talk capacitance has to be substracted

$$C_0 = 1/(w_{C0}*R_T*\sqrt{(U_{bkHI}/U_{C0}-1)}) - C_c \quad [22]$$

Step 3: Measurement of fr and Rr
PI-network with crystal: The generator is set to the "LO"- level and tuned to that frequency, where zero phase (relative to the short circuit calibration with U_{bkLO}) occurs. To speed up the search for f_r, an estimate of f_r can be computed with the help of equation [1] using an estimated value for C_1 and the measured value of C_0.
The resonance resistance R_r can be calculated (as in IEC 444-1) from U_{bkLO} and the voltage U_b at zero phase according to equation [12].

Step 4: Measurement of fL, determination of RL
PI-network with crystal: From the measured R_r the load resonance resistance R_L is determined by equation [2] using the measured C_0 value.
The voltage amplitude at f_L has to be computed from equation [13] using U_{bkHI}. The load resonance frequency f_L is then found by setting the generator to the "HI"-level and tuning to that frequency above f_r, at which the amplitude condition is fulfilled. A first guess f_{L1} is the nominal value f_{nom}. With the amplitude reading at a second frequency a few ppms from f_{nom}, a good second guess f_{L2} can be computed by interpolation. Because in the vicinity of f_L the amplitude vs. frequency response is rather linear, the final value of f_L can be reached by very fast iteration. As this tuning is off resonance, the frequency stepping can be realized very fast without degradation by ringing effects of the crystal.

2. Error considerations

A detailed error analysis has been performed to consider the influence of different measurement errors. A comparison with the conventional method using a physical load capacitor is made in paragraph 2.5.

2.1 Error of the amplitude measurement

Assuming an error of 0.5% for the amplitude measurement of the network analyzer or the vector voltmeter respectively leads to an erraneous location of f_L. The hereby caused frequency error is depicted in Figs. 5a and 5b. For fundamental mode crystals the error is worst at small load capacitances for the crystals at and above 10 MHz. The maximum error is less then 3 ppm, the average is at about 1.5 ppm. Overtone crystals show a much smaller error sensitivity of less then 0.3 ppm.

2.2 Error of the phase measurement

Because of the fact that in the vicinity of f_L the phase angle is close to constantly 90°, the frequency sensitivity to phase errors is very high as shown in Fig.6a and 6b. Therefore it is strongly recommended to use only the amplitude criterion of equation [13] and not the phase criterion of [14] for the search of f_L. The phase information should only be used as a rough indicator to verify that the measured data are regular.
The following charts are all under the assumption, that the amplitude-search method is used.

2.3 Error of R_r and DLD

If the determination of R_r at zero phase is erraneous, this leads to a frequency error for f_L. For an assumed R_r-error of 1% the results are shown in Fig.7a and 7b for the fundamental and overtone mode crystals resp. In both cases the error increases with frequency and with C_L. It is in the order of a tenth of a ppm and can therefore be neglected compared with the other influences. The same graphs can also be used for a judgement about the sensivity of this method to drive level dependencies (DLD) inherent in real crystal resonators. A 10 MHz/ 30 pF crystal unit with a drive level dependent resonance resistance variing 10% has an additional frequency error due to this DLD effect of 0.1 ppm, which certainly can be neglected.

2.4 Error of C_0

Because C_0 is needed to compute R_L, which itself is a term in the amplitude equation [13], the measurement error of C_0 influences the frequency accuracy of f_L.
Fig. 8 shows the computed frequency error for a C_0 error of 2%. For fundamental as well as 3rd overtone crystals this error is only in the order of 10^{-8} and is therefore negligible.

2.5 Comparison with the classical method

The classical method of measuring f_L uses a physical, i.e. real, load capacitor as an adaptor switched in between the measuring jig and the crystal unit under test. This is the method described in IEC report 444-4, which did not find general acknowledgement due to its insufficiencies. The achievable tolerances claimed in this document are by far not reached in practice.
More realistic figures about the possible accuracy in realizing load capacitance adaptors at different places are as follows[3]:

C_L			
up	to 9.9 pF	: tolerance	±0.15 pF
10 ...	19.9 pF	:	±0.20 pF
20 ...	39.9 pF	:	±0.25 pF
40 ...	99.9 pF	:	±0.30 pF
above	100 pF	:	±0.40 pF

Dividing the sample crystal types of table 1 into only three groups

group 1 (planoparallel) : C1 = 25 fF; C0 = 6 pF
group 2 (convex, contoured) : C1 = 10 fF; C0 = 5 pF
group 3 (third overtone) : C1 = 2.2fF; C0 = 6 pF

the computed measurement error induced by the above mentioned tolerances of the load capacitors is depicted in Fig.9.

Furthermore an important frequency dependent error is inherent in every load capacitor adaptor, which is caused by the inevitable inductance of the connecting leads of the adaptor. Assuming this inductance to be totally 10 nH this frequency dependent error is in the order of several ppm as depicted in Fig. 10, which has to be added to the error numbers of Fig.9.
It can be seen that the frequency error of the classical method is

at least three times higher than the most dominant error of the new method (Fig.5a).
An important advantage of the new method is, that it is completely software-based, and therefore independant of the mechanical construction and calibration accuracy of any C_L-adaptor. So basically the same measurement results can be achieved at different places with different equipments, which are only limited by the accuracy of the instrumentation.

2.6 Additional aspects

2.6.1 Measurement of load resonance frequency vs. temperature

Besides the measurement at room temperature, an important factor in practice is the measurement of the load resonance frequency in the temperature range. The classical approach with load capacitor adaptors yields several problems in the practical application, as it is necessary to place the capacitors outside the temperature chamber, which causes considerable problems with lead length etc. Measuring crystals with different load capacitance at the same temperature run is also difficult to handle etc.
The new method allows to measure the f_L vs. temperature characteristics easily without hardware problems. Because the crystal is measured directly at the same operating point, where it works with the load capacitance in the final circuit, all changes of this point in the temperature range can be tracked, the regular $f_L(\theta)$ response as well as irregularities due to activity dips, spurious responses etc.
Existing software methods[4] which compute the load resonance frequency from the measurement of f_r, C_1 and C_0 do not yield the exact $f_L(\theta)$ response (which is different to the $f_r(\theta)$ response), and cannot detect irregularities exactly, because they occur differently at f_r and f_L.

2.6.2 Application to error-corrected measurement techniques

Further improvements can be achieved, if the PI-network is used together with an error correction technique as described in several publications[5,6,7,8]. The algorithm described in this paper can be implemented easily into that method.
As the error corrected PI-network technique is basically an impedance measurement method, the load resonance frequency can be found at that frequency, where the crystal impedance magnitude has the value given in equation [10]. With error correction the measurement results will be nearly completely independant of the used measurement jig (PI-network or others). This would e.g. allow to replace the IEC PI-network by a modified construction with lower attenuation to overcome limitations regarding the maximum applicable drive level at lower frequencies.
In the same way an implementation into other impedance measurement methods such as the s-parameter technique[9,10] can be performed, which actually do not offer any load resonance measurement feature.

3. Summary

A new software-assisted measurement method for the load resonance of crystal units based on the PI-network method of IEC 444 could be presented. The article describes in detail the fundamentals as well as all equations necessary for the implementation of the method. It includes the results of a detailed error analysis, which shows the superiority of the new method over the existing classical method of IEC 444-4 and over previously published methods , which calculate the load resonance from normal resonance frequency measurements.

The method will be included in a forthcoming proposal for IEC.

---

[1] In the following formulas, the series resonance frequency f_s is replaced by the resonance frequency f_r (at phase zero). This approximation is sufficently accurate in the considered frequency range.

[2] An optimum choice for f_{Co} is given in a proposal agreed at the IEC TC49 / WG 6 meeting (Brighton, July 1989) as follows:
For crystals up to 30 MHz, the measurement is made at 5 frequencies slightly above 30 MHz; i.e. 30.1, 30.2, 30.3, 30.4, and 30.5 MHz; and the average of the three values nearest the mean is used as the best estimate.
For crystals above 30 MHz, measurements are recommended at three paired frequencies, each equidistant from fs, i.e., $f_s*(1\pm0.05)$, $f_s*(1\pm0.06)$, and $f_s*(1\pm0.07)$. For each pair, the value of C_0 is calculated of the lower and higher frequency, and a value of $C_0(i)$ is obtained as the mean of the two C_0' values. This gives three estimates, $C_0(i)$, i=1,2,3. The best value to be used for C_0 is the average of the two values which are closest to each other.
In either case, care should be given that the measurement frequencies are not coincidental with an unwanted mode of vibration

[3] Neubig, B.W.: Measurement of Quartz Crystal Units with Load Capacitance ; TELE QUARZ GmbH, D-6924 Neckarbischofsheim 2 (W.-Germany), Extra print 10/89

[4] Sauerland, F.+ Saunders, J.: General Xtal Test Methods Related To Oscillators and Transmission Instruments; Proc. 3rd Quartz Crystal Conference, Kansas City (Aug. 1981), p.210-228

[5] Neubig, B.W.: Measurement of Quartz Crystal Units up to 500 MHz and Above by the Use of a PI-Network with Error Correction; Proc. 11th Quartz Devices Conference,Kansas City (Aug.1989), Vol.1, p.60 ff.

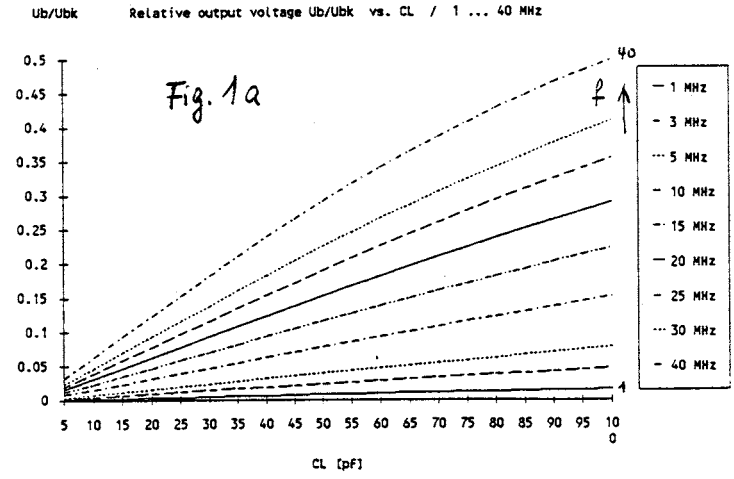
[6] Neubig, B.W.: Fortschritte auf dem Gebiet der Schwingquarz-Meßtechnik bis 500 MHz; Tagungsband ZVEI-Symposium Schwingquarze Frankfurt/M. (14.6.1989), pp.106-126

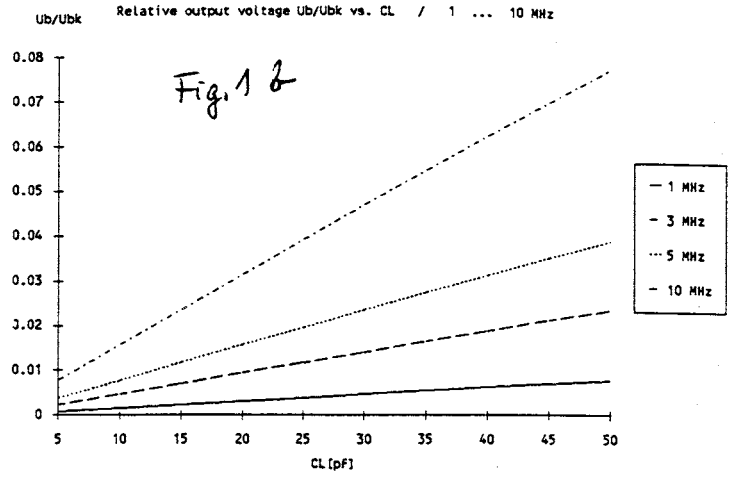
[7] Neubig, B.W.+ Zimmermann, R.: Recent Advances on the Measurement of Quartz Crystal Units up to 500 MHz and Above by the Use of a PI-network with Error Correction; Proc. 2nd International Conf. on Frequency control and Synthesis; Leicester/UK (April 1989), p. 11 ff; IEE Conference Publ.No.303

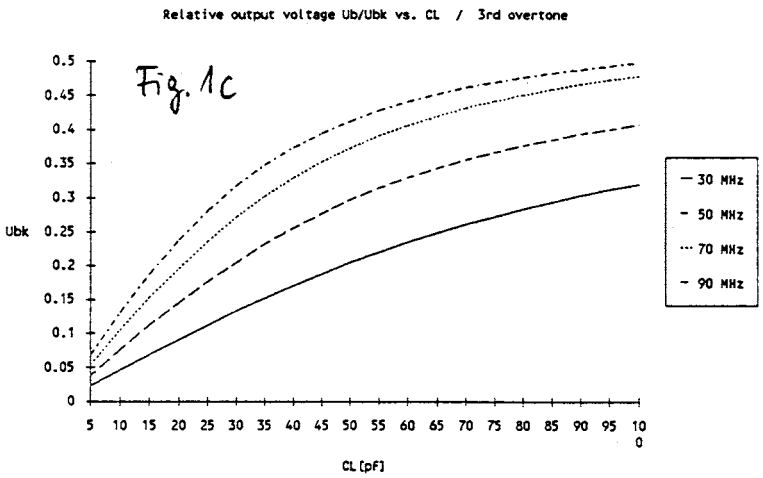
[9] Peach, R.C. + Morris, S.E.: A System for Precission Measurement on Quartz Crystal Resonators And Bipoles; Proc. 39th Annual Symposium on Frequency Control Philadelphia/USA (1985), pp. 527-534

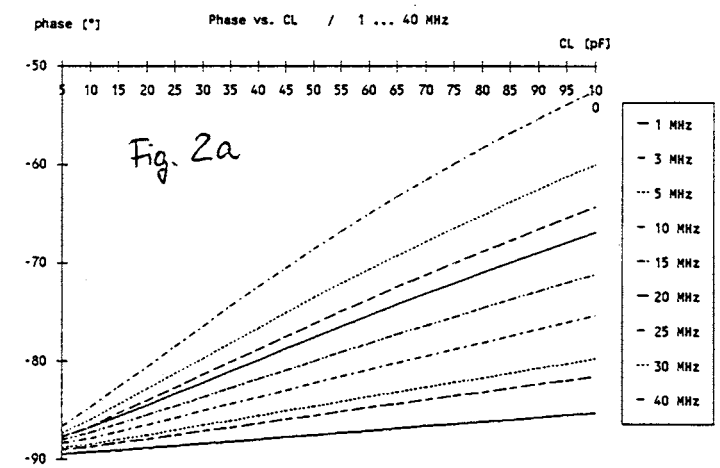
[10] EIA-512: Standard Methods for Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units, 1kHz to 1 GHz; Electronic Industries Association, Washington/USA (1985)

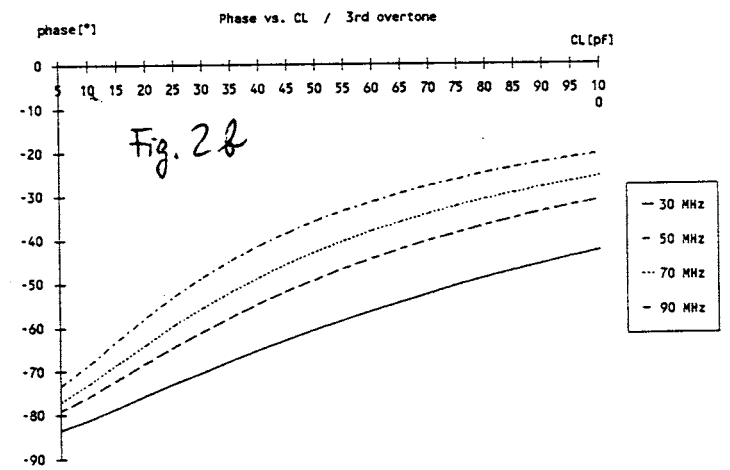
[11] At the presentation of the paper during the 4th EFTF we had to learn from R.C.Smythe that the idea was already suggested earlier in:
Horton,W.H., Payne,T.S., Smythe,R.C. + Symonds,D.A.: Comparison of Methods for Measurement of Quartz Crystal Resonantors with Load Capacitance; Proc. 35th Annual Symposium on Frequency Control Philadelphia/USA (1981), pp.271-279.
Nethertheless, the detailed description suitable for direct application to the PI-Network method as well as the extensive error analysis in our actual paper are unique.

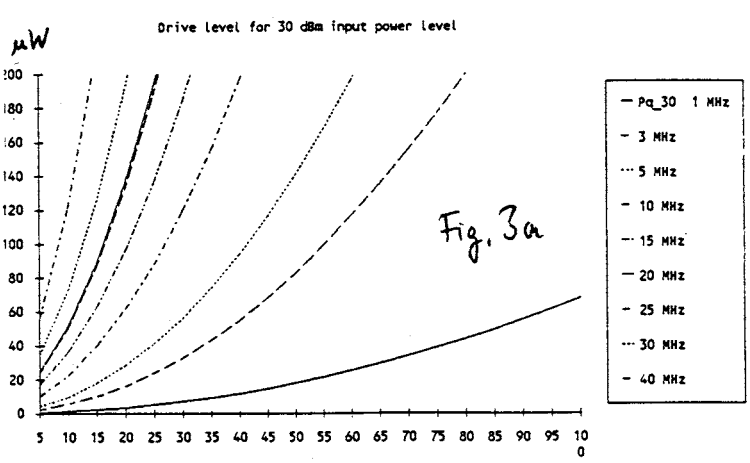

Ub/Ubk
Relative output voltage Ub/Ubk vs. CL / 1 ... 40 MHz
Fig. 1a
f
40
1
1 MHz
3 MHz
5 MHz
10 MHz
15 MHz
20 MHz
25 MHz
30 MHz
40 MHz
CL [pF]

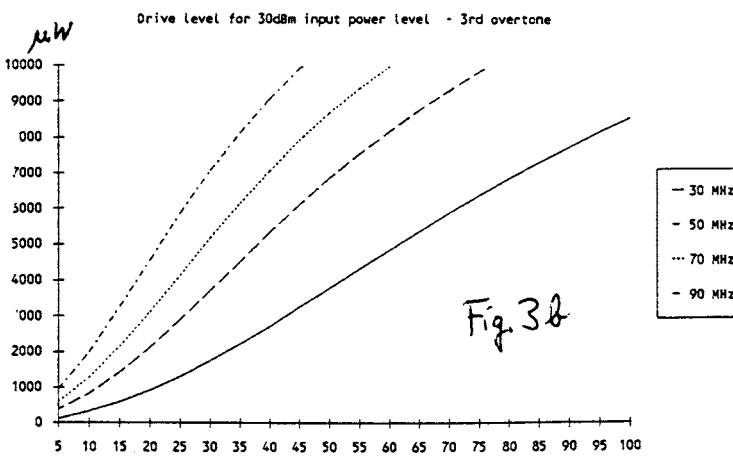

Ub/Ubk
Relative output voltage Ub/Ubk vs. CL / 1 ... 10 MHz
Fig. 1b
1 MHz
3 MHz
5 MHz
10 MHz
CL [pF]


Relative output voltage Ub/Ubk vs. CL / 3rd overtone
Fig. 1c
Ubk
30 MHz
50 MHz
70 MHz
90 MHz
CL [pF]


phase [°]
Phase vs. CL / 1 ... 40 MHz
CL [pF]
Fig. 2a
1 MHz
3 MHz
5 MHz
10 MHz
15 MHz
20 MHz
25 MHz
30 MHz
40 MHz


phase [°]
Phase vs. CL / 3rd overtone
CL [pF]
Fig. 2b
30 MHz
50 MHz
70 MHz
90 MHz


µW
Drive level for 30 dBm input power level
Fig. 3a
Pq_30 1 MHz
3 MHz
5 MHz
10 MHz
15 MHz
20 MHz
25 MHz
30 MHz
40 MHz
CL [pF]


µW
Drive level for 30dBm input power level - 3rd overtone
Fig. 3b
30 MHz
50 MHz
70 MHz
90 MHz
CL [pF]

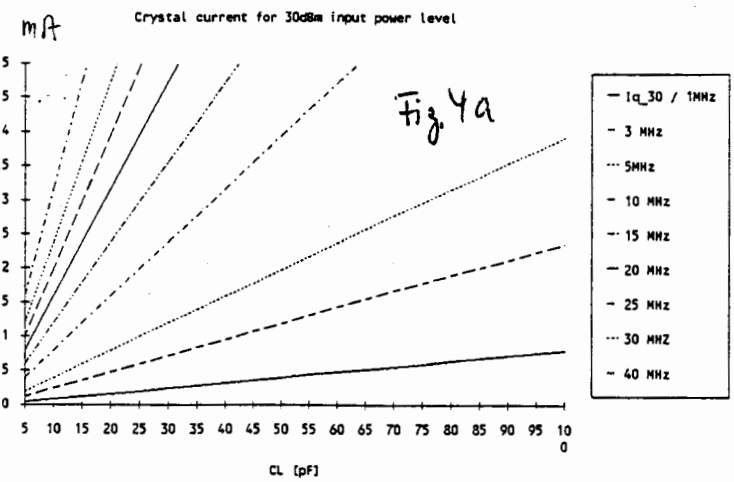


Fig. 4a

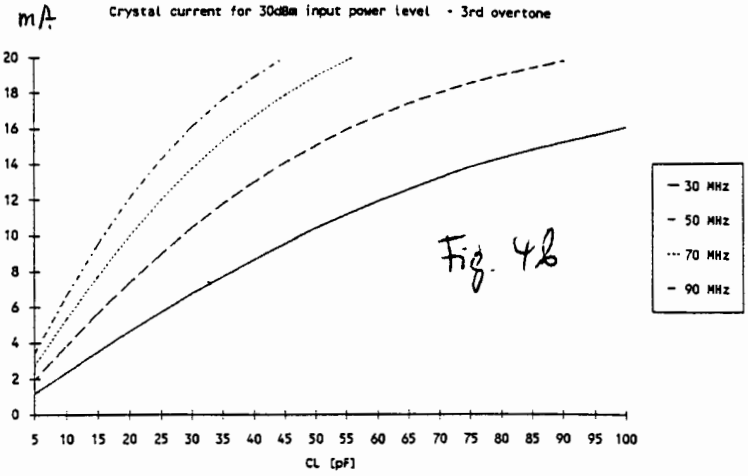


Fig. 4b

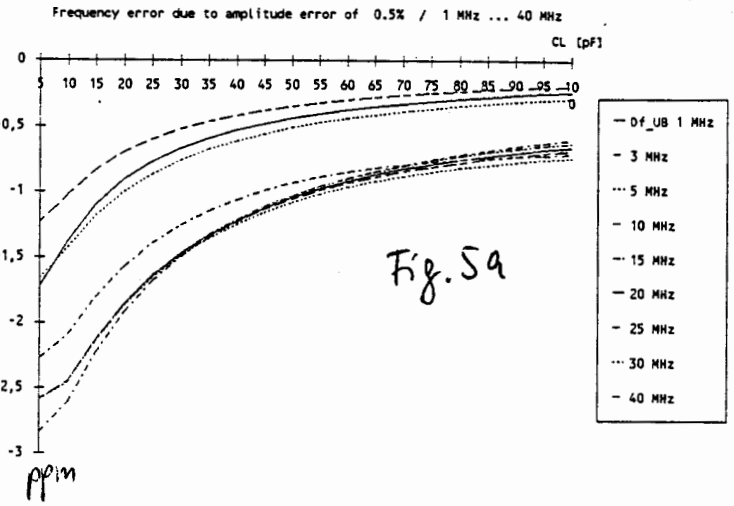


Fig. 5a

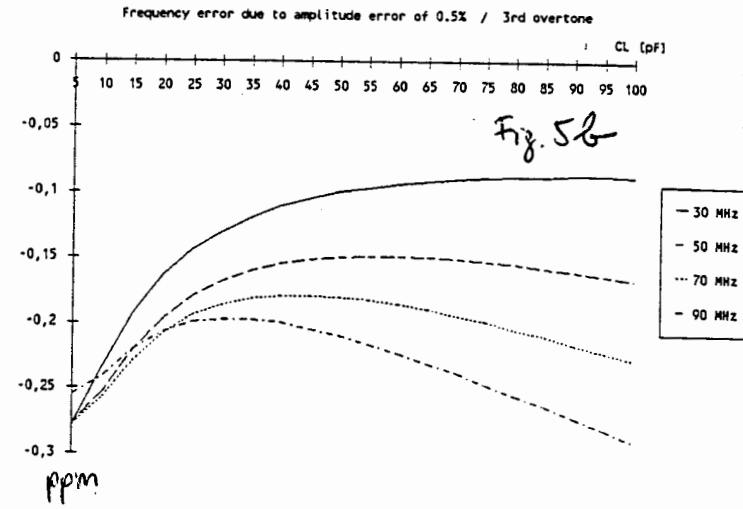


Fig. 5b

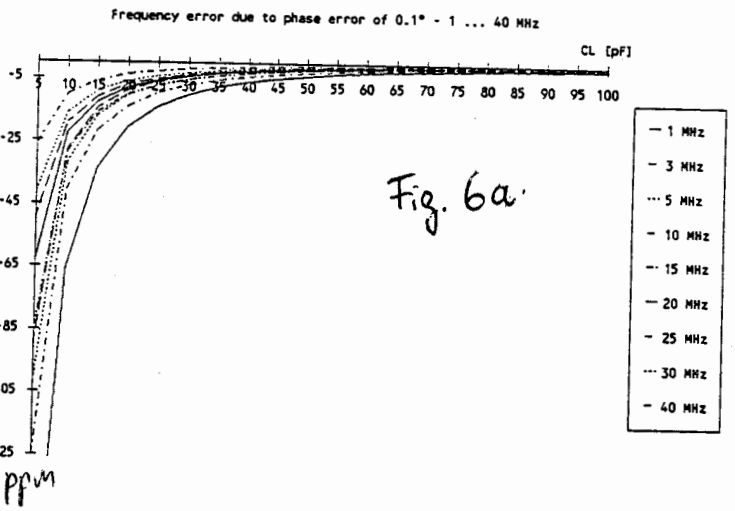


Fig. 6a

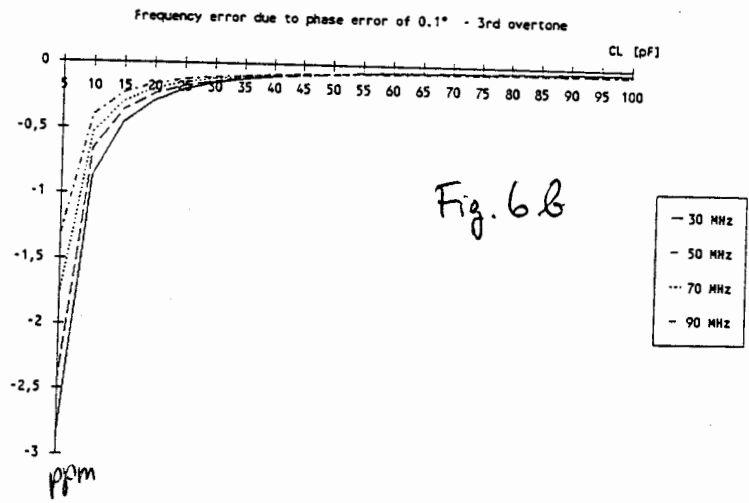


Fig. 6b

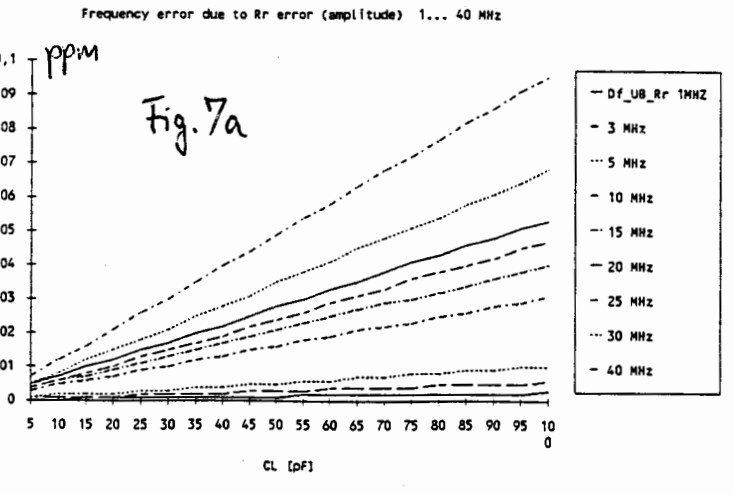


Fig. 7a

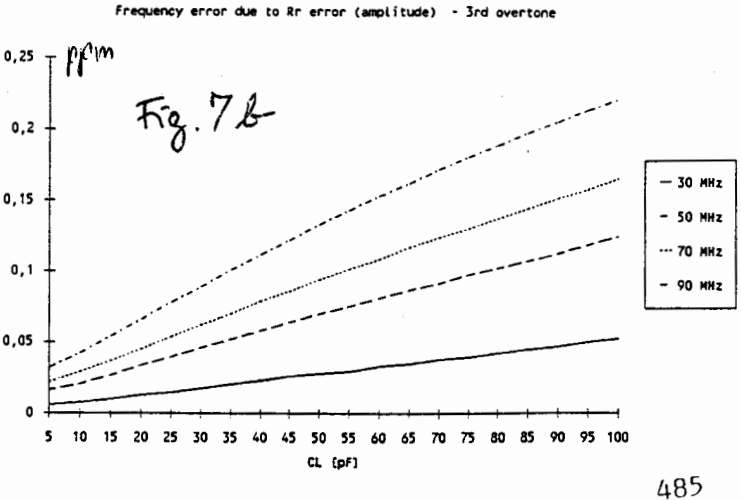


Fig. 7b

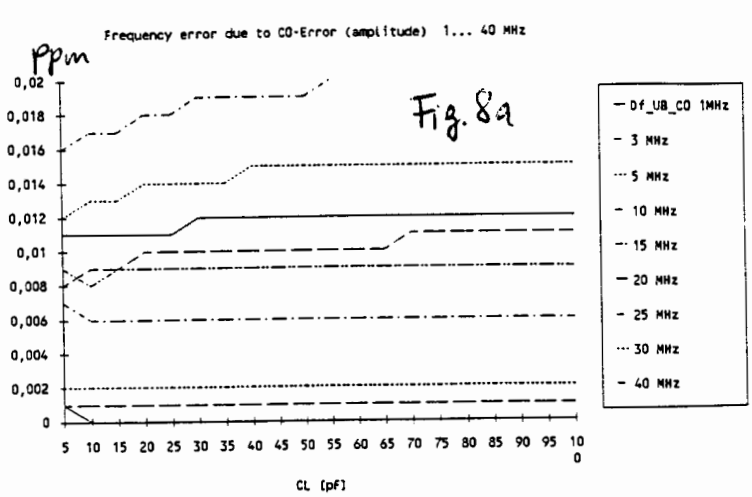


Fig. 8a

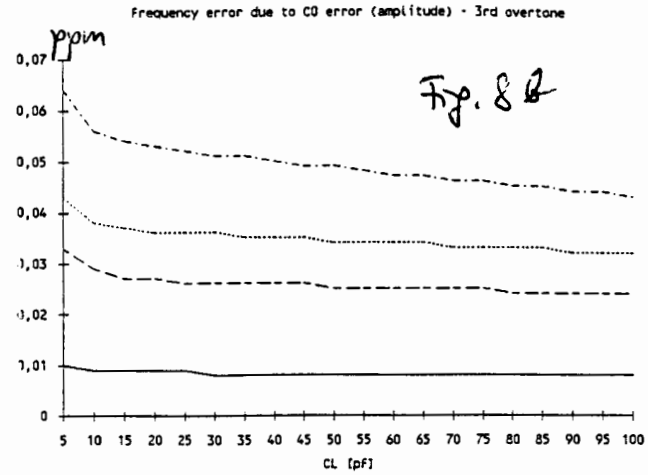


Fig. 8b

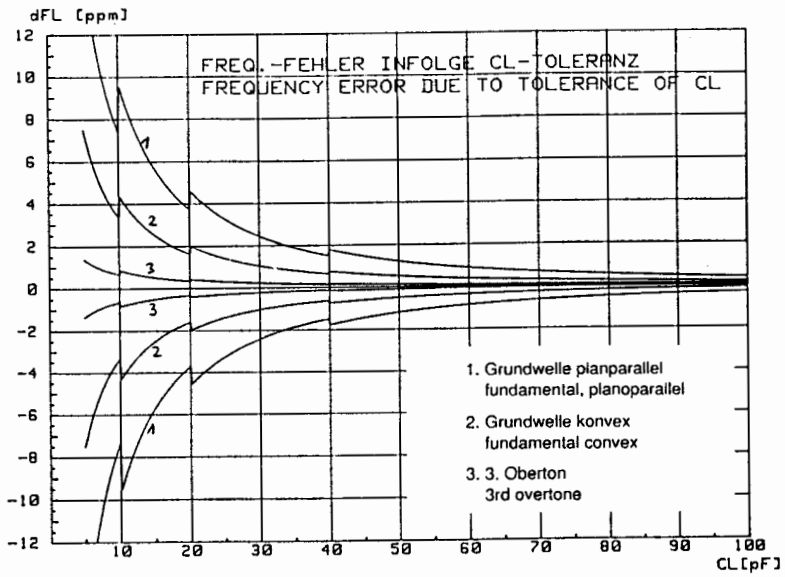


Fig. 9

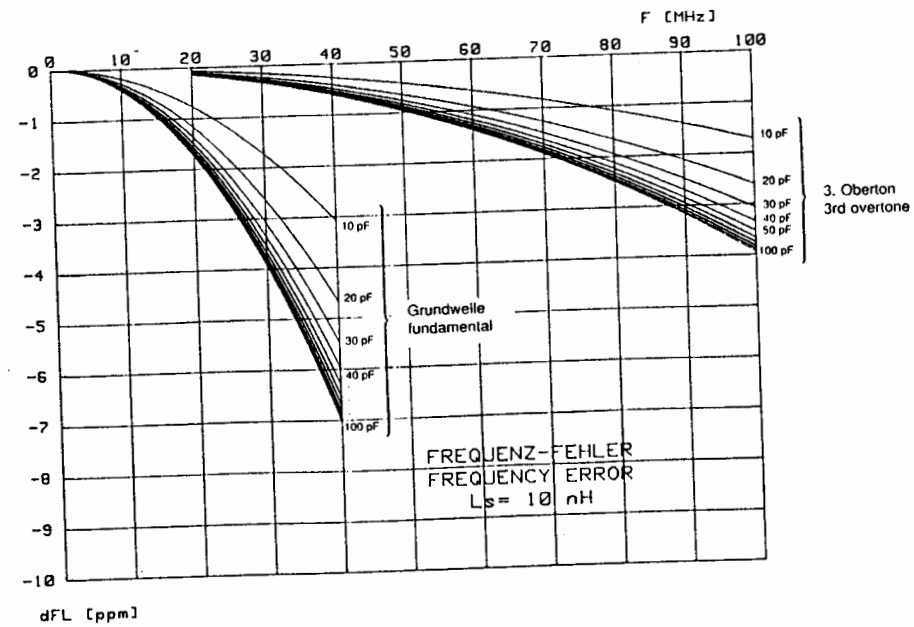


Fig. 10

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-03 Number 2

THEORETICAL ANALYSIS OF INFRARED OVENS FOR TACTICAL OSCILLATORS

S. GALLIOU, M. MOUREY, R. DELAITE*, J.P. VALENTIN

ECOLE NATIONALE SUPERIEURE DE MECANIQUE ET DES MICROTECHNIQUES
Laboratoire de Chronométrie, Electronique et Piézoélectricité
Route de Gray - La Bouloie - 25000 BESANCON - FRANCE

* INSTITUT DE GENIE ENERGETIQUE
Parc Technologique - Avenue Jean Moulin - 90000 BELFORT

ABSTRACT

The feasibility of crystal resonator heating by thermal radiation is examined. A theoretical analysis and simulations show the heating of quartz crystal resonators by infrared oven is feasible. The corresponding warmup times are computed and the advantages of the heating by radiation are pointed out.

INTRODUCTION

The goal of this work is to carry out theoretical analysis of an infrared oven for future realizations of miniature fast warmup oven controlled oscillators. Those oscillators must correspond to the following main specifications :

- thermal range (1^{rst} step) : - 40°C + 90°C
- thermal range (2^{nd} step) : - 55°C + 100°C
- frequency : 10 MHz
- stability ($\Delta f/f$) : 10^{-8} for full range
- aging : 10^{-8}/year
- power consumption : < 500 mW
- warmup time : < 2 minutes at 25°C
- size (without battery) : < 30 cm^3
- mass (without battery) : < 50 g

The principal difficulty is warmup, so that paper is turned to thermal problems.

THERMAL PROBLEMS

Thermal problems are connected to thermal range, warmup time, energy and power delivered by the battery, thermal hysteresis, thermal shock and retrace.

In order to find solutions for those thermal problems we have choosen to improve three areas : crystal cut, oven and mounting. For the crystal cut we have choosen SC cut, of course, and for oven and mounting new designs have been performed.

OVENS FOR FAST WARMUP

To perform an oven for fast warmup, we have considered two possibilities : heating by conduction and heating by radiation. The following figures show schematically those two possibilities.

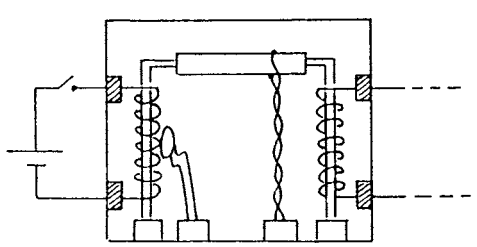


Figure 1 : Conduction heating design

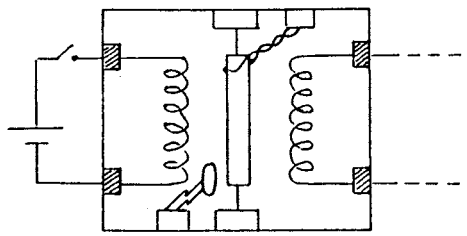


Figure 2 : Radiation heating design

By conduction, the quartz crystal is heated by the means of two heaters rolled on the quartz supports. The schematical figures show the electrodes wires and the thermal sensor.

By radiation, the quartz crystal is heated by the means of two heaters settled opposite to the quartz.

ELECTRICAL ANALOGY

The electrical analogy is very suitable for the study of thermal analysis : thermal power becomes a current, temperature becomes a voltage, thermal resistance a resistor and thermal capacity a capacitor.

Figure 3 presents the electrical analogy corresponding to the conduction heating design.

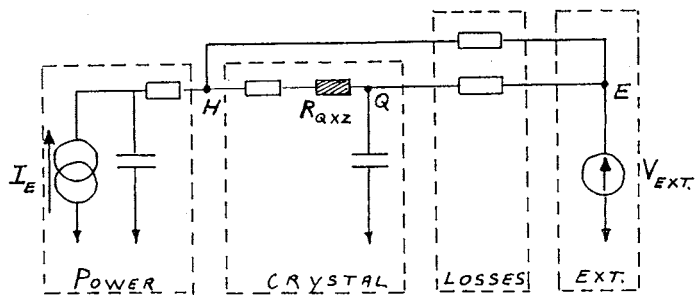


Figure 3 : Electrical analogy for conduction heating design.

The power part points out the power supply including the two heaters by the means of the current supply I_E. The crystal part points out the quartz crystal with the thermal resistances of supports and quartz, R_{QXZ} being the resistance in the crystal main plane. The losses part takes into account the heaters wires and electrodes wires losses and the last part presents the external conditions by the voltage supply V_{EXT}.

Figure 4 shows the electrical analogy for the radiation design.

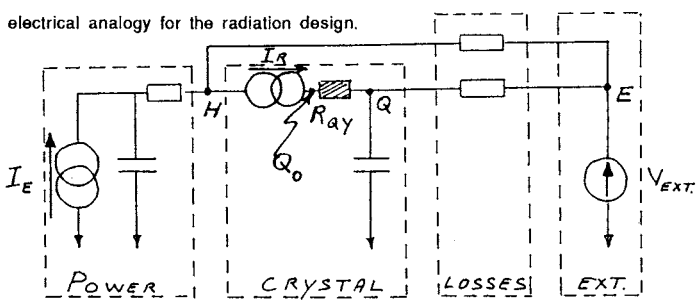


Figure 4 : Electrical analogy for radiation heating design.

In the case of radiation, we have the same parts. The difference is on the resistance RQY, perpendicular to the crystal main plane, and on the current supply IR which denotes the radiative power gathered by the quartz surfaces from heaters. That current is given by $I_R = (\theta_H^4 - \theta_{Qo}^4)$. In that equation, θ_H and θ_{QD} are respectively the temperature of H and Q_0 points. As consequence of the strong non linearity of that expression, the computation is necessary. We have choosen an electrical simulation in the case of proportional controlled loop, i.e. with $I_E = I_0 + G(\theta_{REF} - \theta_{SEN})$ where I_E is the current indicated on the figures 3 and 4, I_0 a constant, G the gain, θ_{REF} the reference temperature and θ_{SEN} the thermal sensor temperature. In the case of heating by conduction the thermal sensor is sticked on the heater ($\theta_{SEN} = \theta_H$), and in the case of heating by radiation the sensor is the B mode ($\theta_{SEN} = \theta_{QD}$). The results are given by the following graphs.

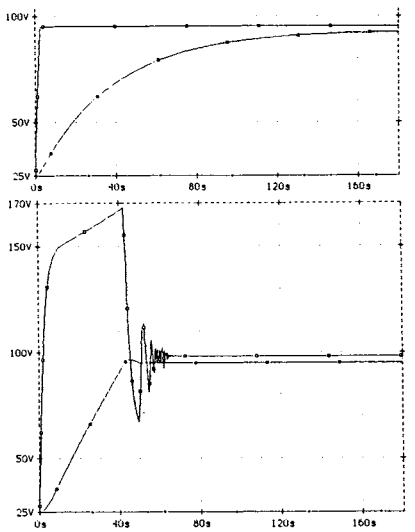


Figure 5 : Warmup with external temperature = 25°C.

Results for conduction are given at the top and at the bottom for radiation. The white squares are the temperature variations versus time for the heater and the black squares for the quartz crystal. On the figure 5, the warmup time is 1 minute for radiation and about 3 minutes for conduction.

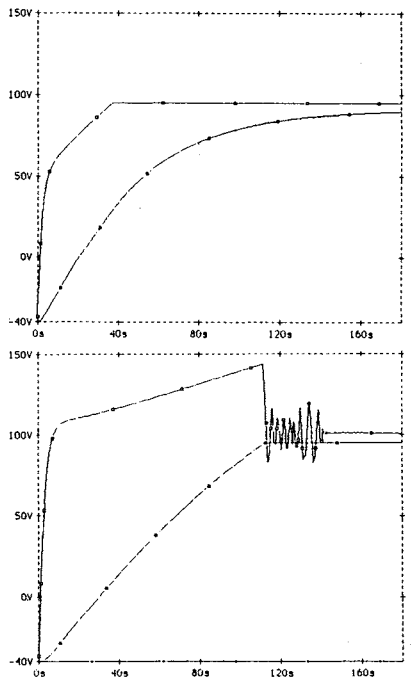


Figure 6 : Warmup with external temperature = - 40°C.

On the previous figure the simulations are obtained for an external temperature equal to - 40°C. Warmup time is 2 minutes in the case of radiation and 3 minutes for conduction.

At last, for an external temperature equal to 90°C, the warmup time is the order of 1 minute for radiation and 3 minutes for conduction (see Figure 7).

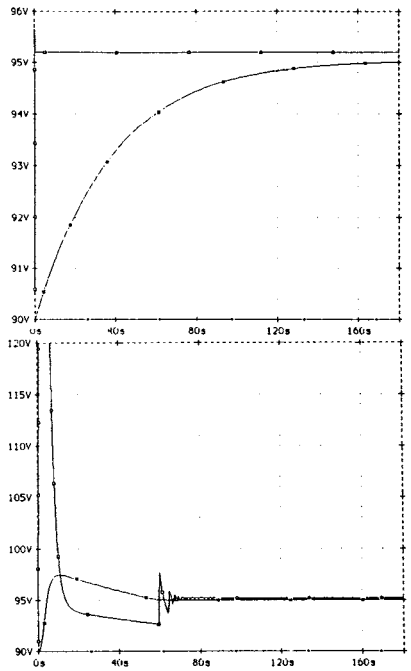


Figure 7 : Warmup with external temperature = 90°C.

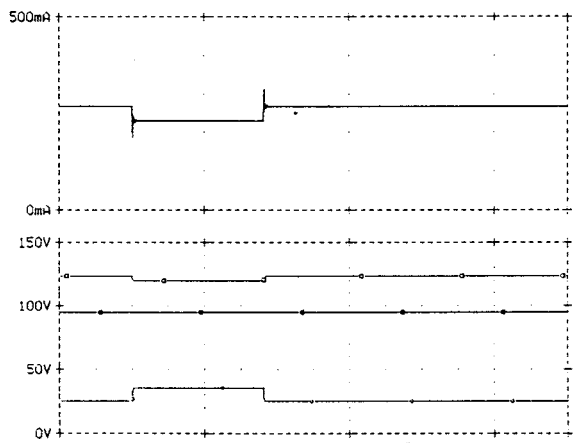


Figure 8 : Regulation with proportional control system for heating by radiation.

Figure 8 shows the regulation with proportional control system is possible in the case of heating by radiation. The curve at the bottom is the external temperature v. time : a disturbance of 10°C is applied. The curve at the top is the current variation in the heater v. time. The white squares are the heater temperature and the black squares the quartz crystal temperature.

ADVANTAGES OF RADIATION

In abstract, the advantages of heating by radiation can be summed up so :

- a faster warmup is theoretically obtained (factor 3 at 25°C and 90°C) and as consequence warmup times are reduced,
- heating is more homogeneous. The thermal flux flows through the whole surfaces of the crystal and not only by two points (the holders). So the thermal dynamic effects are reduced,

- the ratio RQY/RQXZ < 10^{-3} makes the thermal shock cancelled in the case of radiation when a large heating stop is applied on a cold crystal. So the break down risks are very reduced,

- in the case of heating by radiation, the holders are not used for transfer in the thermal flux and, as consquence, specific mounting is autorized. So mechanical stresses, and by the same way crystal quartz thermal losses, can be reduced.

In the goal to perform the crystal suspension, we have realized two new designs which are based on two main ideas : the use of the contact thermal resistance and the use of a stressless mounting.

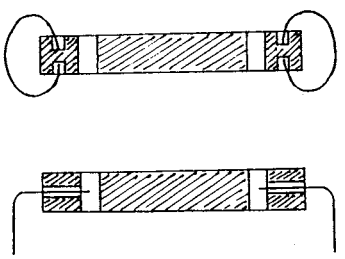


Figure 9 : Two new designs for thermal isolation and stressless mounting.

The first design (Fig. 9) uses an isolation by bridges as for BVA resonators and elastic mounting made by phynox ^{R} wire.

The second design shows lateral holes, and phynox ^{R} wire, which can slip inside the Roles, providing the free dilatation of the crystal.

SPECIFIC PROBLEMS

Before examining a precise design for heating by radiation, let's see the specific problems of heating by radiation and their solution :

1) Degazing by hot wires : using a bulb which retains the gases inside.

2) Quartz is transparent for light : working in infrared with wavelength greater than 4 µm.

3) The WIEN law imposes the wavelengh λ_{max} for maximal emittance at the θ temperature : at 500° K, λ_{max} equals 6 µm, therefore largely greater than 4 µm.

4) Gilded electrodes deposited on the quartz surface are reflectors for radiation : making pierced electrodes as a wire-lattice.

5) Radiation losses by absorption on the different surfaces are very reduced by polishing.

6) Thermal sensor cannot be given up inside enclosure without caution. It must be closely set at a thermal support. That support can be bulb, or the crystal itself, when the B mode is used.

SIMPLIFIED DESIGN

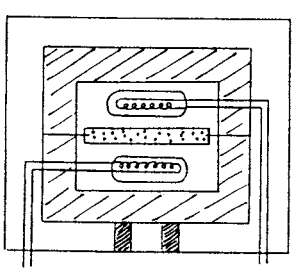


Figure 10 : Simplified design

The previous figure shows schematically our design approved for heating by radiation. We consider a quartz crystal maintained by the free dilatation mounting previously presented, and heated by the means of two bulbs. Reflecting shells surround quartz and bulbs and the whole is set inside a can, by an isolating leg. The simplified design clearly shows the mounting symmetry, except for the leg (the thermal conductance of this last on will be distributed on all the outer shell surface in the following thermal exchange equations).

Using the first letter of the name of each solid to denote its temperature (see fig. 10), we consider that one solid behavior is defined by its surface temperatures and a mean inside temperature. Moreover we assume that surfaces are isothermal. According to the notations of fig. 11 where we take the example of a half-shell, we can write :

* **Energy conservation :**

$$(-S_{Si}\overset{R}{\varphi}_{Si}) - (S_{S0}\overset{R}{\varphi}_{S0} + S_{S0}\,k_{SC}\,(\theta_{S0} - \theta_c)) = \rho_S\,C_S\,V_S\,\frac{d\theta_S}{dt}$$

input influx output influx

* **Boundary conditions :**

On the outer surface : $k_{SC}\,(\theta_{S0} - \theta_c) + \overset{R}{\varphi}_{S0} = -\lambda_S\,\frac{\theta_{S0} - \theta_S}{l_S}$

On the inner sufarce : $-\overset{R}{\varphi}_{Si} = -\lambda_S\,\frac{\theta_S - \theta_{Si}}{l_S}$

considering an opaque solid and where ρ_S, C_S, λ_S, V_S are respectively the density, the specific thermal capacity, the thermal conductivity, the volume of the half-shell.

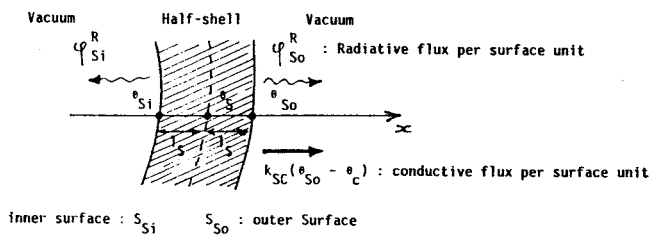


Figure 11

All the thermal transfers are summed up in Fig. 12

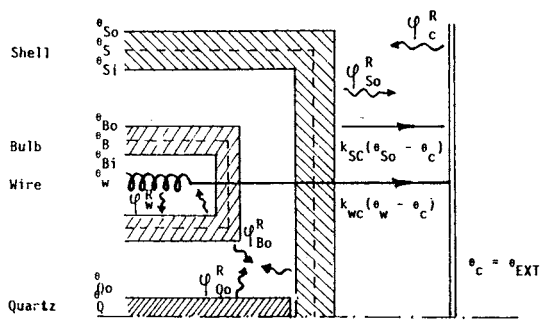


Figure 12

The analysis of those thermal transfers leads to a nine-equations system whose variables are the wire temperature θ_W, the bulk temperatures θ_{Bi}, θ_B, θ_{Bo}, the quartz temperatures θ_{Qb}, θ_Q and the half-shell temperature θ_{Si}, θ_S, θ_{So} versus the electrical applied power P and the can temperature θ_c (supposed to be equal to the ambient temperature). In fact the main problem is the calculation

of the radiated flux which are functions of the surface temperatures for opaque solid. Moreover, quartz and glass are not opaque for all the wavelengths. They may be considered as transparent solids from 0 to 4 µm : this is expressed by the Z(θ) function in the flux expression. Z(θ) is the ratio of the emitted energy in O, 4 µm on the total emitted energy by a surface at the temperature θ .

As an example we mention the wire radiated flux :

$$S_W \overset{R}{\varphi_W} = S_W F(f_{ij}, \varepsilon_{Si}) \sigma [Z(\theta_W)\theta_W^4 - Z(\theta_{Si})\theta_{Si}^4] + S_W \sigma [(1 - Z(\theta_W))\theta_W^4 - (1 - Z(\theta_{Bi}))\theta_{Bi}^4]$$

where $F(f_{ij}, \varepsilon_{Si})$ is a function of the shell surface emissivity ε_{Si} and of f_{ij} which are shape coefficients of bulb, quartz and shell surfaces. Here the wire and bulb surface emissivities when the bulb is opaque are supposed to be equal 1, the above expression points out the transparency of glass and quartz in the band 0, 4 µm where the wire surface (S_W) exchanges only with the inner surface of the shell (S_{Si}).

In the vacuum inside the shell the radiated flux of each surface are functions of θ_{Qo}, θ_{Si}, θ_{Bo} and θ_W (in the band of transparency) because of the multiple reflections.

Without anymore details let us only say that the radiated flux expressions are strongly non-linear. So it is convenient to use the electrical analogy for the computer-aided analysis, as previously.

Figure 13 shows the basic electrical equivalent circuit of those thermal transfers

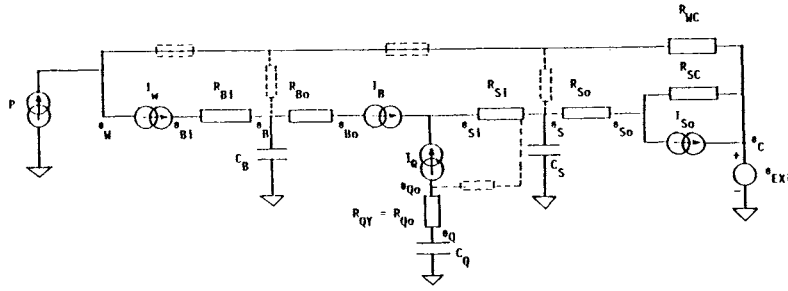


Figure 13

where $S_W \overset{R}{\varphi_W} = I_W = f_W(\theta_W, \theta_{Bi}, \theta_{Si})$ $\quad S_{Bo} \overset{R}{\varphi_{Bo}} = I_B = f_B(\theta_W, \theta_{Bo}, \theta_{Qo}, \theta_{Si})$

$S_{Qo} \overset{R}{\varphi_{Qo}} = I_Q = f_Q(\theta_{Bo}, \theta_{Qo}, \theta_{Si})$ $\quad S_{So} \overset{R}{\varphi_{So}} = I_{So} = f_{So}(\theta_{So}, \theta_c)$

From that figure we notive that the wire thermal capacitance is neglected. Here, R_{Qo} denotes the quartz thermal resistance in the Y'-quartz direction. Dotted line resistances may be added in order to take account conduction of heat such as feeding wires-glass bulb transfer, quartz-shell transfer through the quartz-plane and suspension resistances.

WARM-UP WITH REALISTIC DATA

Usually, the higher the power step the faster the warm-up is. It is not that way in our case because of the band of transparency for glass and quartz (see fig. 14). When increasing the power supply the wire temperature goes up inducing larger losses accross RWC (Fig. 13) and else a bigger heat absorption by the shell (whose polished surface emissivity is only ε_{Si} = 5 % in our case !).

This is not a good manner to have fast warm-up ! Black bulbs may be a solution to that problem. In such a way, all the wire radiated energy is trapped in the bulb and almost entirely usable by the quartz crystal.

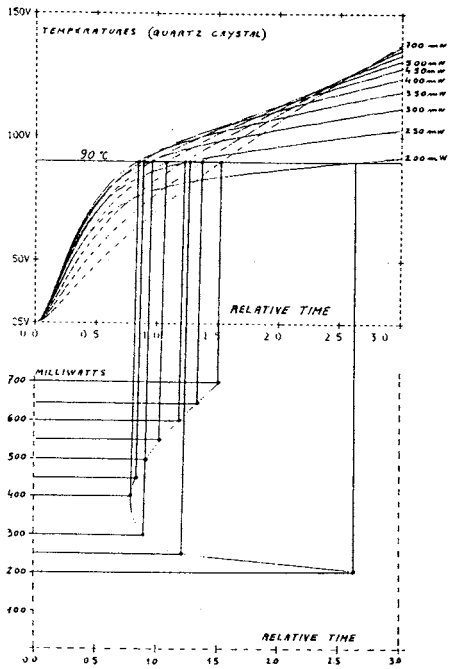


Figure 14

Figure 15 shows the simulation results with black bulbs, the other components being the same as above.

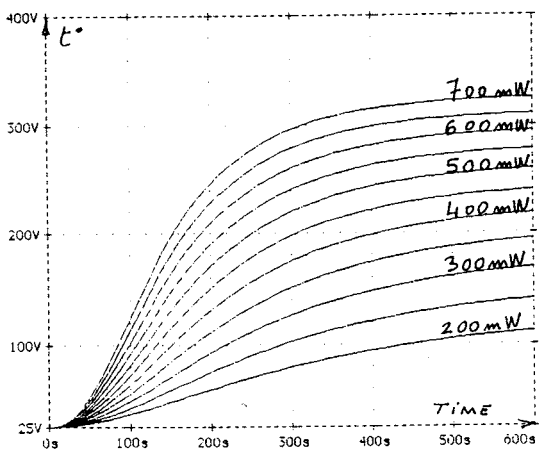


Figure 15

CONCLUSIONS

The theoretical analysis and the simulation with electrical analogy nettworks show the heating of quartz crystal resonator by infrared oven is feasible. The computed warmup time is in the order of one minute at 25°C or 90°C, and two minutes at - 40°C. The advantages of heating by radiation are the thermal dynamic effect decreases, the break down risks are reduced and a free dilatation mounting can be used. At last, the B mode of the SC cut can serve as thermal sensor.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the efforts of Eric ANDREY and Alain ROYER in prociding assistance in performing of test units and Boutros ABBOUD for mounting and testing of the first prototypes.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-03 Number 3

LOW PHASE NOISE, HIGH FREQUENCY TCXO
Actual state of Art

Louis BIDART, Emmanuel GIRARDET, Jean HUET, Patrick RENOULT

Q K PIEZOELECTRONIQUE, 30 Bld Galliéni, F 92390 VILLENEUVE LA GARENNE

INTRODUCTION

In today's technologies, a lot of telecommunications, localisations, radar systems need high stability frequency sources which provide a very good phase noise at frequencies that go as high as possible (typically around 100 MHz or more).

The best performances for both stability and phase noise are reached with ovened oscillators (OCXO). However, when the size, the power consumption, the "start up " become critical problems, the temperature controlled oscillator (TCXO) can give the best compromise.

The most important characteristic of TCXOs is that the stability (especially versus temperature) and the phase noise are tightly connected.

This report will present the different aspects that must be taken into account and how to improve them. Different designs will be shown and the performances will be compared. There is not simple solution. Each oscillator must be designed regarding a precise and complete specification.

1. LOW NOISE, HIGH FREQUENCY TCXO - PRESENTATION

The only objective in this report is the consideration of oscillators which require the following characteristics or performances :

(in brackets : targets)

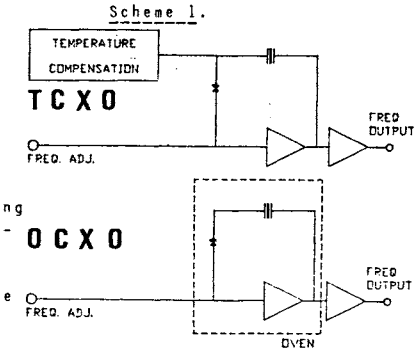
Frequency range	: 80 - 120 MHz	
Temperature range	: - 20 $\leq \Theta \leq$ + 70°C	(-40 $\leq \Theta \leq$ + 85°C)
Stability vs Temp.	: $\pm$ 1.10^{-6}	($\pm$ 5.10^{-7})
Long term stability	: $\pm$ 1.10^{-6}/year	($\pm$ 5.10^{-7}/year)
Phase noise (dBc/Hz)	: 10 Hz -75	(80)
	100 Hz -105	(110)
	1 KHz -135	(140)
	Noise Floor -150	(155)

Of course, other specifications must be taken into account when designing the oscillator, for example :

- Output signal / level
- Size
- Environment strains (g and acoustic sensitivity)
- Supply voltage / consumption

but they will not be discussed in this report.

Scheme 1.



In designing TCXOs, the greatest difficulty is the simultaneous optimization of both parameters stability and phase noise. With the OCXOs, the thermal regulation is completely independent of the performing oscillator. Whereas, in the case of the TCXOs, the temperature compensation works directly in the oscillator loop. (scheme 1)

The analogic compensation process in the TCXO is well known, none the less it is necessary to show what the main parameters are and how to optimize them.

2. ANALOGIC COMPENSATION

2.1. PRINCIPLE - DEFINITIONS - (Curve 2.)

- The used resonator is always an AT Cut Quartz. The angle is chosen in such a way that the frequency drift in temperature range is optimized (typ. $\pm$ 15.10^{-6}, $-20 \leqslant \Theta \leqslant +70°C$).
The frequency is controlled under the voltage applied on the varactor (varicap diode), connected in serie with the resonator in the oscillator loop.

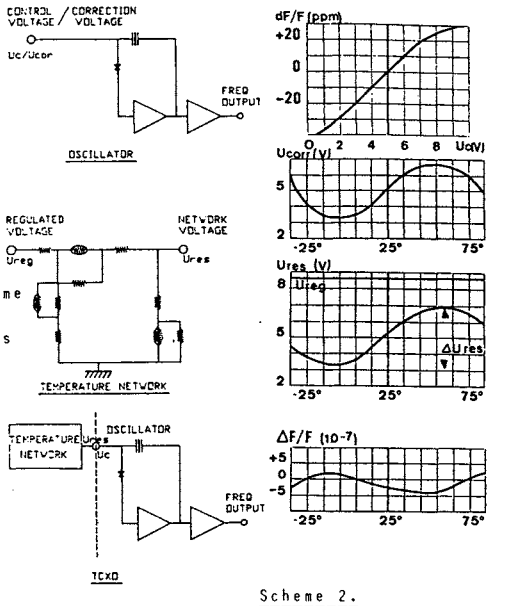
- In a given temperature range, it is possible to conserve a stable output frequency by modifying the control voltage Uc. The curve Ucorr (Θ) has approximatively the same form as the frequency vs tempe. curve of the Quartz (3rd order).

- The analogic compensation network has 3 thermistors and some resistors. They are assembled in such a way that from a regulated voltage, a 3rd order voltage curve Ures (Θ) occurs. The resistors values are calculated so that the Ures (Θ) curve is as close as possible to the Ucorr (Θ) curve.

- The Ures (Θ) voltage is applied on the oscillator. The voltage difference $\Delta U (\Theta)$ = Ucorr (Θ) - Ures (Θ) causes the error frequency vs temperature curve (Typ. $\pm$ 1.10^{-6}, $\pm$ 5.10^{-7}).

The compensation principle is easy to use for the desired frequency stabilities of 1 to 2.10^{-6} ($-20 \leqslant \Theta \leqslant 70°C$).

For wider temperature ranges or best stability values ($\pm$ 1.10^{-6}, $\pm$ 5.10^{-7}), it is necessary to know the requirements of the different parameters in order to determine the limits of this analogic compensation.



Scheme 2.

2.2. LIMITS OF ANALOGIC COMPENSATION NETWORK

The assembly of resistors Ri and thermistors RTHi = $R_0 e^{\beta\left(\frac{1}{T}-\frac{1}{T_0}\right)}$ allows it to generate a 3rd order curve :

Ures (Θ) = a Θ^3 + b Θ^2 + c Θ + d
with a, b, c, d = F (RTHi, Ri, Ureg, $Ures_0$)

This curve is characterized by :

- the average value $Ures_0$
- the dynamic $\Delta Ures = \dfrac{Ures\ max - Ures\ Min}{Ures_0}$

Ures depends on :

- the network design
- the RTHi, Ri values
- the temperature range
- the regulated voltage value
- the chosen average output value $Ures_0$.

For a given configuration, the values of $Ures_0$ and Ures are dependent on each other and are limited.

For example :($-25 \leqslant \Theta \leqslant +85°C$, Ureg = 10 V) :

$Ures_0$ = 5 V Ures max = $\pm$ 1,5 V

$Ures_0$ = 7,5 V Ures max = $\pm$ 1 V

then, one must design the oscillator by considering these limitations.

2.3. COMPENSATION : OSCILLATOR'S PARAMETERS

For both the resonator and the oscillator design, the temperature compensation requires taking into account the contrains and making choices.

2.3.1. Resonator

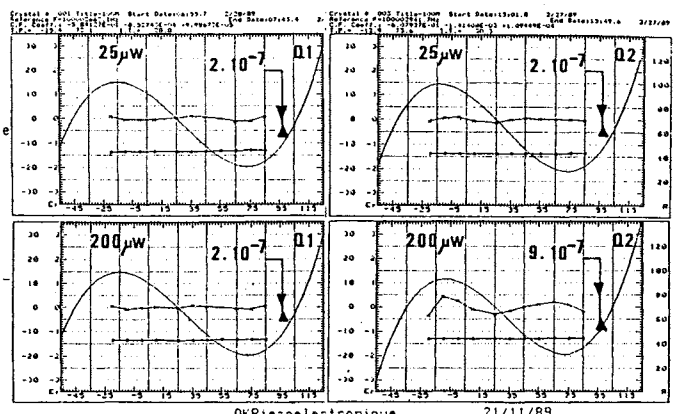
- the Cut angle is chosen for minimizing the frequency deviation versus the given temperature range
- the T. curve is a 3rd order curve which can be somewhat deformed. This is due to the geometrical characteristics of the blank and the mounting structure (strains). This phenomena gets more sensitive as the Fundamental frequency of the resonator increases.

In addition, these disturbances become worse when the drive level increases (anisochronism defect), as shown on curve 3.

- In the same way, there may be some "activity dips" on the frequency vs temperature curve, which come from punctual variations of the Quartz resistance value.

Because of its design, the temperature network however cannot correct these disturbances or activity dips. They can be negligeable if their value are small (some 10^{-7}) and if stabilities of some 10^{-6} are wanted.

These defects become especially critical with stabilities of about 5.10^{-7}.



Curve 3.

2.3.2. Oscillator

In fact, the oscillator is a VCXO, controlled by the temperature network voltage.

- F(Uv) characteristic
(values are given regarding spec. § 1)
The frequency deviation FT must be sufficient in order to compensate :

. the temperature drift $\Delta F_\theta (\pm 15.10^{-6})$

. the long term stability $\Delta Ft (\pm 5.10^{-6})$

$$\Delta FT = \Delta Ft + \Delta F_\theta \quad (\pm 20.10^{-6}) \qquad (1)$$

- The amplitude of the control voltage variation is given by the comp. network (§ 1.1.2.)

$$\Delta UT \leq \Delta \text{ Ures max} \qquad (2)$$

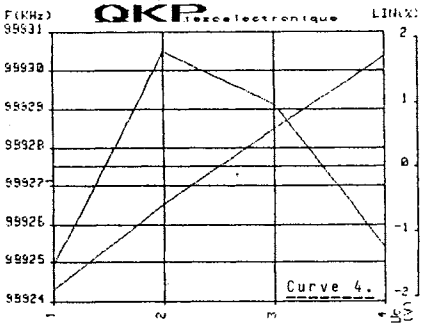
- (1) and (2) define the minimum sensitivity :

$$s \text{ min} = \Delta FT / \Delta UT \quad (\simeq 15.10^{-6}/V)$$

- In order to have the Uc (Φ) curve the same as the resonator one (3rd order), the F(Uc) characteristic must be as linear as possible.

OKPiezoelectronique 21/11/89
SCF No: 1 F: 99.931 MHz
pente moy : 2024 Hz/V (moindres carres)
Fnominale @: 4.3 V
Nb de mesures: 7

No Pt	Uc (V)	Ymes (KHz)	ecart (KHz)	ecart (%)	pente (KHz/V)	deltaP (%)
2	1.0	99924.30	-0.09	-1.5	2.225	94
3	2.0	99926.53	0.11	1.8	2.225	10
4	3.0	99928.50	0.06	1.0	1.975	-2
5	4.0	99930.39	-0.08	-1.2	1.889	-7



In addition, the linearity of the F(Uc) curve must be obtained on the whole frequency range FT in order to conserve a right frequency vs temp. compensated curve on the complete adjustment range (curve 4.)

2.3.4. Practical design

The use of different types of resonators (Fundamental, 3rd, 5th OT) needs adapted oscillators.
Obtaining a sufficient sensitivity and linearity is relatively easy for fundamental crystals (standard varicap diode)
On the other hand, for 3rd and 5th overtones, the using of a single varactor (hyperabrupt diode) can be insufficient, especially for the linearity. It is then necessary to use inductive elements which are mounted in serie with the Quartz. But the operating or loaded Q_L factor will be strongly affected.
An interesting alternative on this can be the use of an op. amplifier (low gain) which is placed between the temp. network and the varactor. Then, the required oscillator sensitivity is divided by the amplifier gain. But of course, the linearity specification remains the same.

3. PHASE NOISE

As for the temperature compensation, the optimization of Phase noise requires specific characteristics for the resonator and especially for the oscillator in the case of TCXO.

3.1. RESONATOR

The specific quality of the crystal can be defined by its Quality factor Qo.

The main factors that affect Qo are the following :

- Frequency, overtone mode
- Blanks geometry / roughness
- Electrodes design / type
- Mounting design / strains
- Impurities / material defects

Qo must be as high as possible.

Table 5 shows values of differenttypes of crystals which can be used in this application.

TABLE 5
TCXO 100 MHz AT QUARTZ
ELECTRICAL PARAMETERS

FREQUENCY MHz	VIBRATION MODE	R Ohms	L mH	Q(x1000) MAX	Q(x1000) AVE
100	3 rd	15	2,5	85	75
100	5 th	40	6.0	290	270
20	FUND	4	2	140	120
33	FUND	4	2	120	100
33	3 rd	11	12	270	240

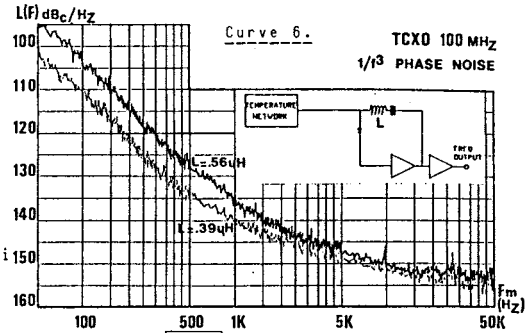
The optimization of signal/noise ratio requires a high oscillation level in the oscillator loop. The crystal drive level must be chosen as high as possible, within the right working range of the resonator (see § 2.3.1.)

3.2. OSCILLATOR CHARACTERISTICS

If the crystal Quality Factor is crucial, the operating QL of the oscillator is even more so.

QL has a direct effect on the phase noise curve close to the frequency carrier (1/F3 noise).

The QL/Qo ratio decreases, so the oscillator's sensitivity increases, especially when such high sensitivity is reached by using inductors, as shown on curve 6.



Curve 6.

TCXO 100 MHz
1/f³ PHASE NOISE

To optimize the Quality Factor QL, the following points must be considered :

- use of abrupt or hyperabrupt varactors
- if the need of inductors is necessary to reach the sensitivity and linearity required (3rd or 5th OT) their values must be minimized
- the use of an op. amplifier between network and varactor allows a reduction of the required sensitivity (ratio 1 to 4).

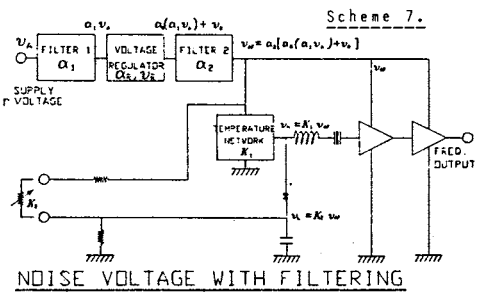
Another important parameter is the Flicker noise (1/F). 1/F noise is also due to the electronics of the oscillator. It is necessary to choose low noise transistors in the working frequency range and to use them correctly.

3.3. PERTURBATIONS INDUCED IN OSCILLATOR'S LOOP

The specific phase noise curve close to the frequency carrier (1/F3, 1/F) can be altered by perturbations induced directly in the oscillator's loop as shown on curve 7.:

- Noise on the oscillator's regulated supply
- Voltage noises on varactor

These last ones have a particulary strong impact in the case of the TCXO where the oscillator's sensitivity is very high.

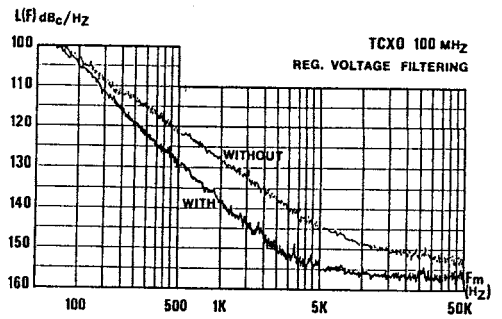


NOISE VOLTAGE WITHOUT FILTERING

Scheme 7.

NOISE VOLTAGE WITH FILTERING

The only way to reduce these perturbations is to filter as best as possible the regulated voltage source, as shown on curve 8.



TCXO 100 MHz
REG. VOLTAGE FILTERING

Curve 8.

4. TEMPERATURE VS PHASE NOISE

The above paragraphs have shown the most important parameters to take into account when designing high temp. stability, low phase noise oscillators.

Some of them can be optimized regarding one parameter (either temp. compensation or phase noise) without influencing other.

For others, the improvement of both parameters requires solutions of contrary senses. It is necessary to find a compromise between these as shown on table 9.
The specification must show the precise operating needs. If not, it could involve technical impossibilities or modest performances regarding one parameter if the other was more important.

For instance, to reach a $\pm 5.10^{-7}$ temp. stability, one must limit the drive level to avoid any perturbations on the temp/frequency curve. the phase noise result is degraded as well.

From the right specification, the oscillator's scheme must be chosen by considering the temp. and Phase noise characteristics, and possibly more other parameters, for example aging, environmental sensitivity, size, etc..

Different solutions can be foreseen :

- direct frequency oscillator/AT 3rd or 5th O.T. resonator
- multiplied Freq. oscillator AT Fund. x 3 /x 5
  AT 3rd O.T. x 3 / x 5

The crystal and oscillator characteristics are very different (Q Factor, Stability, Drive level, Components in oscillator loop, i.e. inductors).

It is interesting to compare the electrical results of different types of oscillators.

TABLE 9

T.COMPENSATION VS PHASE NOISE

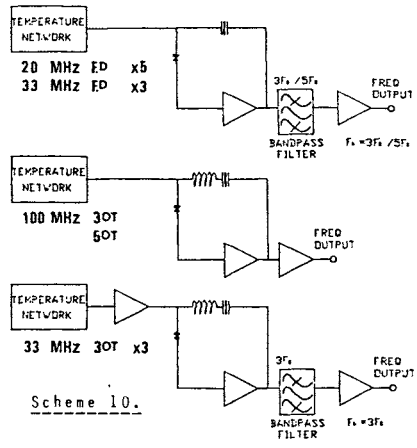
PARAMETER	TEMPERATURE COMPENSATION	PHASE NOISE	COMPROMISE
F(Uc) SENSITIVITY LINEARITY	as high as possible as good as possible	as low as possible not critical	15-20 ppm/V * 0.01 - 0.05
QUARTZ DRIVE LEVEL	low level (activity dips)	high level (signal/noise ratio)	50-100µW **
* without op. amplifier ** depends on type of crystal/vibmode			

5. PRACTICAL EXPERIMENTS - RESULTS

From the specification shown in § 1., the following oscillators have been designed and compared (scheme 10).

- direct frequency TCXO : 100MHz AT 3rd OT /AT 5th OT
- multiplied Freq. TCXO : 33MHz AT Fund X3 /20 MHz AT Fund X5
  33MHz AT 3rd OT x 3

The frequency multiplication is obtained by using a band pass filter (L/C or Quartz) which selects the 3rd or 5th harmonic of the oscillator signal Fo. The output level is obtained through an amplifier. When using a Quartz filter, the bandwidth at 3 dB is about $\pm$ 5 KHz or more. The phase noise close to the carrier ($1/F^3$, $1/F$) is not modified. The choice between Quartz or L/C Filter depends on the required harmonic/subharmonic levels.



Scheme 10.

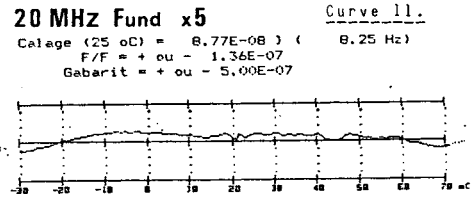
The results are not exhaustive. Otherwise, the oscillators designs are very different regarding the components and adjustments parameters. The interpretation of variances between the results (especially phase noise characteristics) can be complex.

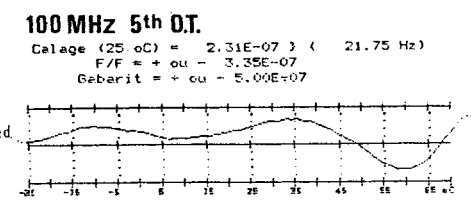
However it is possible to draw some conclusions by considering the different parameters.

5.1. TEMPERATURE COMPENSATION (SEE CURVE 11)

The different oscillators that have been presented reach stabilities of about $\pm 1.10^{-6}$ ($-25 \leq \Theta \leq 75°C$) rather easily. The $\pm 5.10^{-7}$ class needs the use of resonators with very low temperature defects (some 10^{-7}). In this case, the 20 MHz/33 MHz At Fund. crystals give the most constant results. In the frequency range 100 MHz, stability of $\pm 5.10^{-7}$ represents the limit of the analogic compensated oscillator (standard production).

Curve 11.





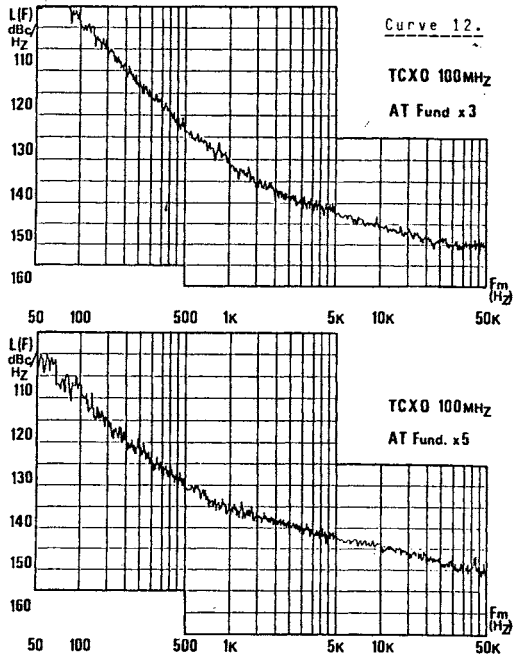
5.2. PHASE NOISE

- 20 MHz Fund X 5 / 33 MHz Fund X 3 TCXOs (Curve 12)

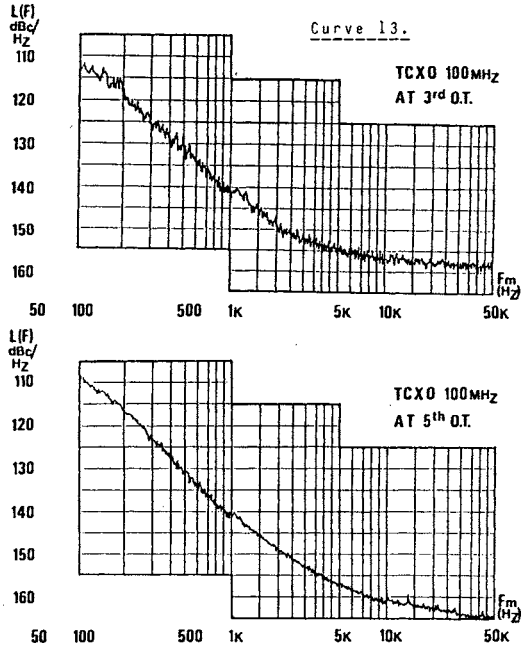
The phase noise at 100 MHz corresponds with the 20 MHz phase noise, + 14 dB (resp. 33 MHz + 9 dB).

In both cases, the 1/F3 (Flicker frequency noise) is limited by the Qo Factor of the crystals. The 1/F (Flicker Phase noise) is degraded by the oscillator's noise at 20 or 33 MHz.

Curve 12.



Curve 13.



- 100 MHz 3.OT/AT 5.OT TCXOs (Curve 13)

The higher sensitivity of the AT3 oscillator is compensated by the higher Qo Factor of the AT5 resonators.

The results are rather similar. The 1/F Flicker noise is quite good.

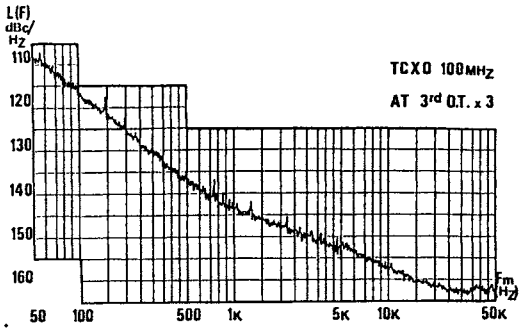
- 33 MHz AT 3.0T X 3 TCXOs. (Curve 14)

The tested oscillators use an op. amplifier in the compensation network (gain=1.8)

This allows the oscillator's sensitivity to be reduced by minimizing the serial inductor.

Phase noise results close to the carrier (1/F3) are very good. The 1/F section comes from the Flicker noise of the 33 MHz oscillator and it limits the performance over 1 KHz - 10 KHz.

Comparing with the OCXOs, the phase noise close to the carrier (Fm ≤ 5 KHz) is limited by the sensitivity required on the analogic TCXOs. In any case, these results cannot be reached with Digital TCXO, because of their principe.



Curve 14.

5.3. LONG TERM STABILITY

The aging depends essentially on the specific crystal characteristics. However, it can be altered by the long term behavior of the series-connected components, especially for the TCXO where the required sensitivity is obtained through the use of inductors. Otherwise, the increase of crystal drive level degrades the aging.

- Although the resonators and the oscillators are quite different, the direct frequency TCXOs (AT 3.0T, AT 5.0T) and the multiplied frequency TCXOs (AT Fund X 3, X 5) give similar aging characteristics:

$$\Delta F/F\ (t) \simeq 1.10^{-6}/\text{year}$$

The reduction of this value seems to be difficult in standard series.

- On the other hand, the AT 3.0T X 3 TCXOs give the best compromise : they guarantee (for industrial series) an excellent long term stability :

$$\Delta F/F\ (t) \leq 5.10^{-7}/\text{year}.$$

5.4. ENVIRONMENTAL CONDITIONS

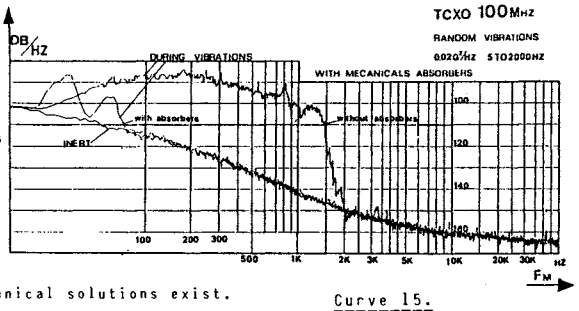
For on board equipments, the initial specification is completed by environmental requirements. Phase noise characteristics are required under mechanical vibrations and acoustic perturbations.

The resonator will be chosen taking into account the following complementary aspects :

- mechanical frequency resonances
- acceleration sensitivity (g sensitivity)
- acoustic sensitivity

In some cases, it could be necessary, in addition, to neutralize these disturbances by using shock absorbers and/or elastomer foams.

An example of Phase noise under mechanical and acoustic strains is given on curve 15.



Curve 15.

6. CONCLUSION

With the specification in question, different technical solutions exist. The following results can be achieved :

Temp. Stability		standard	industrial limit
$-25 \leq \Theta \leq 75°C$		$\pm 1.10^{-6}$	$\pm 5.10^{-7}$
Phase noise (dBc/Hz)	10 Hz	75	80/85
	100 Hz	105	110/115
	1 KHz	135	140/145
	10 KHz	150	155/160

Each solution has its own characteristics and its specific operating domains, regarding others parameters :

AT 3.0T AT 5.0T direct	easy electronic concept very good 1/F3, 1/F	Quartz difficult to produce ($\Delta F/F\ (\Theta)$)
AT 3.0T X 3 multiplied	very good 1/F3 Excellent aging	1/F limited at 30 MHz more complex scheme
AT Fund. X 3 X 5	natural sensitivity . possibility to develop VTCXO (low shift . temp. compensation more easy	more complex scheme

Around 100 MHz these different solutions can be compared.

The multiplication frequency concept allows Low Phase noise TCXOs to be developed in extended frequency range (100 MHz up to 350/450 MHz).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-03 Number 4

ELECTROSTATIC EFFECTS ON QUARTZ CRYSTAL RESONATORS

R.J.Williamson
STC Technology Ltd
London Road, Harlow
Essex CM17 9NA
England

A.M.Harrison
STC Components
Quartz Crystal Division
Edinbugh Way, Harlow
Essex CM20 2DE
England

SUMMARY

Initial experiments into the effects of voltage surges applied to quartz crystal resonators showed that both frequency and ESR shifts occurred and in some circumstances these effects could be reversed.
A more systematic and detailed study has identified a number of effects in quartz crystal resonators from electrostatic applied voltages and three distinct regions dependant on voltage level have been identified.
In the first region, with voltages up to 10^6 V/m, there is no measurable effect on the quartz parameters. In the third region from about 2 kV upwards, electrical breakdown phenomena occur both across the surface and through the bulk of the quartz, clearly causing irreversible damage. In the region between these two voltage levels the effect is erratic yet reversible and apparently permanent.
It is in this region that most work has been concentrated.

INTRODUCTION

For many years the quartz crystal industry has regarded their product as being essentially non-static sensitive. However, particularly with regard to higher frequency fundamental crystals, this assumption has recently increasingly been questioned. There is a considerable amount of published work (Refs. 1-16) on voltage/frequency effects in relatively low fields, up to 10^6 V/m studying the first derivative of the elastic constants with the applied voltage. This effect is, according to the papers, entirely reversible. Furthermore, in theory, singly rotated cuts such as the AT and BT cuts should have no electrostatic effect and therefore most of the reported work concentrated on double rotations such as SC cut(Refs. 3-16).
Initial studies indicated that above these voltage levels, and sometimes below, crystals exhibit both erratic jumps in frequency and shifts in frequency which are dependent on the direction of the applied voltage in relation to the crystallographic axes.
The purpose of the current work was to investigate and characterise these effects with a view to discovering the mechanisms involved and to establish their degree of permanence. With a better understanding of these mechanisms it is hoped that it may be possible to develop a test method to detect crystals which have been damaged by static.

TRIALS

Trials were carried out using two different sets of apparatus to apply voltage to crystals. As it is expected that contact between a charged person and a crystal would be a major cause of electrostatic damage, the first series of trials were conducted using the equipment as illustrated in Figure 1. This equipment simulates the human body as a high voltage source. There is some confusion as to the most appropriate model to use. BS6667 Part 2 and IEC 801/2 (Ref. 17) recommend the use of a 150 pF storage capacitor and a 150 ohm discharge resistor, while DOD-STD 1686 (Ref. 19) and ESA Specification 238000 (Ref. 18) use 100 pF with 1.5 kohm. More complicated modelling is also recommended (Ref. 20) which includes inductances in the circuit. For these trials, however, the BS model was adopted throughout.
A second series of experiments was carried out using the equipment illustrated in Figure 2. This equipment enabled the frequency to be monitored while the field was applied. It was possible to use this apparatus over part of the voltage range, at lower voltage levels where the mechanisms do not appear to be dependent on the rise time of the applied field. Thus a large resistance of 1 Mohm can be placed in series with the voltage source both to reduce the initial current surge and hence to protect the measurement equipment, also to eliminate the effect of voltage source on the crystal measurement. Both sets of equipment were such that the voltage could be switched to either side of the device under test.
Trials using the "human body" apparatus were carried out on unelectroded blanks and electroded blanks mounted on standard crystal headers. To detect the frequency response of the unelectroded blanks, an air gap jig was used on which the air gap could be accurately and repeatably set to ensure that the frequency and amplitude changes could be accurately recorded and compared. The trials on the plated unencapsulated crystals were carried out in a vacuum system as shown in Figure 1. The system enabled the atmosphere to be controlled so that ageing due to airborne contamination could be reduced.
A further series of trials were conducted using plated fully encapsulated crystals using the second apparatus.

© 1990 STC PLC

RESULTS

The results show that the effects can be conveniently categorised into three regions of applied voltage level each with its own distinct behaviour.

Region I

This region for fields up to 10^6 V/m there is no observed remnant effect. For singly rotated cuts, there is no frequency shift even where the field is maintained. For doubly rotated cuts,the effect is claimed to be closely linear with field ,although a quadratic component has been detected(Ref. 1). It is generally agreed that the frequency shift is caused by the change in the first derivative of the elastic constant with field. These effects were not studied and all experiments were carried out on AT cut crystals.

Region II

In this region where higher voltages above 10^6 V/m were applied permanent frequency shifts occur. A number of swept and unswept crystals at about 75 MHz with differing amounts of plateback, have been subjected to a variety of electric fields.
Figure 3 shows a plated crystal where a voltage (1.5 kV) has successively been applied, removed and then reversed with the indicated effects. Figure 4 shows the same experiment performed at -8°C.
There are occasional jumps in frequency in both directions, and
the jumps are worse (more frequent and larger) at lower temperatures and with unswept devices.The frequency shifts are very fast (milliseconds) and generally stable but there is no progressive additive frequency shift.
In order to throw more light on the underlying mechanisms, crystals were subjected to a ramp voltage with and without a zero field relaxation period, so that the onset of the phenomena could be observed.
Figure 5 shows the frequency changes with applied voltage for an unswept 73 MHz crystal. Each measurement was taken with the field applied. Each point on the graph is largely repeatable during the reapplication of a particular voltage and the voltage is varied in discrete steps to complete a cycle of measurements.
Apart from some discontinuous frequency jumps, the general shape of the resulting curve reveals a predominant quadratic characteristic, with a small linear form and considerable hysteresis in both directions.
Figures 6 and 7 show the same curve for a swept crystal at room temperature and at 125°C. The remnant frequency shift when the voltage is removed is also shown. When the same voltages were re-applied small changes in the frequency shift occurred.
The plots of Figures 6 and 7 show little difference due to the effect of temperature.
Figure 8 shows the general shape of the frequency shift, with the change in frequency whilst the voltage is applied always negative and generally quadratic in nature with the remnant (zero field) frequency shift being predominantly linear.

Region III

This region is much more straight forward. Two destructive breakdown mechanisms are observed both having a permanent and progressive frequency shift and a worsening of the ESR.
Figure 9 shows the result of applying +/- 4 kV using the "human body" apparatus on the frequency of a 132 MHz blank. A number of effects can be seen. Firstly, the multiple peaks from coupling to plate modes; secondly the resultant frequency has moved either side of the initial frequency on reversing the field; thirdly, from the repetitive sweeps there is evidently some relaxation after the application of the voltage and finally, the crystal broke down leading to a low Q response. The field seen by this crystal was in the in the region of 3.2×10^8 V/m. Figure 10 shows the damage caused by breakthrough on this blank. Figure 11 shows the effect on an electroded device subjected to a similar voltage with a piece of quartz ejected together with twinning. The field of about 3×10^8 V/m is about half that quoted by Cady (Ref. 21) for the breakdown of quartz.
The second destructive mechanism which can be observed is surface breakdown in which flashover occurs across the surface of the quartz, often forming a corona around the periphery of the electrode. The damage thus caused is illustrated in figure 12.
The cleanliness and morphology of the surface as well as the rise time of the applied voltage alter the point of breakdown.For example, for a particular design at 21 MHz, the breakdown occurred at 800 Volts for a lapped finish but at over 1.2 kV for a polished finish. Careful cleaning also raises the breakdown voltage as does a slow voltage rise time.

POSSIBLE MECHANISMS

The mechanisms responsible for the Region III effects are those of dielectric and surface breakdown which are fairly well understood. The damage to the crystal is clearly irreversible.
It is probable that dislocations or etch tunnels are the site of breakdown in the bulk of the quartz as subsequent voltage stressing of the device reveals a relaxation of the frequency shift with a time constant of several seconds (Fig. 13). This is compatible with the effects noticed by others (Refs. 8,13,15) which is attributed to the mobility of ions within the quartz. It is possible that this increased mobility is caused by surface diffusion down the hole created by the breakdown.

The surface breakdown mechanisms are dependent on the rise time of the applied voltage and it is possible that fast rise times could drive the crystal into resonance. This could produce very high stresses which would in turn produce high piezoelectric voltages across the crystal just outside the electrodes. These voltages may assist in the surface breakdown and this is a possible explanation of the way in which the damage to the electrode often develops around its periphery (Fig 12). Clean polished crystals breakdown at higher voltages because they do not provide any surface peaks to provide shorter paths for flashover.
As far as Region II is concerned it is possible that the frequency shifts are caused by mechanisms similar to those involved when a crystal is irradiated.
The observed frequency shifts, as a result of subjecting a crystal to a sufficiently large field, could be caused by the movement within its site of one of the defects present in the quartz with a resultant alteration of oxygen bond lengths and, hence, elastic modulus. There may be more than one effect involved to explain the results since there is both a relatively smooth and repeatable variation together with some random jumps. The latter which are worse for both unswept quartz and at low temperatures are most likely to be associated with the change of a macroscopic defect since it seems unlikely that isolated point defects would act in a concerted but random fashion as observed in the jumps. The most obvious candidate for such a defect would be dislocations. The more repeatable effect which is present in all devices may be associated with the movement of one of the other impurities trapped within the lattice such as aluminium, iron, water or hydrogen.

CONCLUSIONS

It has been established that both parametric and catastrophic damage occurs to crystals subjected to electrostatic voltages.
Figure 14 illustrates the frequency variation caused by the application of positive and negative voltages on 21 MHz crystals. Region I is unaffected, Region II shows progressive frequency shift and Region III shows the catastrophic breakdown.
Two catastrophic damage modes (Region III) have been identified, a field-related dielectric breakdown and a voltage-related surface flashover. During surface breakdown the electrodes are damaged and there is evidence that the noise performance of the device is degraded. This effect requires further investigation.
For dielectric breakdown on electroded devices, there is short term relaxation during and after voltage stress, possibly caused by impurity movement along the damage.Both these effects are apparent on subsequent measurement of the crystal and would cause the crystal to be rejected.
The parametric changes (Region II) are of uncertain origin.There seem to be two mechanisms; erratic frequency jumps worse in unswept material and a remnant shift repeatable in direction if not magnitude, dependent on the direction of the applied field with respect to the crystallographic axes. It could be that both of these mechanisms are caused by impurities and defects in the quartz being rearranged, the erratic jumps being associated with dislocations. Neither of these mechanisms leads to a progressive shift in frequency on successive application of the same voltage.Having established that there are changes to the crystal it is important to be able to assess the electrostatic history of a device.
For these mechanism it is not clear whether there will be a means of monitoring the crystal condition and what the long term implications are.Is there a change in the noise spectrum of the device; is the long or short term ageing affected?
These areas still remain to be investigated.

ACKNOWLEDGEMENTS

This work was largely carried out at STC Technology Ltd, funded in part by the Procurement Executive, Ministry of Defence (Royal Signals and Radar Establishment), sponsored by Admiralty Research Establishment (Portsdown), and in part by STC Components.

REFERENCES

1) Hruska, C.K. IEEE Trans. S & U Vol SU 28 No. 2, 1981, p.108-110

2) Hruska, C.K. IEEE Trans. S & U Vol SU 27 No. 4, 1980, p.221-223

3) Kusters, J.A. Proc. 24th AFCS 1970 p.46-54

4) Hruska, C.K. Proc. 31st AFCS 1977 p.159-170

5) Brendel, R. and Gagnepain, J.J. Proc. 36th AFCS 1982 p.97-107

6) Brendel, R. J. Appl. Phys. 15(9) 1983 p.5339-5346

7) Hruska, C.K. J. Appl. Phys. 61(3) 1987 p.1127-1129

8) Brendel, R., Gagnepain, J.J., Aubry, J.P., Proc. 40th AFCS 1986 p.121-126

9) Kittinger, E., Tichy, J., Freidel, W. J.Appl. Phys. 60(4) 1986 p.1465-1471

10) Hruska, C.K. Proc. 40th AFCS 1986 p.54-62

11) Kittinger, E. and Tichy, J., 1st EFTF 1987 p.222-225

12) Hruska, C.K. and Brendel, R. ibid. p215-221

13) Brendel, R., ibid p15-158

14) Hruska, C.K. and Brendel, R. J. Appl. Phys. 65(2) 1989 p.715-717

15) Brendel, R. 3rd EFTF 1989

16) Hruska, C.K. IEEE Trans U, F and FC Vol 35(5) 1988 p.637-640

17) British Standards Institution : Methods of Evaluating Susceptibility to Electric Discharge. Bs 6667 part 2 1985 and IEC 801-2 1984

18) European Space Agency : Electrostatic Discharge Sensitivity Test Method. Spec. No. 23800, July 1989.

19) Department of Defense : Electrostatic Discharge Control Program for Protection of Electrical and Electronic Parts, Assemblies and Equipment. DoD-Std. 1686, May 1980

20) Boxleitnor, W. Electrostatic Discharge and Electronic Equipment. IEEE Press 1989

21) Cady, W. Piezoelectricity. McGraw-Hill 1946

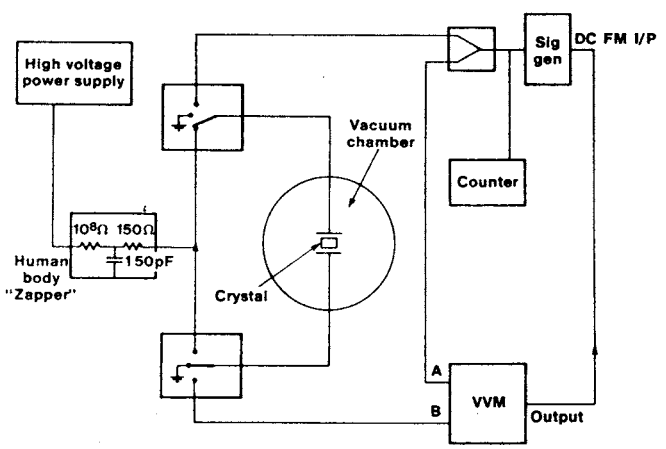


Fig. 1 ' Human- Body ' Test Equipment

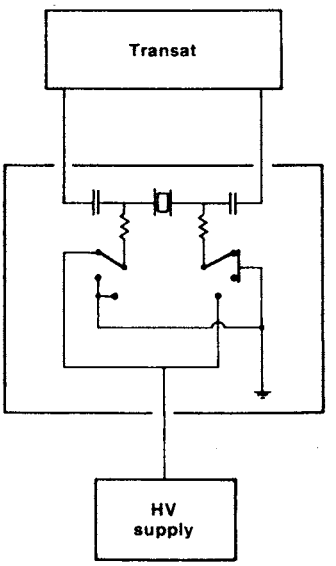


Fig. 2 Applied Voltage Test Equipment

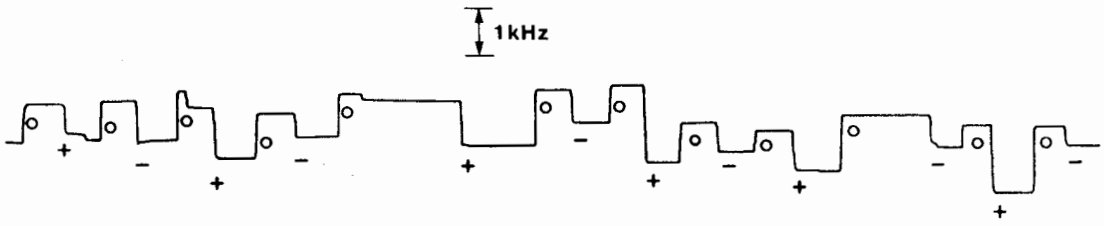


Fig. 3 Repeated application of voltage (+/- 1.5 kV) unswept 73 MHz device at room temperature

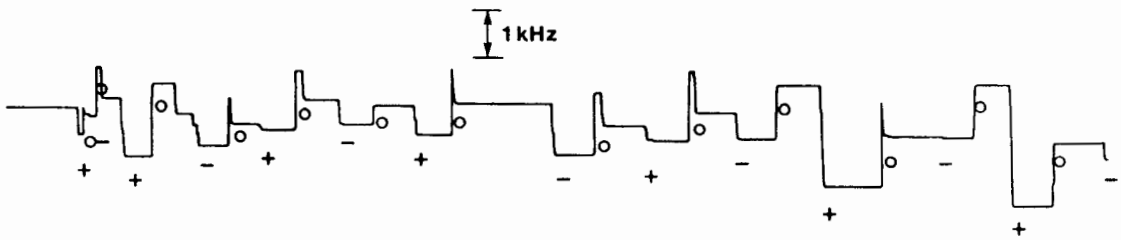


Fig. 4 Repeated application of voltage (+/- 1.5 kV) unswept 73 MHz device at -8°C

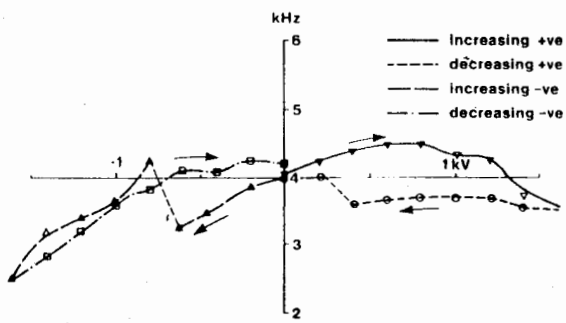


Fig. 5 Plot of frequency change with applied voltage. 73 MHz unswept crystal at room temperature

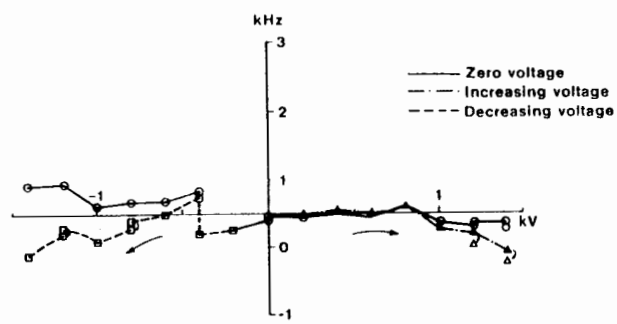


Fig. 6 Plot of frequency change with and after applied voltage. 73 MHz swept crystal at room temperature

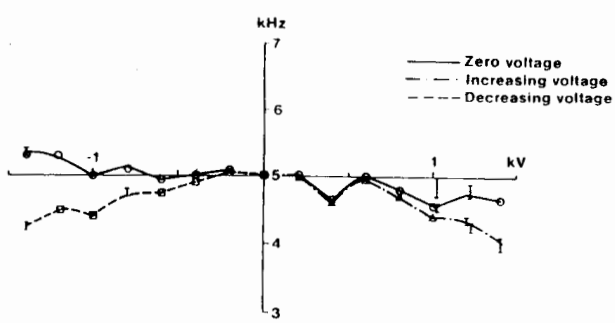


Fig. 7 Plot of frequency change with and after applied voltage. 73 MHz swept crystal at 120°C

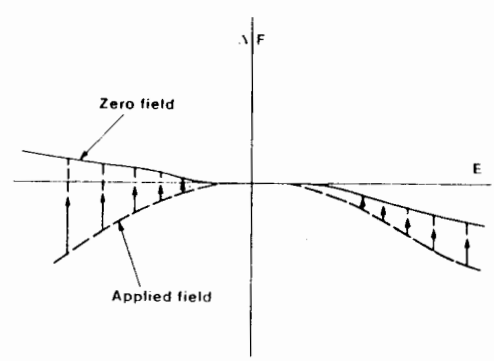


Fig. 8 General shape of voltage/frequency shift

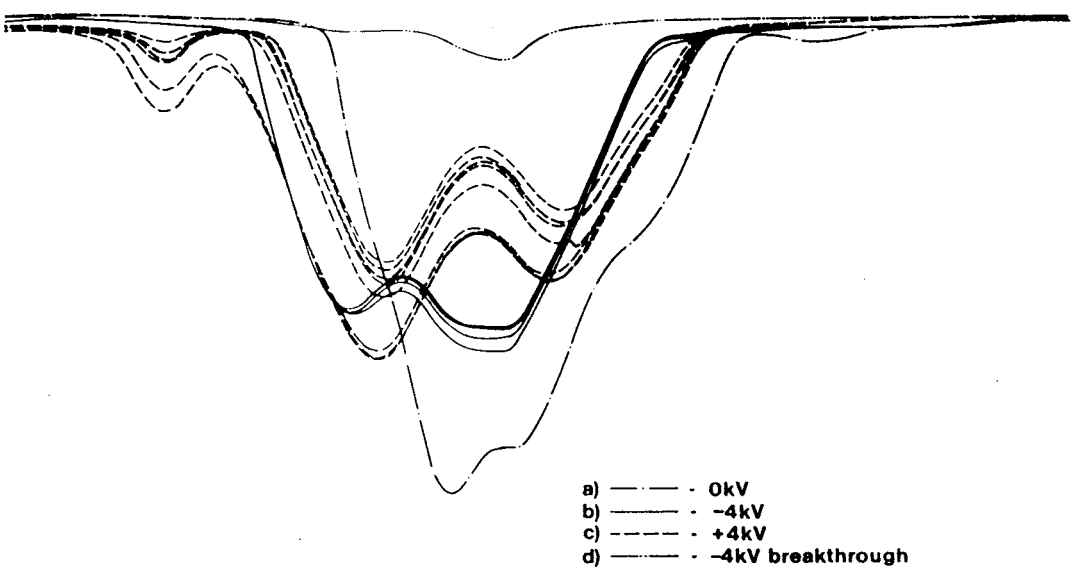


Fig. 9 Air gap response of 132 MHz fundamental (frequency span +/- 100 kHz)

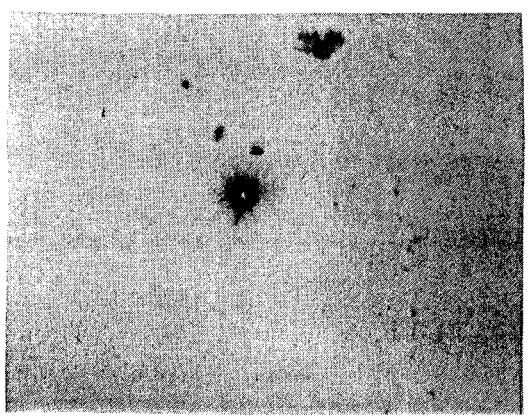


Fig. 10 Photomicrograph of a quartz blank damaged by breakthrough under 3 x 10^8 V/m (x 200)

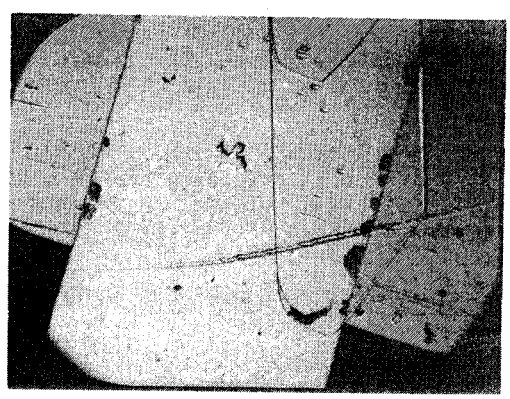


Fig. 11 An unelectroded blank showing a piece of quartz ejected and evidence of twinning (x 70)

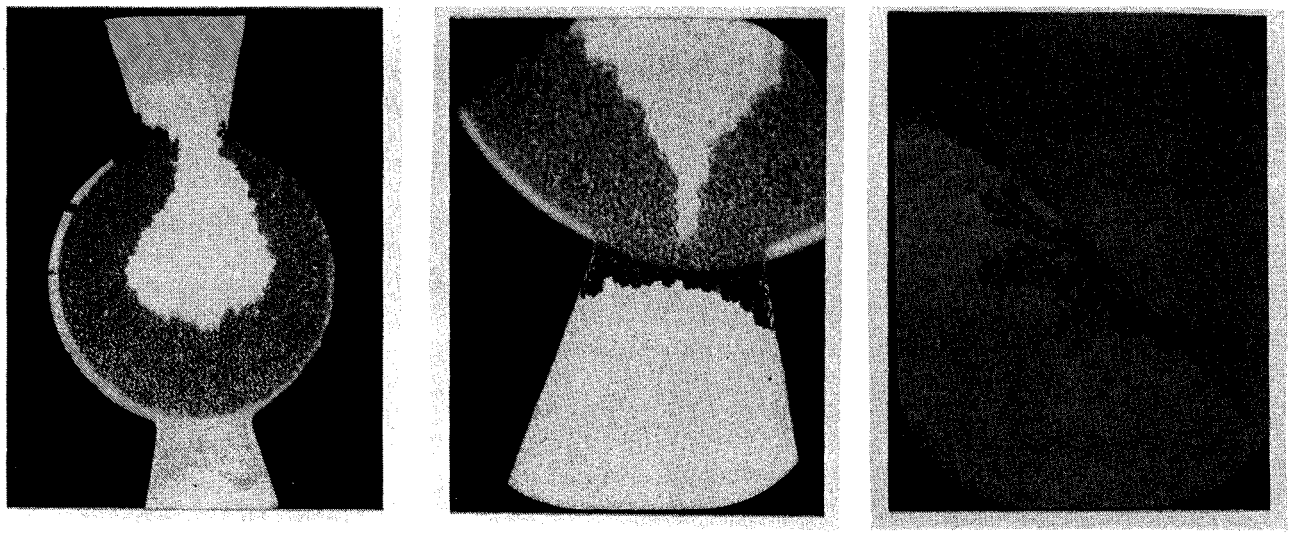


Fig. 12 Damage to an electroded device caused by flashover

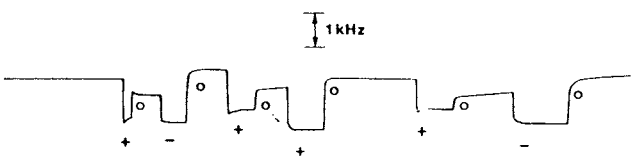


Fig. 13 Frequency deviation with field after breakdown

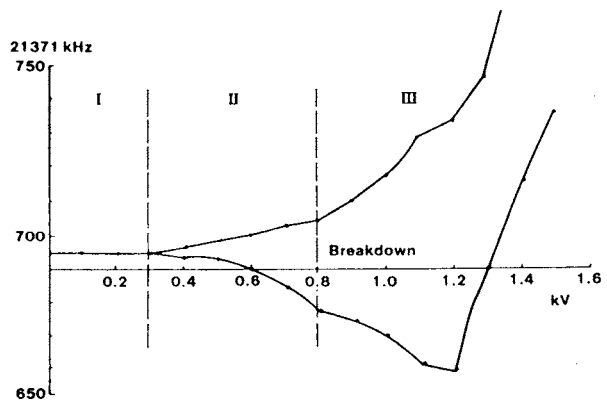


Fig. 14 Variation of frequency of 21 MHz crystal with positive and negative applied voltages

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-03 Number 5

AMPLITUDE AND PHASE INSTABILITIES INDUCED BY NONLINEARITY AND SPATIAL DISPERSION IN A QUARTZ CRYSTAL RESONATOR

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

It is well known that low frequency modulations with 1/f spectrum are the ultimate limit in the stability of quartz crystal resonators. Despite numerous works, 1/f noise has not been well understood in most solid state physics systems [1, 2]. In an attemps to improve our knowledge of these fluctuations experimental [3] and theoretical work [4] were reported.

Since the wave function for a dispersive wave decreases in time like $t^{-1/2}$ in the stationary phase approximation (i.e. large times or low frequencies), its power spectrum will have the hyperbolic shape around the carrier frequency [4]. In this paper we show that dispersion in quartz originates in the polarisation microstructure.

Mindlin's equations of continuum elasto-optics [12] are used to derive the dispersion relation for thickness shear vibrations in a Y-cut quartz resonator.

Accounting for both dispersive and nonlinear effects, the Korteweg de Vries equation for only phase modulations is established. Periodic solutions are computed and it is shown that the period and the wave shape change like 1/f at low frequencies.

1 - Introduction : Thickness shear vibrations in the linear nondispersive approximation

Thickness shear vibrations in a rotated Y-cut quartz plate excited by a shearing force traction are conventionaly written by using the linear theory of piezoelectricity in the quasistatic approximation. The shear mechanical displacement u_1 and the electrical potential ϕ obeys the following set of equations [5]

$$\begin{aligned} C_{66}\, u_{1,22} + e_{26}\, \Phi_{,22} &= \rho\, \ddot{u}_1 \\ e_{26}\, u_{1,22} - \varepsilon_{22}\, \Phi_{,22} &= 0 \end{aligned} \tag{1}$$

and when $y = \pm h/2$

$$\begin{aligned} C_{66}\, u_{1,2} + e_{26}\, \Phi_2 &= 0 \\ \phi &= \pm \phi_o / 2 \cos \omega t \end{aligned} \tag{2}$$

From the equation for the electrical displacement

$$D_2 = e_{26}\, u_{1,2} - \varepsilon_{22}\, \Phi_{,2} = -Q/A$$

where $Q = C\, \phi_0 \cos \omega t$ is the charge on the electrode, the effective dielectric constant $\bar{\varepsilon}$ and the electrical current i can be expressed with the relations $C = \bar{\varepsilon}\, A/h$ and $i = A\, \dot{D}_2(h)$.

Solving system (1)-(2) it is found [5, 6]

$$\bar{\varepsilon}\,(\omega) = \varepsilon_{22} \Big/ \left| 1 - K^2\, tg\left(\frac{h\omega}{2V_o}\right) \Big/ (h\omega/2V_o) \right| \tag{3}$$

where $\bar{K}^2 = e_{26}^2 / \bar{C}_{66}\, \varepsilon_{22}$ is the electromechanical coupling constant and $\bar{C}_{66} = C_{66} + e_{26}^2/\varepsilon_{22}$. Owing to the smallness of $\bar{K}^2$, resonance frequencies are approximately $\omega_n = (2n+1)\, \pi\, V_T / h$, $n = 0, 1, 2, 3 \ldots$

Dissipation effects are usually introduced by a term $r_{66}\,\dot{u}_{1,22}$ in the right hand side of eq. (1b), leading to the replacement of ω by $\omega + j\omega/2Q$ in the expression for $\bar{\varepsilon}$, where $Q = \rho\,\bar{C}_{66}\,h / \pi\, r_{66}$ is the quality factor of the resonator. From the expression for $\bar{\varepsilon}$, the equivalent electrical circuit near resonance is established in the conventional form.

II - Introduction of spatial dispersion

In the classical theory of piezoelectricity no coupling between acoustic and electromagnetic modes is considered. The hypothesis lies upon the big ratio (about 10^5) between the velocity of light and that of sound. Such a coupling was considered by Mindlin [7, 8], leading to the prediction of electromagnetic radiation from a rotated Y-cut quartz plate vibrating in a single thickness-shear mode. The coupling only slightly modifies the frequency of acoustic and electromagnetic modes which does not exhibit any dispersion of their phase velocity.

A different kind of coupling occurs if we account for microscopic effects. Since the electronic polarization is proportional to the relative displacement between the massive ion core and the electron shell and that the long wave expansion involve gradients of displacements, polarisation gradients are involved in the long wave expansion [9]. In centrosymmetric crystals polarisation gradients are responsable for an electric boundary layer effect leading to an anormalous capacitance of thin films [10]. In noncentrosymmetric crystals of the ferroelectric type they explain the existence of a new mode of propagation : the soft electric mode [11, 9]. In addition any crystal lacking of a center of symmetry exhibits optical and acoustical activity. Like in the corresponding optical case, it has been shown that the degeneracy is removed for transverse waves propagated along the acoustic axes, so that instead of one degenerate wave we have two normal circular waves with opposite directions of rotation of the polarisation vector. The simplest consequence is that the plane of polarization of the shear wave rotates as it propagates. This was accounted for in the case of shear waves propagating along the Z-axis [12] in accordance with Brillouin scattering experiments [13].

In the general case, spatial dispersion may couple all acoustic waves travelling in a specified direction leading to elliptical polarization of the waves. However since the eccentricity of ellipses is small, waves remains practically linearly polarized and they experience a dispersion of their phase velocity [14]. For waves propagating in a singly rotated Y-cut quartz resonator we can restrict to the coupling between mechanical transverse and longitudinal displacements u_1, u_2, electromagnetic fields E_1, E_2 and polarizations P_1, P_2. Using Mindlin's theory [12], and since u_i, P_i and Ei are functions of x_2 and t only we obtain

$$
\begin{aligned}
&C_{66}\,u_{1,22} - f_{11}\,P_{2,2} + d_{66}\,P_{1,22} = \rho\,\ddot{u}_1 \\
&C_{11}\,u_{2,22} - f_{11}\,P_{1,2} + d_{11}\,P_{2,22} = \rho\,\ddot{u}_2 \\
&d_{66}\,u_{1,22} + b_{66}\,P_{1,22} + f_{11}\,u_{2,2} - a_{11}\,P_1 + E_1 = 0 \\
&d_{11}\,u_{2,2} + b_{11}\,P_{2,22} + f_{11}\,u_{1,2} - a_{11}\,P_2 + E_2 = 0 \\
&E_{1,22} = \varepsilon_o\,\mu_o\,\ddot{E}_1 + \mu_o\,\ddot{P}_1 \\
&\varepsilon_o\,\mu_o\,\ddot{E}_2 + \mu_o\,\ddot{P}_2 = 0
\end{aligned}
\qquad (4)
$$

In system (4), C_{66}, C_{11}, $a_{11} = \varepsilon_{11}^{-1} = \varepsilon_{22}^{-1}$ and $f_{11} = -\,e_{11} / \varepsilon_{11} = e_{26} / \varepsilon_{22}$ belongs to the classical theory of elasticity so that they are only four new constants d_{11}, d_{66}, b_{11} and b_{66} involved when microstructure is taken into account in the long wavelength approximation.

Looking for waves of elliptical polarization [12] and since we are interested in the propagation of acoustic waves of very low velocity comparatively to the velocity of light, simple vectorial algebra leads to the following dispersion relation

$$\begin{vmatrix} \rho\omega^2 - C_{66}k^2 & 0 & -d_{66}k^2 & f_{11}k \\ 0 & \rho\omega^2 - C_{11}k^2 & -f_{11}k & -d_{11}k^2 \\ -d_{66}k^4 & -f_{11}k^3 & -(a_{11} + b_{66}k^2)k^2 + \mu_o\omega^2 & 0 \\ f_{11}k & -d_{11}k^2 & 0 & -(\varepsilon_o^{-1} + \varepsilon_{11}^{-1} + b_{11}k^2) \end{vmatrix} = 0 \tag{5}$$

An estimate of constants appearing in the dispersion relation can be given. The scale of the spatial inhomogeneity is that given by the dimension a of the unit cell. In quartz, the distance between Si and O atoms is about half the periodicity of the lattice, then

$$\frac{d_{66}}{f_{11}} \simeq \frac{d_{11}}{f_{11}} \simeq \frac{f_{11}}{a_{11}} \simeq \frac{b_{66}}{f_{11}} \simeq \frac{b_{11}}{f_{11}} \simeq a \simeq 2.5 \times 10^{-10} m \tag{6}$$

where $d_{66} \simeq d_{11} \simeq 0.25$ volts ; $f_{11} \simeq 6$ m^2/F and $b_{11} \simeq b_{66} \simeq 15 \times 10^{-10}$ m^3/F.

An order of magnitude of dispersion relation has been obtained by neglecting the influence of piezoelectricity ($f_{11} \simeq 0$). In this case the thickness shear mode of velocity V_T couples only to the electromagnetic mode of velocity V_e and the dispersion relation for both modes is

$$\omega = k \left[\frac{V_e}{V_T} \right] \left| 1 + \frac{1}{2} \varepsilon_{22} k^2 \left(b_{66} \pm \frac{d_{66}^2}{\rho V_e^2} \right) \right| \tag{7}$$

Using the estimates given in (6), we see that the term in d_{66} is about 10^{13} smaller than the first one. The contribution of the first one is about 2×10^{-12}. On the other hand, if we neglect the constants d_{11} and d_{66} and we take full account of piezoelectricity ($f_{11} \neq 0$), the thickness shear may couple only to the b_{11} constant and this leads to a correction of order 10^{-16}. We see from eqs. (4) that the piezoelectric constant f_{11} ensures the coupling between shear and longitudinal modes, leading to elliptic polarization of both modes. However according to the discussion above, the main dispersion arises from the coupling with the electromagnetic mode through polarization gradients and the coupling with the longitudinal mode can be neglected.

III - Nonlinear effects

Nonlinear propagation effects in quartz resonators have been studied for almost three decades [15, 16, 17]. Much of this work is based on early fundamental work on nonlinear elastic wave propagation [18].

If the nonlinear interaction with the longitudinal mode is neglected, the dispersion relation for the thickness shear mode in an untrapped resonator can be written as [15 - 17]

$$\omega = k V_T \left| 1 + \frac{9}{32} \frac{\gamma}{C_{66}} k^2 |\Psi|^2 \right| \tag{8}$$

where $\gamma = 1/2\, C_{22} + C_{266} + 1/6\, C_{6666}$. Since $\gamma = 10^{12}$ N/m^2, the order of magnitude of the correcting term is 5×10^{-12} at $|\Psi| = 1$ Å.

IV - Phase modulations due to nonlinear dispersion

From (7) and (8), we can write the nonlinear dispersion relation for the thickness shear wave as

$$\omega = k V_T \left| 1 + a k^2 + \nu k^2 |\Psi|^2 \right| \tag{9}$$

where $a = 1/2\, \varepsilon_{22} b_{66}$ and $\nu = 9\gamma / 32\, C_{66}$.
Note that the two effects have the same order of magnitude when $|\Psi| \simeq 1$ Å.

The envelope equations for a quasimonochromatic wave propagating inside the resonator are established by rewritting (9) approximately as

$$k = k_o + \chi$$
$$\omega = \omega_o + \chi \omega'_o + \frac{1}{2} \chi^2 \omega''_o + q\,[|\Psi|^2 - \Psi_o^2] \tag{10}$$

with

$$\omega = \omega(k_o)\ ,\ \omega'_o = \left(\frac{\partial \omega}{\partial k}\right)_{k_o, \Psi_o}\ ,\ \omega''_o = \left(\frac{\partial^2 \omega}{\partial k^2}\right)_{k_o, \Psi_o} \text{ and } q = \frac{\partial \omega}{\partial |\Psi|^2}$$

The mechanical displacement becomes

$$u(y,t) = exp\,j(k_o y - \omega_o t)\,\Psi_I(y,t) + exp\,j(-k_o y - \omega_o t)\,\Psi_I(-y,t) \tag{11}$$

where the two terms correspond respectively to the incident and reflected waves inside the resonator and the complex envelope Ψ_I is written in the moving frame $\bar{y} = y - \omega'_o t$ as

$$j \frac{\partial \Psi_I}{\partial t} + \frac{1}{2} \omega''_o \frac{\partial^2 \Psi_I}{\partial y^2} - q\,(|\Psi_I|^2 - \Psi_o^2)\,\Psi_I = 0 \tag{12}$$

Once the envelope Ψ_I for the incident wave is found from the nonlinear Schrödinger equation (12), the solution for the resonator is found from (11).

In the following we consider the case when $q\omega''_o > 0$, i.e. $b_{66} > 0$ so that modulations are stable. Amplitude and phase modulations may decouple in this case. Looking for solutions of (12) in the form

$$\Psi_I = \sqrt{\rho(y,t)}\ exp\,[j\theta(y,t)] \tag{13}$$

and using Karpmann's approximation, $\partial^2/\partial_y{}^2\,(\rho^{1/2}) / \rho^{1/2} \simeq \partial^2\rho/\partial_y{}^2 / 2\rho_o$ it can be shown [19] that amplitude and phase modulations decouple as

$$\rho_t + (\rho\,\theta_y)_y = 0$$
$$v_\tau + \left(\chi_o + \frac{3}{2} v\right) v_y + \beta\, v_{yyy} = 0 \tag{14}$$

where $v = \theta_y$, $\tau = \omega''_o t$, $\beta = -1/(8\chi_o)$ and $\chi_o{}^2 = q\rho_o/\omega''_o$.

The gradient v of phase modulations obeys a Kort eweg de Vries equation and as soon as (14b) is solved the amplitude modulation is obtained from (14a).

Following Witham [20], periodic solutions of equation (14b) are found to be

$$\frac{\partial \theta}{\partial y} = v = -a_1\, Cn^2\,(\Omega t - Ky) \tag{15}$$

with the dispersion relation

$$\Omega = \omega''_o\, K \left[\chi_o + \frac{K^2}{2\chi_o}\,(1 - 2s^2)\right] \quad , \quad 0 < s < 1 \tag{16}$$

In (15), (16) Cn is the Jacobian elliptic function with the modulus s.

An estimate of that modulus s representing the distortion of the wave shape can be obtained at low modulation frequencies Ω (or low wavenumbers K). Since by integrating (15), the amplitude of phase modulations is given by $\theta_1 = s^2 K/\chi_o$ and that at low values of K and Ω (when $\Omega < q\rho_o$), we have $\Omega = \omega''_o\,\chi_o\,K$. We can relate the modulus s to the amplitude frequency effect $q\rho_o$, the modulation frequency Ω and to θ_1 with the simple relation

$$s^2 = \frac{q\rho_o}{\Omega}\theta_1 \qquad (17)$$

For $\rho_0 = 100$ Å ($q\rho_0 \simeq 1.6$ rad.Hz), $\theta_1 = 1° \simeq 1.7 \times 10^{-2}$, $\Omega = 0.1$ rad.Hz this give the estimate $s = 0.5$. Formula (17) means that the wave profile looses its sinusoïdal shape ($Cn \to \cos$ when $s \to 0$) and tend to a cnoidal waveform at decreasing modulation frequencies. Remember that $Cn \to$ Ch when $s \to 1$ which is a localized solution called fundamental soliton.

With decreasing modulation frequencies the wave tends to the cnoidal shape with longer and longer periods given by $T \simeq 2\,K(s)/\Omega$, where K(s) is the complete integral of first kind with modulus s.

In a future paper, we will examine the role of dissipation in destabilising these long period periodic orbits. In our present understanding a $1/\Omega$ spectrum of phase modulations should merge.

Bibliography

[1] Dutta P. and Horn P.M., "Low frequency fluctuations in solids : 1/f noise", Rev. Mod. Phys., 53, 497 (1981).

[2] Weissman M.B., "1/f noise and other slow, nonexponential kinetics in condensed matter", Rev.Mod. Phys. 60, 537 (1988).

[3] Planat M., Gagnepain J.J., "1/f noise in quartz crystal resonators in relation with acoustic losses and frequency dispersion", Appl. Phys. Lett. 50, 510 (1987).

[4] Planat M., "Nature of 1/f phase noise in quartz crystal resonators", IEEE 1988 Ultrasonics Symposium, Chicago, p. 379.

[5] Tiersten H.F., "Thickness vibrations in piezoelectric plates", J. Acoust. Soc. Am. 35, 53 (1963).

[6] Nelson D.F., Electric, optic and acoustic interactions in dielectrics. J. Wiley, New York (1979, Sec. 11.7).

[7] Mindlin R.D., "Electromagnetic radiation from a vibrating quartz plate", Int. J. Solids Structures, 9, 697 (1973).

[8] Ballato A., "Mindlin's elastoelectrodynamics problem", 42nd Ann. Freq. Cont. Symp., 1988, p. 6.

[9] Maugin G.A., Continuum mechanics of electromagnetic solids, North Holland (1988).

[10] Mindlin R.D., "Continuum and lattice theoreis of influence of electromechanical coupling on capacitance of thin, dielectric films", Int. J. Solids Structures, 5, 1197 (1969).

[11] Anderson P.W. in Fizika Dielectrikov. Ed. G.I. Skamavi, Akad. Nauk. SSSR, Fizikoestrii Inst. Lebedeva, Moscow (1960).

[12] Mindlin R.D. and Toupin R.A., "Acoustical and optical activity in alpha quartz", Int. J. Solids Structures, 7, 1219 (1971).

[13] Pine A.S., "Direct observation of acoustical activity in α-quartz", Phys. Rev. B, 2, 2049 (1970).

[14] Vuzhva A.D. and Lyamov V.E., "Acoustic activity and other effects due to spatial dispersion in crystals", Sov. Phys. Cryst. 22, 73 (1977).

[15] Gagnepain J.J. and Besson R., "Nonlinear effects in piezoelectric quartz crystals" in "Physical Acoustics" (W.P. Mason and R.N. Thurston, eds.), vol. 11, Academic Press, N.Y., p. 245-288 (1975).

[16] Tiersten H.F., "Analysis of nonlinear resonance in thickness-shear and trapped-energy resonators", J. Acoust. Soc. Am. 59, 866 (1976).

[17] Planat M., Théobald G. and Gagnepain J.J., "Propagation non linéaire d'ondes élastiques dans un solide anisotrope", L'Onde Electrique, 60, 36 (1980).

[18] Thurston R.N., "Waves in Solids", in Handbuch der Physik, vol. VIa.4 (1974).

[19] Karpmann V.I., "Nonlinear waves in dispersive media", Pergamon Press, Oxford (1975), p. 126.

[20] Witham G.B., "Linear and nonlinear waves", John Wiley & Sons (1974), p. 467.

Frequency Standards

FS-05 Cesium Standards

Chairman: P. Thomann

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-05 Number 1

IMPROVING THE LONG TERM STABILITY OF CESIUM BEAM FREQUENCY STANDARDS

Andrea De Marchi

University of Ancona, Ancona, Italy.

Abstract

Reference standards of physical quantities that are based on microscopic intrinsic properties of Nature should in principle show a stability that increases indefinitely with the averaging time if they were not affected by systematic biases. In fact, under the assumption that such properties are not changing with time, a long enough average should be able to attain whatever level of precision in that case.

Unfortunately in the real world all devices have more or less built-in systematic effects, if only those produced by the perturbations introduced in the system to allow the observation of the considered fundamental property. The overall uncertainty in the evaluation that can be made of all systematic biases in a particular standard is called its Accuracy, and is clearly an upper limit to its expected lifetime variations.

Instabilities greater than what expected from time-averaging of existing noise processes can be introduced in the standard by variations in long term of the systematic biases, but can obviously never exceed the Accuracy.

Following this lead, interesting results have been obtained toward the improvement of long term frequency stability in commercial Cs beam frequency standards. Accuracy analysis shows that only few systematic effects stand in the way of the realization of accuracies of the order of 10^{-13} and long term stabilities of the order of 10^{-14}.

As a consequence the CCDS has recommended that laboratories that contribute to the international time scale consider seriously the improvement of the long term frequency stability of their standards.

In this paper a quick review is given of the relevant systematic effects existing in a commercial standard. However the accent is on reviewing the techniques that can be used to improve their accuracy and as a consequence their long term stability and their environmental sensitivity.

Introduction

It is assumed in this paper that fundamental physical properties of Nature are constant with time and location. If this is true, they constitute the ideal basis for the realization of working standards for reference in a measurement system. Among such properties are those which determine the laws governing the energy levels of unperturbed single atoms. This is the reason for the actual definition of the Second.

Possible limits to the stability of Nature's fundamental constants have been suggested to exist, due to variations of the energy density in the universe, however their projected level is much smaller than the observed frequency variations in Cesium beam frequency standards. Therefore, apart from the question whether long term changes in Nature, if they existed, would be observable at all, instabilities measured in these sources must be generated inside the system that should make the frequency of the slave oscillator equal to the hyperfine splitting of the unperturbed Cesium atom.

Noise and systematic effects are present in this system, and prevent it from being perfect. Random noise is the most relevant limitation to stability in the short term, but is averaged out for long integration times due to its random nature. In well built Cesium beam standards shot noise in the beam is the dominant process of this sort, producing white frequency noise which gives a $\tau^{-1/2}$ slope to the $\sigma(\tau)$ plot. Continued integration makes in principle its contribution smaller than any desired level of stability, but at some level the latter starts to be limited by fluctuations in systematic effects, and the $\sigma(\tau)$ plot then flattens out. This level is usually called the "flicker floor", because its contribution resembles that of a flicker process. It must be noted however that its appearance is caused by long term instability of systematic biases, and it can hardly be assumed that they would be necessarily driven by a flicker process. In fact it's easy to imagine that they could be driven by environmental effects, which usually show a random walk behaviour. That the asymptotic slope of $\sigma(\tau)$ be $\tau^{1/2}$, indicating the presence of random walk of frequency, is in fact clear in the most unstable commercial standards, for which the slope change takes place at short averaging times (curve a in Fig.1) so that good statistics with sufficient data can still be done. In the best commercial standards (curve b in Fig.1), or in laboratory clocks, the presence of random walk is much less obvious, and the stability may well be limited by a flicker process. Whether it's just poor statistics that prevents unambiguous identification, or random walk is really not showing in very good commercial clocks, is in most cases difficult to say, since the standard dies too soon. On the other hand also for laboratory clocks it's difficult to make statements about long term behaviour, due to poor ensemble statistics.

We will leave this question unanswered at this time, because our object here is to state that poor commercial clocks, which do show the effect of long term changes in systematic biases with the typical signature of a random walk of frequency process, suggesting that they may be environmentally driven, are most likely suffering from high sensitivity to variations of microwave power, a problem that can be cured in a number of ways. That microwave power variations be the culprit for excessive long term frequency instability in such devices has been recognized for a long time [1], although only in the recent years the systematic effects which transfer power variations into frequency variations have been clearly identified to be pulling by a cavity mistuning and by the effects of neighbouring transitions.

The rest of this paper is devoted to quickly reviewing the systematic biases which could be responsible for long term frequency instabilities, above the level of 10^{-14}, in commercial Cesium beam frequency standards, and to analyzing the strategies that can be adopted to improve the situation.

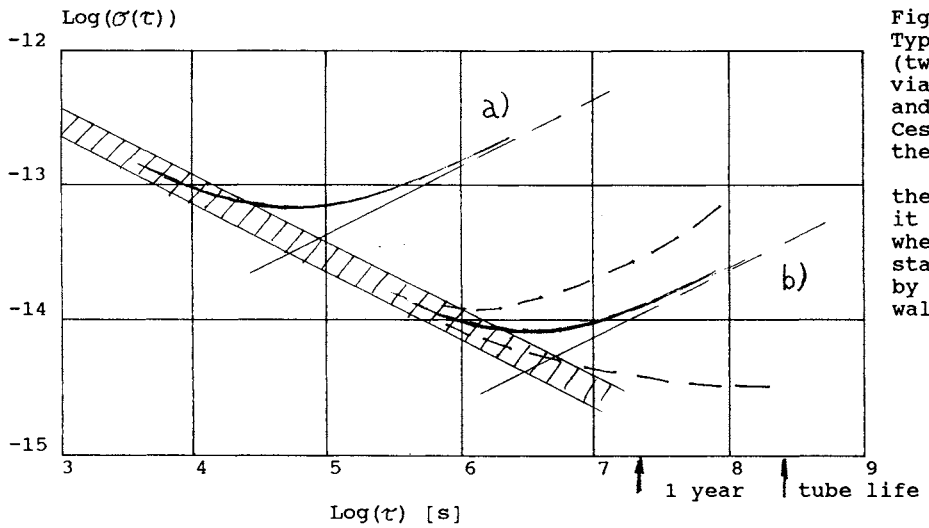


Fig.1
Typical stability plots (two-sample standard deviation) for a poor (a) and good (b) commercial Cesium beam standard of the same model.
It is shown that for the very good standard it is difficult to tell whether the long term stability is dominated by a flicker or a random walk process.

Review of systematic biases in short tubes

From the preceding discussion it appears of primary importance, in order to understand the origin of observed long term instabilities in commercial Cs beam frequency standards, to study in detail the systematic biases that are present in the system and could produce them by fluctuating. This has been the author's main occupation in recent years [2-4] and only the results of this analysis will be summarized here. The main result of interest here is that only very few systematic effects are relevant in short tubes for an accuracy level of 10^{-13} and a long term stability of 10^{-14}. In fact most of the bias effects which are traditionally considered for the evaluation of primary standards, where atomic beams are usually fast (200 m/s) and beam apertures are $\lambda/2$ from the end of the microwave cavity so that they can be made wider and allow the use of bigger beams, become smaller in short tubes, where beams are slow and pass close to the end of the cavity. These include 2^{nd} order Doppler effect and end-to-end and distributed cavity phase shifts, while 1^{st} order Doppler increases from the case of long beams due to the less favorable aspect ratio (Rabi cavity length over Ramsey separation), but still remains very small.

After this screening only three types of pulling are left for consideration, that can be relevant at the mentioned target levels of accuracy and stability. These are the C-field offset, cavity pulling and pulling by neighbouring transitions. To these, biases coming from technical imperfections in the electronics should be added, but they are usually of the order of 10^{-13} at most in a good system, and stable enough not to constitute a worry. For this reason, and because they can hardly be improved by the user, we will not address these problems.

What can be improved by the user, if necessary, are cavity pulling and pulling by neighbouring transitions, which are responsible for the coupling between power and frequency instability.

The latter has been shown to have a damped oscillating behaviour versus the quantization field B_c, so that field values usually exist for which the effect vanishes. Operation of a number of standards with high "flicker floor" at one of these zero crossing points for power sensitivity has resulted in long term stability improvement down to the 10^{-14} level. Two distinct mechanisms appear to contribute to this bias. The effect of the background slope from the asymmetric wings of neighbouring Rabi lines (Rabi pulling) is well known and was shown to be the main one in some cases [5,6]. It vanishes with the use of servo schemes which reject a background slope, like sinewave modulation with detection on the third harmonic [7] or slow square wave frequency modulation with variable depth [6]. The other mechanism is much less clear and is under study. Its existence is denounced by the fact that in some standards the period of the measured oscillations versus C-field is the double of what whould be expected from the Rabi pulling theory, and for the same standards the effect is not dependent on the type of modulation [8]. For these reason it is thought that it may be related to coherent interaction between the m=0 and the m=±1 Zeeman

sublevels. Both mechanisms are power dependent and have zero crossing C-field values, so that they can be treated in the same way, except for the fact that servo schemes which reject a background slope work only for pure Rabi pulling. For example reduction of bias below 10^{-13} and of its power induced variations to the 10^{-14} level can be done by setting the C-field at the zero crossing of power sensitivity.

Cavity pulling is well known [7] to go through zero near the power that maximizes the signal at center frequency (optimum power), the precise point depending mainly on the used modulation scheme. Therefore it can be easily made very small, as a bias effect, if power is close to the level for which it vanishes. This is not necessarily easy to obtain, since optimum power, which is the easy reference level, may be 0.5 dB or more away from this point, but cavity tuning good enough to push the bias below 10^{-13} can be performed with some care. A problem arises here from the fact that a small bias does not automatically mean that the power sensitivity which is induced by this effect be irrelevant. In fact a bias of less than 10^{-13} may easily induce a power sensitivity greater than 2×10^{-13}/dB. In Fig.2 cavity pulling curves, calculated for a real tube and sinewave modulation with first harmonic detection, are reported as a function of microwave power, with the fractional cavity detuning as a parameter. The solid lines indicate positive cavity detuning and the dashed lines indicate negative cavity detuning. It can be observed that the only situation which yields no power sensitivity is perfect cavity tuning. While a bias below 10^{-13} is easy to obtain, power induced long term instabilities above 10^{-14} should not be surprising. Possible improvements on this problem include the use of different servo loop schemes, for example third harmonic detection [7], and compensation of power sensitivity with the residual sensitivity introduced by the effect of the neighbouring transitions. Careful tweaking of both cavity, for best tuning, and C-field, for a zero of the pulling by neighbouring transitions, has been shown to improve the long term stability down to the 10^{-14} level in standards that previously showed a factor of 3 to 10 worse variations [2-4]. This shows that, at least above that level, these two effects are the only source of environmental sensitivity [2] and long term instability [3] in short tubes. The overall accuracy can also be improved to 10^{-13} by this procedure.

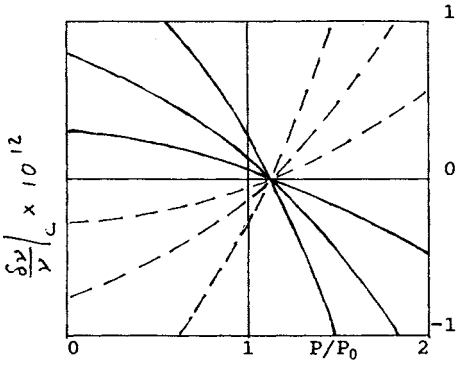


Fig.2 - Cavity pulling versus power with cavity detuning as parameter, in 1st harm. loop.

In Fig.3 (a and b) an example of such an improvement is reported. The stability seen before tweaking with f_z = 53 kHz (Fig.3a) is compared with the stability measured at 39 kHz after tweaking (Fig.3b). Random walk is clearly seen only in the first.

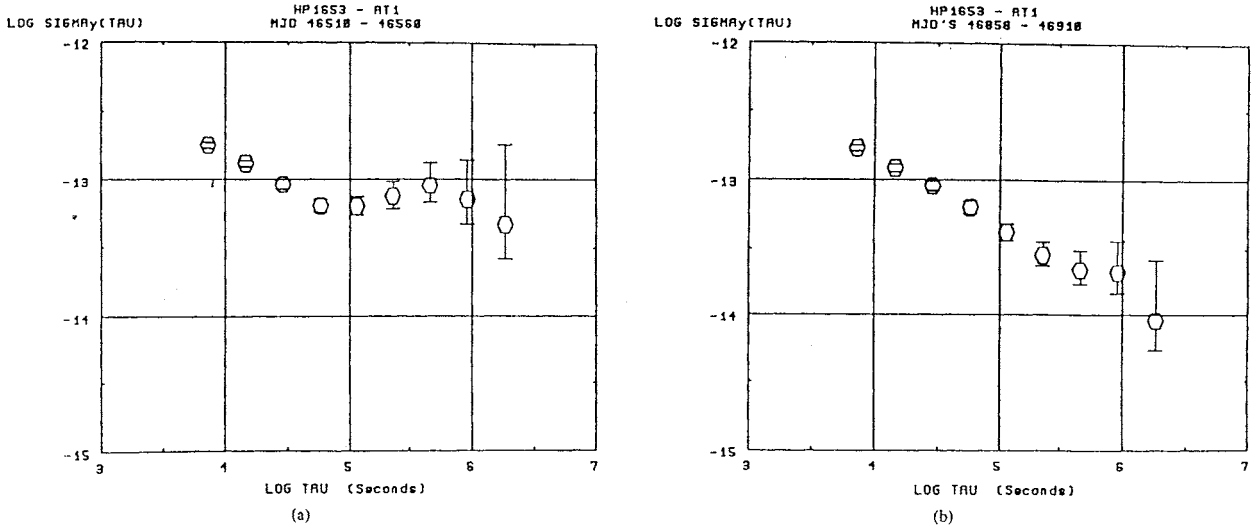


Fig.3 - Measured frequency stability for a high performance Cesium beam standard operated at two different C-field values, which give two different power sensitivities. In Fig.3a the latter is 10^{-12}/dB, and in Fig.3b it is less than 2×10^{-13}/dB. The improvement in long term stability is of a factor of 5 from a) to b), indicating that power sensitivity is the driving effect for the instability in a). In fact the C-field is lowered only a factor of 1.36 from a) to b), and could account only for an improvement of $1.36^2=1.85$ (less than 2) if the driving effect were magnetic field instability.

By far the biggest bias present in the system is the offset introduced by the magnetic field which is necessary to define the quantum states in the experiment. This bias is of the order of 2×10^{-10} for normal operating conditions, and the formula $8(f_z/\nu)^2$ is good to 10^{-14} for its evaluation if the field is reasonably homogeneous (Ramsey patterns must be clearly visible for the $|m|=3$ transitions). The Zeeman frequency f_z used here is the distance at microwave frequency between the m=0 and m=1 central Ramsey peaks. The stability of the field must then be very good if the mentioned target levels are to be attained. In fact the fractional frequency instability induced by field variations is proportional to the fractional field instability $\delta B_c/B_c$, the proportionality constant being $(4f_z/\nu)^2$, that is 4×10^{-10} for f_z around 50 kHz. A quite possible flicker level of 3×10^{-5} in the C-field current supply would introduce a real flicker floor at 10^{-14} for the frequency stability. Random walk and drift of course would do the same. Reduction of the operating C-field value is an obvious way to improve this sensitivity. For example, going from 50 kHz to 25 kHz gives a factor of 4 more rejection of field instabilities, and could be helpful in case the latter were critical. It should be underlined that this argument is in favour of choosing a low field zero crossing of power sensitivity as an optimum C-field setting. Stabilization of the current source based on measurement of the Zeeman frequency would be another way to overcome the problem if better sources were nowhere to be found and operation at relatively high field were desired for any reason. Long term changes in the magnetic shields may also introduce drift or random walk because of the varying shielding factor, and outside compensation of the Earth's magnetic field may be the only solution to this. It should be pointed out that these latter changes induce frequency changes that don't depend on f_z^2 but only on f_z, and their rejection therefore does not increase as fast by lowering the field.

As a conclusion of this discussion it appears that accuracy and long term stability of commercial Cesium beam standards are limited by cavity pulling and pulling by neighbouring transitions, and their power induced variations, at least if they are worse than 10^{-13} (accuracy) and 10^{-14} (long term stability), and that there may be other limitations due to the C-field at levels below that.

A summary of this scenario is given in Table 1, where a list of the main systematics present in a Cesium beam tube is given. The expected size of the bias in a well built commercial tube is indicated. An estimate of the expected uncertainty with which each bias can be evaluated, and of its expected stability, are also given. This estimate should be provided with a complete set of conditions for the environment and the electronics, which ensure that variations don't exceed the indicated figure. This set should include indication of the allowed range for external B field, temperature, humidity and possibly more. It is not our purpose here to specify these ranges, but rather to point out that operation outside of specified ranges may induce variations outside of the indicated limits. In this perspective it can be easily understood how a better accuracy can turn into better long term stability in a specified environment, or relaxation of environmental specifications for a given target frequency stability.

Bias effect		Bias	Bias Uncertainty	Bias Stability	Frequency Uncertainty $\times10^{14}$	Frequency Stability $\times10^{14}$
Quantization B field	*	$<3\times10^{-10}$	10^{-4}	$<3\times10^{-5}$	3	1
Doppler 2nd order		$<1\times10^{-13}$	0.03	<0.01	0.3	0.1
Doppler 1st order		$<3\times10^{-14}$	1	<0.01	3	0.3
Neighbouring Transitions	*	$<1\times10^{-13}$	1	<0.1	10	1
Cavity Pulling	*	$<1\times10^{-13}$	0.3	<0.3	3	1
End-to-end Phase Shift		$<1\times10^{-13}$	0.3	<0.03	3	0.3
Distributed Phase Shift		$<3\times10^{-14}$	1	<0.01	3	0.03
Electronic offsets	*?	$<1\times10^{-13}$	0.3	<0.1	3	1

Table 1
Typical bias and accuracy budget for a well built commercial Cesium beam standard. The second and third columns are uncertainty and stability of the bias effect, and the last two are contribution to accuracy and long term frequency stability. It is assumed that the standard is operated at a zero crossing of power sensitivity with a well tuned cavity.
The estimate of accuracy is thought to be very conservative in this table.
The asterisk "*" indicates that the corresponding effect is relevant for the desired levels of accuracy (10^{-13}), and long term stability (10^{-14}).

Strategies to improve long term stability and environmental sensitivity.

The analysis reported above of possible accuracy and resulting stability for Cesium beam standards of the commercial type (i.e. with a short tube and traditional state selection) carries at least two important consequences.

One is that it's possible, with relatively low effort, to improve their long term stability, when it's not good, so that all standards perform at the level of the best ones. This has caught the attention of the BIPM because of the potential improvement that might induce in the long term stability of the international time scale. It has also interested users who don't wish to build costly and cumbersome environmental stabilization apparatuses around their standards, but would still like to have lower environmental sensitivity.

The other is that it may become possible in the near future to have hundreds of Cesium standards in the world with high reliability and 10^{-13} accuracy. An improvement in the realization of the SI second would then come from the possibility of meaningful statistics, which is at this time a dream since the operating primary standards are very few. This perspective makes the realization of a primary standard economically attractive also for

laboratories and countries who cannot think of undertaking the big project of building a laboratory tube. At the same time it should force the laboratories that are contemplating the idea of starting the big project to aim at an accuracy 10 times better, that is 10^{-14}, than is feasible with a commercial tube. The cost of such a project in fact is around a factor of 100 higher than the cost of a commercial standard.

Strategies which can be followed to improve the long term stability are all centered on the idea of either making the system insensitive to the existence of cavity pulling and pulling by neighbouring transitions, or removing the effects themselves.

A number of the possible strategies are applicable only by the manufacturer at a more or less serious level of redesign. Among these are tube redesign either introducing optical pumping or by the use of computer interactive design of the beam optics for minimization of the pulling, and smart servo loop design to make the system power insensitive. Third harmonic detection is an example of such a servo, for all the cases in which only Rabi pulling and cavity pulling must be taken care of, but more sophisticated digital servos can be fancied. Servo control of the power level is also a possibility. From the user's point of view only two possibilities are really open: one is microwave power stabilization by environmental control of the power source; the other is tweaking of the C-field for minimum power sensitivity, as it was done in the work to which we refer.

All these strategies have their advantages and their disadvantages, but the simplest for the user is without doubt the last one, which was also the first one. In fact its implementation is very simple, as only a reference Cesium standard, some measurement system for 5 MHz with 10^{-13} resolution, and a 12 MHz synthesizer with Hz resolution are required beyond rather common equipment for any laboratory. The reference standard is necessary for the frequency measurements that are needed to determine the power insensitive field, and the synthesizer is necessary to accurately measure the Zeeman frequency by substituting for the internal synthesizer. If accuracy is not a concern the external 12 MHz can even be given up, if necessary, by subrogating with C-field current evaluation. In this case only power sensitivity can be measured for each C-field setting. The search for a zero crossing is done by trial and error once the cavity has been tuned carefully on frequency to ensure that zero crossings exist. Cavity tuning is best done using the beam itself as a power detector, which is possible by working at a power level lower than optimum power, where the beam signal is proportional to power. The provided attenuator must be inserted sufficiently into the waveguide for this operation. In this way the cavity can be tuned by maximizing the beam signal. Care must be taken to ensure that at all times impedance is matched at the input of the cavity, that is to say that small adjustments must be done in the matching section after each little step of cavity tuning.

The only real problem with this alignment procedure for maximization of long term stability is that after the C-field has been moved from its original value to a point where power sensitivity is minimized, the output frequency of the standard suffers from a huge offset because the internal synthesizer is set for the exact compensation of the original C-field offset. It is true that this offset can be predicted with high accuracy, because in that situation all the systematic effects are known to the degree shown in Table 1, but the output frequency will still be some 10^{-10} off, and this may give at least number truncation problems in the data handling system. A more agile frequency synthesizer seems to be the sole solution to this problem. Another minor annoyance of this method may be the fact that the standard must be realigned each time the tube is replaced due to end of life.

Conclusions and acknowledgements.

This paper is of course not providing any new material, because all the research which has spurred it has been already published or is still the subject of current study. Instead it wants to be an overview of the possibilities opened for users and manufacturers of Cs beam standards, and for all the time and frequency community, by some recent developments in the field. It is the author's opinion that the 10^{-13} accuracy level given here as a target for short tubes may well be the ultimate goal for such devices.

The encouragement of many colleagues throughout the recent years is gratefully acknowledged, and a special thank goes to Dr. Claudine Thomas of the BIPM, for suggesting the realization of this review.

References.

[1] D.W.Allan, H.Hellwig, S.Jarvis, D.A.Howe, R.M.Garvey - Proc. 31st FCS (1977), p.555.
[2] A.De Marchi - Proc. 1st European Time & Freq. Forum, Besancon (1987), p.288.
[3] A.De Marchi - Proc. 41st FCS (1987), p.53.
[4] A,De Marchi - Proc. 4th Symposium on Frequency Standards and Metrology, Ancona (1988). Spinger Verlag, 1989. A.De Marchi, Editor. p.52.
[5] A.De Marchi, G.D.Rovera, A.Premoli - Metrologia 20 (1984), p.37.
[6] A.Bauch, K.Dorenwendt, T.Heindorff - Metrologia 24 (1987).
[7] A.De Marchi, A.Premoli, G.D.Rovera - IEEE Trans.UFFC-34,6 (1987), p.582.
[8] L.Cutler, R.Giffard - Private comm. (1988).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-05 Number 2

Experimental data on a new Ramsey cavity for cesium atomic clocks

H. de Boer, B. Fischer, T. Heindorff, R. Schröder
Physikalisch-Technische Bundesanstalt, Braunschweig

Summary

The PTB's experimental cesium beam apparatus CSX has been equipped with ring cavity end parts and experimentally tested. The frequency shift due to the distributed cavity phase difference was found to be reduced by at least half an order of magnitude to $2 \cdot 10^{-13}$/mm compared with the traditionally designed cavity end parts. Ultimate verification of the theory failed in consequence of the insufficient sensitivity of our apparatus.

1. Introduction

The contribution of the distributed cavity phase difference to the total uncertainty of primary and commercial clocks is still the largest one, although mechanical imperfections in the fabrication of conventional cavities were removed and the precision of beam reversals in laboratory standards increased. In parallel with this development the atomic beam velocity was lowered even further, with the result that in the PTB a new generation of upright standing atomic beam tubes is operating with an average beam velocity of 69 m/s (FWHM = 14 m/s), albeit at the expense of a reduced beam intensity /1/. Elsewhere, laser cooling of an atomic cesium beam has been demonstrated /2/ but more research is necessary before this method can be used on an existing clock. However, a limit seems to have been reached, and only a better cavity design promises further progress in the future. De Marchi, Shirley, Glaze and Drullinger have proposed a ring cavity /3/. They checked the validity of their assumptions by experimenting with a prototype ring cavity, but ultimate verification of its function in an atomic cesium clock was not possible.

With its experimental atomic beam apparatus CSX, among other things, the PTB determined the distributed cavity phase difference of a traditionally designed microwave cavity /4/. It was therefore logical to continue this line of thought and construct new cavity end parts as ring cavities and substitute them for the CSX ones long used. In the following these ring cavities will be described and the measurement results of our experimental investigations be presented and discussed.

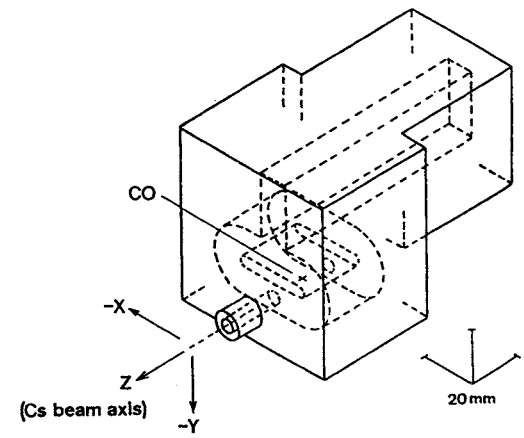


Fig. 1: Artist's impression of one of the end cavities proposed by A. De Marchi (see ref. /3/) and tested in CSX. CO centre of irradiation zone and origin of X, Y coordinate.

2. Ring cavity

Fig. 1 shows the end cavity in front of the analyzer magnet. It consists of three parts milled in copper: A horizontal cut halves the height of the upper waveguide, while a vertical one is perpendicular to the beam axis and passes through the centre CO of the intrinsic ring cavity. All three parts are held together by pins. To prevent rf leakage the outer slits were carefully closed by means of light surface soldering. The inner surface of the cavity was therefore not penetrated by soldering material but contained open slits. During the evacuation of CSX these slits were the source of gas bursts. A procedure of heating the whole apparatus up to 70°C which took several days was necessary to eliminate these.

As described in /4/ the cavity is rotatable around the center of the first irradiation region. The second irradiation region can then be displaced relative to the X and Y axis of the cesium beam-connected coordinate system.

3. Measurement results and discussion

The operating parameters of CSX deviate in two points from those in /4/: The total beam intensity increased from 100 pA to 118 pA and the resonance signal from 13 pA to 15 pA. It was also necessary to change the C-field strength. Fig. 2 depicts the microwave spectrum of CSX for a low C-field and a high C-field. In the upper part of the figure the C-field strength is adjusted to 5.1 A/m or a Zeeman frequency of 45 kHz while 12.5 A/m or 110 kHz are applied in the lower part. The latter Zeeman frequency was chosen in order to separate the Rabi pedestals and to reduce to a minimum the possible frequency shift of the (4,0)-(3,0) transition frequency caused by neighbouring π-transitions ($\Delta F = 0$, $\Delta m_F = \pm 1$). Since in our microwave and C-field configuration, along off-axis trajectories the cesium atoms are irradiated by transversal field components, the excitation of π-transitions is unavoidable.

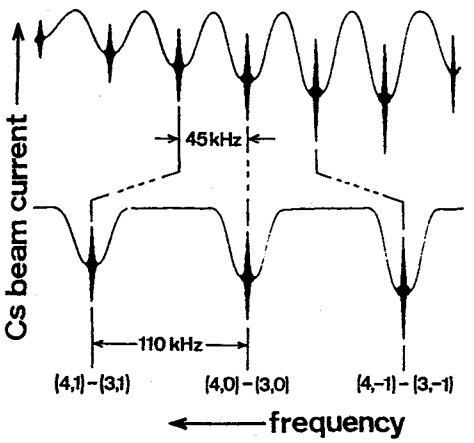


Fig. 2: Part of the microwave spectrum of CSX for a low C-field (Zeeman frequency 45 kHz) and for a high C-field (Zeeman frequency 110 kHz).

Fig. 3 shows the different power dependence of the CSX resonance frequency for the two C-field values. For power values more than 2 dB above optimum power P_0, frequency deviations of some parts in 10^{13} appear. The reason is clearly the differing positions of the Rabi pedestals and π-transitions.

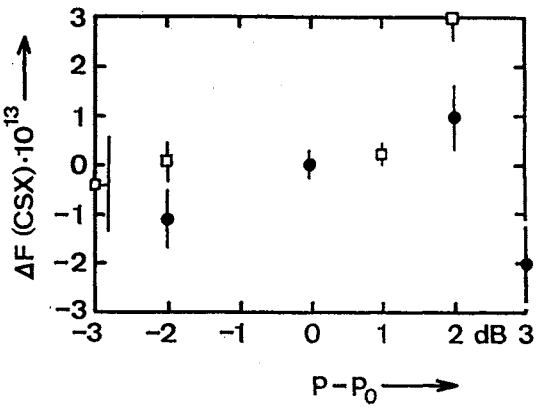


Fig. 3: Power dependence of the relative CSX frequency. P_0 is the optimum microwave power and ΔF is the frequency difference $F(CSX)_P - F(CSX)_{P_0}$ for low C-field (open squares) and for high C-field (full dots).

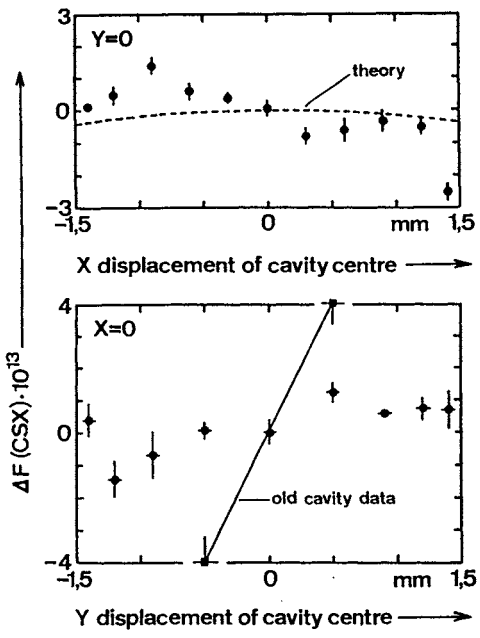


Fig. 4: Relative CSX frequency F as a function of the cavity position at a Zeeman frequency of 110 kHz. Here ΔF is the frequency difference $F(CSX)_{X,Y} - F(CSX)_{X=0,Y=0}$ at optimum microwave power. The theoretical curve was calculated with the help of De Marchi's results (see ref. /3/). The "old cavity data" were taken from ref. /4/ in which an H_{301}-type cavity end part was studied.

Fig. 4 shows the ring cavity measurement results. The upper diagram illustrates the influence of displacements of the second ring cavity centre along the X axis, and in the lower diagram along the Y axis on the output frequency of CSX.

Over a period of 24 hours the CSX frequency was compared with that of the best high performance clock (HP 004) for a fixed displacement. This reference clock was calibrated in parallel with CS1, the PTB's primary clock. A mean value of the statistical measurement uncertainty of CSX versus CS1 was $6.5 \cdot 10^{-14}$. The new displacement was adjusted from day to day; only at weekends did it remain unchanged for 3 days. A sequence of arbitrary positive and negative displacements followed, whereby the mechanical uncertainty of an X displacement was 0.03 mm, and that of a Y displacement 0.05 mm.

If we start from X=0 and move the cavity centre in the X direction (y coordinate in /3/, p. 187), a weak parabolic frequency shift of CSX as indicated by the dashed line in Fig. 4, is expected. The measurement values obviously do not confirm the theory. The effect of Y displacements on the frequency difference between CSX and CS1 shows the same discrepancy between the expected independence of the CSX frequency from Y displacements and the shape of the measurement values. It is evident from Fig. 4 that ring cavities have almost no distributed cavity phase difference and that other parameters control the CSX frequency. On the basis of Fig. 4 an upper limit for the distributed cavity phase difference frequency shift $\Delta F(CSX)_{DCPD}$ is estimated:

$$\begin{aligned} |\Delta F(CSX)_{DCPD}| &\leq 2 \cdot 10^{-13} \; |\Delta X/mm| \\ &\leq 2 \cdot 10^{-13} \; |\Delta Y/mm| \quad . \quad (1) \end{aligned}$$

In the lower diagram in Fig. 4 the frequency shift due to the distributed cavity phase difference, measured for CSX equipped with a conventional cavity, is drawn as a straight line and is $10 \cdot 10^{-13}$/mm /4/. Taking into account the different beam velocities (CSX:416 m/s; CS1/CS2: 93 m/s) this result was transferred to CS1 and CS2 and used to estimate the distributed cavity phase difference uncertainty contributions of CS1 and CS2. If it is assumed that this method is still applicable, it follows that these contributions could be reduced by a factor of 5 when operating PTB's CS1 and CS2 /5/ with ring cavities. In the case of CS2 the C-field uncertainty contribution of $5 \cdot 10^{-15}$ would become the largest of all, while the distributed cavity phase difference would be reduced from 10 to $2 \cdot 10^{-15}$, i.e. the total uncertainty of CS2 could break the 10^{-14} barrier.

In the future CS2 will be the first of our primary clocks to run with the new ring cavities being manufactured.

4. Conclusions

Our measurement results demonstrate that a Ramsey resonator equipped with ring cavities is capable of reducing the total uncertainty of primary clocks to below the 10^{-14} barrier.

References

/1/ Fuhrmann, H.; T. Heindorff: The new generation of atomic cesium clocks. PTB Jahresbericht (Annual Report) 1988, p. 17 - 18

/2/ Watts, R.N.; C.E. Wieman: Manipulating atomic velocities using diode lasers. Opt. Letters 11, p. 291 - 293 (1986)

/3/ De Marchi, A.; J. Shirley; D.J. Glaze; R. Drullinger: A new cavity configuration for cesium beam primary frequency standards. IEEE Trans.Instr.Meas. IM-37, p. 185 - 190 (1988)

/4/ Bauch, A.; T. Heindorff; R. Schröder: Measurement of the frequency shift due to distributed cavity phase difference in an atomic clock. IEEE Trans.Instr.Meas. IM-34, p. 136 - 138 (1985)

/5/ Bauch, A.; H. de Boer; B. Fischer; T. Heindorff; R. Schröder: Long-term performance of the PTB's primary clock CS2. Proceed. 2nd EFTF, p. 425 - 435 (Neuchâtel, 16. - 18.3.88)

Session FS-05 Number 3

STUDY OF THE CORRELATIONS AMONG THE FREQUENCY CHANGES OF THE CONTRUBUTING CLOCKS TO TAI

Patrizia TAVELLA*
Istituto Elettrotecnico Nazionale Galileo Ferraris
Strada delle Cacce, 91
10135 Torino
Italy

Claudine THOMAS
Bureau International des Poids et Mesures
Pavillon de Breteuil
92312 Sèvres Cedex
France

* Guest worker at the BIPM from October to December 1989

Abstract

The Comité consultatif pour la définition de la seconde declared at its last meeting (1989 April 19-20) that:

"considering that correlation among the frequency changes between contributing clocks can degrade the long-term stability and accuracy of TAI,

recommends

- that the BIPM studies and reports on correlation in frequency changes among the clocks contributing to TAI,

and

- that the BIPM acts with contibuting laboratories towards improving statistical independence of the clocks."

This paper is related to the first part of this declaration.

The variations of bimonthly frequencies, relative to TAI, of the clocks maintained in continuous operation for a long time, have been analysed by several methods, for evidence of correlation. The analyses show that strong and persisting correlations do exist among the clocks of individual laboratories. Typical examples where the existing correlations are due to the time scale TAI itself, to environment of the clocks or to the time link the clocks are seen through are shown.

One particular study is devoted to two clocks known to react differently to variations of humidity.

Correlation arising from the seasonal variation of TAI is demonstrated by a study of the frequency variations of three very stable clocks located in different parts of the world: the Australian H-maser AUS 44 2, the American H-maser APL 42 13 and the German primary standard PTB Cs2.

A typical example of "fugitive" correlation, appearing only during a short period of the year is also given. This involves a large number of clocks belonging to the same laboratory and can perhaps be attributed to a particular change in their environmental conditions.

In one case, a non detectable correlation is reported.

Introduction

International Atomic Time, TAI, is derived from a free atomic time scale EAL (échelle atomique libre), the frequency of which is steered retrospectively for accuracy. The EAL is a weighted average of the readings of about 200 free running atomic clocks spread accross the world. One of the basic hypotheses of the EAL algorithm is that the contributing clocks are statistically independent, the weight of each being assigned following its long-term stability [1].

An outside event can influence simultaneously the behaviour of a set of clocks and then the resulting time scales. It can create correlation among the clock frequencies and degrade the stability and accuracy of TAI.

This paper exhibits some results of a study about the correlation appearing between the frequencies relative to TAI of the contributing clocks. It follows one of the declaration of the last CCDS meeting (1989, April 19-20) [2].

This study involves all the contributing laboratories keeping at least two clocks with continuous operation exceeding twenty months. Only typical examples are given here. The methods used are described in the first section.

Apparent persistent correlations in the set of clocks belonging to a given laboratory or among clocks belonging to different laboratories have been detected. They may be ascribed to three main causes, which may occur simultaneously:

- correlation due to TAI time scale itself reinjecting its own seasonal variation on the contributing clocks. Section 2 presents examples corresponding to this case: apparent correlation between the two primary standards PTB Cs1 and PTB Cs2 frequencies (see the acronym list at the end) or among hydrogen masers and cesium primary standards located in different parts on the world.

- correlation due to changes in the clock room environmental conditions, in particular humidity and temperature variations. Examples showing clocks kept by the NAOM, CH and IEN are given in section 3,

- correlation due to the time link the clocks are seen through: when this link is noisy it may occult the true own qualities of the clocks. This is the case of the Russian masers, the frequencies of which appear to be strongly correlated because of the Loran C link between Moscou and Paris (section 4).

Section 5 describes an example of "fugitive" correlation. It involves the USNO clock set and appears only during a short period (November 1989) as if a local and brief event had influenced the global behavior of the clocks.

The example of TAO where no correlation is detectable over the recent years is described in the last section.

For each case, we specify the averaged relative weights of the involved clocks in the TAI computation, so as to emphasize in the conclusion, the response of TAI to the correlations which have appeared.

I - Methods used to detect out correlations among the clock frequencies

The clock H_i frequency relative to TAI is:

$$B_i(t) = \frac{\beta(t) - \beta_i(t)}{\beta(t)} \qquad (1)$$

where $\beta(t)$ and $\beta_i(t)$ are the mean frequencies of TAI and of clock H_i over the interval, of basic duration τ = 60 days, ending at date t.

Since the quantities $\beta(t)$ and $\beta_i(t)$ are inaccessible separately, only the normalized frequency departure $B_i(t)$ from TAI can be estimated. $B_i(t)$ then keeps memory of the contribution of TAI itself.

Correlation among the variations of bimonthly frequencies relative to TAI can be shown out in different ways.

As a first approach, one can graph the frequencies $B_i(t)$ versus time for a given set of clocks H_i (i = 1, --, n) and compare their variations.

In a statistical and more quantitative approach the quantities $B_i(t)$, obtained for clock H_i over a long period (t - mτ, ---, t - τ, t) are expressed as a time series B_i. The correlation between the clock H_i and clock H_j frequencies is then given by the cross-correlation function $\mu_{ij}(k)$, at lag k, of the bivariate process (B_i, B_j)[3], defined as:

$$\mu_{ij}(k) = \frac{\overline{[\,(B_i(t) - \bar{B}_i) \cdot (B_j(t + k\tau) - \bar{B}_j)\,]}}{\sigma_i \cdot \sigma_j} \qquad (2)$$

where B_i and σ_i are the mean value and the standard deviation of the time series B_i.

Here $\mu_{ij}(k)$ is estimated by computing the means and standards deviations over a range of 10 to 20 values.

In most cases, k is choosen equal to zero. If the resulting $\mu_{ij}(k)$ are close to + 1 or - 1, the two clocks H_i and H_j are highly correlated or anticorrelated at lag k. If $|\mu_{ij}(k)|$ is close to zero, the two clocks H_i and H_j are uncorrelated.

An alternative way to vizualise correlation between H_i and H_j frequencies relative to TAI is to graph $B_j(t)$ versus $B_i(t)$. The obtained points are no more ordered in time.

Computing the regression line of B_j on B_i, one obtains:

$$B_{j\ r.l.} = s_j \cdot B_i + \text{constant} \qquad (3)$$

In the same way, the regression line of B_i on B_j gives:

$$B_{i\ r.l.} = s_i \cdot B_j + \text{constant} \qquad (4)$$

It can easily be shown [4] that:

$$[\mu_{ij}(0)]^2 = s_i . s_j \qquad (5)$$

s_i, s_j and $\mu_{ij}(0)$ having the same sign.

On the graphs of B_j versus B_i, the two regression lines defined by (3) and (4) are represented with the slopes s_j and $1/s_i$. If $\mu_{ij}(0)$ is close to ± 1, the two regression lines are nearly superposed. If $\mu_{ij}(0)$ is near zero, the two regression lines lie close to, respectively, the horizontal and vertical axes.

II - Apparent correlation due to TAI itself

Example 1: The PTB clocks

Fig. 1 shows the two primary standard PTB Cs1 and PTB Cs2 frequencies relative to TAI for a period of over 3 years.

Annual variations can be seen. They have common phase and amplitude, the peak-to-peak amplitude being about 4 ns/day ($4,6 \times 10^{-14}$).

It is likely that this seasonal variation is due to TAI itself. Previous studies [5] have already pointed out that EAL has an annual fluctuation with respect to the primary frequency standards.

The strong correlation observed between the two PTB primary standards with

$$\mu_{ij}(0) = 0,72$$

is therefore more apparent than real.

Because of their very good long-term stability, the PTB primary standards have the maximum relative weight (≃ 2 %) in the TAI computation.

Figure 2 shows $B_{Cs2}(t)$ versus $B_{Cs1}(t)$ together with their corresponding regression lines as defined in section 1 (eq. 3 and 4).

Note that these two straight lines necessarily cross at the origin because the frequency values are shown in reduced form.

The two slopes are positive and are given by:

$s_1 = 0,49$ and $1/s_2 = 1/1,07$.

This confirms the value $\mu_{ij}(0) = 0,72$.

Figure 3 represents the frequencies $B_i(t)$ relative to TAI versus time for six PTB clocks which operated continuously from January 1988 to October 1989. We do not observe a global movement involving the whole set of clocks but two of them, PTB 14 394 and PTB 14 867, exhibit an annual variation with the same phase. This fluctuation has a peak to peak amplitude equal to 12 ns/day ($1,4 \times 10^{-13}$), which is much larger and in opposite phase to the Cs1 and Cs2 frequency variations noted above. So it cannot be attributed entirely to TAI itself. The graph of B 14 867 versus B 14 394 and the two regression lines are shown on Fig. 4: the corresponding cross-correlation coefficient $\mu_{ij}(0)$ is 0,53.

The correlation between these two clocks is the highest observed for PTB commercial clocks and may perhaps be assigned to environmental conditions of the clocks, these two clocks reacting more strongly than others, and in the same sense, to some humidity or temperature changes.

These two clocks were introduced into the TAI computation with averaged relative weights, for 1989, of about 1,6 % and 0,4 %. Their seasonal variations thus contribute to the TAI seasonal variation.

Example 2: Apparent correlation between the two APL Hydrogen masers 42 13 and 42 14.

Fig. 5 shows the frequency variations versus time of the APL H-masers 42 13 and 42 14 for a 20 month period. (They were taken out of operation at the beginning of 1989).

One can observe that these two H-masers are affected by a drift and a seasonal variation. Again an apparent correlation exists between these clocks: the cross-correlation coefficient $\mu_{ij}(0)$, calculated over the period under study, is 0,81.

Taking into account the value of the peak to peak amplitude ($4,7 \times 10^{-14}$), and the phase of the apparent seasonal variation, one may think that TAI is mainly responsible for that correlation. This interpretation is then similar to the apparent correlation between PTB primary standards (example 1).

The two APL H-masers had for that period the maximum weight in the TAI computation.

Example 3: Apparent correlation between three very stable clocks located in different parts of the world.

To point out again the apparent correlations induced in clocks by the seasonal variation of TAI, we choose here to look at the frequency variations of three different clocks, chosen among those of maximum weight. They are located in three different parts of the world, that is to say in laboratories experiencing very different climates:

- PTB primary standard Cs2, for Europe,
- APL H-maser 42 13, for North America,
- AUS H-maser 44 2, for Australia.

Fig. 6 shows the bimonthly frequencies, relative to TAI, for a 28 month period ending in October 89. These three clocks exhibit seasonal variations of nearly same amplitudes and same phases, even although one is located in the southern hemisphere. This means that what we mainly see here is the seasonal variation of TAI itself. The cross-correlation coefficients obtained are:

i = PTB Cs2	j = AUS 44 2	μ_{ij} (0) = 0,80
i = PTB Cs2	j = APL 42 13	μ_{ij} (0) = 0,77
i = AUS 44 2	j = APL 42 13	μ_{ij} (0) = 0,70

In Fig. 7, are shown the regression lines of AUS 44 2 and PTB Cs2 frequencies versus APL 42 13 frequency. They are close to each other and close to the straight line of slope + 1. This means that the three clocks frequencies, relative to TAI, are highly correlated, their frequency reactions to the correlating event being of the same order of magnitude.

This example of apparent correlation allows an estimate to be made of the TAI seasonal variation. Its peak to peak amplitude appears to be of the order of 5×10^{-14} with a maximum at the end of winter. In the following sections, a larger value or a different phase in the annual fluctuation of a clock frequency relative to TAI will be mainly related to the clock itself rather than to TAI.

III - Correlation due to changes in the clocks environment

Example 4: Humidity variations in NAOM

We have already presented an example of apparent correlation for PTB 14 394 and 14 867 (section II), due perhaps to changes in the environment of these clocks.

Another example, perhaps more typical is given on Fig. 8. It concerns two cesium clocks 14 1315 and 14 885 kept by NAOM in Mizusawa (Japan). These two clocks seem correlated by a seasonal variation (μ_{ij} (0) = 0,80). Its peak to peak amplitude is about 15 ns/day ($1,7 \times 10^{-13}$) which is much bigger than that observed for TAI.

The relative humidity variation was recorded in this laboratory and reported to the BIPM for 1988 and 1989. It varied from 35 % in winter to more than 70 % at the end of summer, the temperature meanwhile being constant to within 1°.

Fig. 8 shows that the frequencies, relative to TAI, of these two clocks follow the humidity variations, and so display high correlation. The clocks were entered in the TAI computation with relative weights of about 0,7 % for 1989.

Example 5: Swiss clocks

In Fig. 9, the frequencies relative to TAI of four clocks kept in Switzerland, are reported. They present seasonal variations ranging from 10 to 30 ns/day ($1,1 \times 10^{-13}$ to $3,3 \times 10^{-13}$). Again we see a case of clocks correlated by environmental conditions.

The cross-correlation coefficients μ_{ij} (0) have been computed for these four clocks and are reported in the following matrix (table 1):

table 1

	CH 12 863	CH 16 64	CH 21 179	CH 21 243
CH 12 863	1			
CH 16 64	0,61	1		
CH 21 179	0,38	0,41	1	
CH 21 243	0,80	0,35	0,5	1

Two of these clocks are highly correlated (21 243 and 12 863). It can be stressed that they belong to the same Swiss laboratory, located in Neuchâtel, their relative weight in the TAI computation for 1989 being about 0,2 %.

Example 6: IEN clocks

Another case analysed is shown in Fig. 10a. This concerns clocks 14 893 and 16 84 kept by IEN in Turin (Italy) and observed for a two year period. These clocks present high seasonal variations but have different peak to peak amplitudes (30 ns/day and 45 ns/day) and different phases.

The relative humidity recorded in the room where the clocks are kept is reported in Fig. 10a; the temperature was kept constant within ± 0,5 °.

It is evident that humidity variations influence the frequencies of these two clocks but the effect differs in amplitude and phase. We have computed the cross-correlation function μ_{ij} (k) at different lags k to point out the existing correlation, the results are reported in Fig. 10b.

High correlation coefficient (negative in this case) appears when one B series is shifted of one bimester (k = 1) showing a difference in the reaction time constant of the two clocks of about 2 months.

Between the maximum of anticorrelation (k = 1 or k = 7) and the maximum of correlation (k = 4), there are six months as would be expected for a seasonal variation.

In the TAI computation, these two clocks had relative weights of about 1 % and 0,15 %, in the period examined.

IV - Apparent correlation assigned to the time links

Example 7: Russian Hydrogen masers

The most obvious case of correlation involves Russian clocks: the two H-masers SU 381 and SU 382 (Fig. 11). They appear to be totally correlated because they are seen through a Loran C time link, very noisy in this case, which completely occults their long-term qualities. We obtained: μ_{ij} (0) = 0,999.

The graph of $B_j(t)$ versus $B_i(t)$ (Fig. 12) shows good alignment of the points on the straight line of slope + 1.

These two H-masers are intentionally weighted to zero in the TAI computation.

This study which has considered only bimonthly frequencies, cannot point out correlations due to the GPS time transfers whose noise at this time interval is smoothed out.

V - "Fugitive" correlation

Example 8: Global variation of USNO clock frequencies

Table 2 gives the variations of the frequencies, between consecutive 60-day samples, for the complete clocks set kept by USNO, for the year 1989.

The reported quantities are:

$$\delta B_i(t) = B_i(t) - B_i(t - \tau) \qquad (6)$$

where τ = 60 days.

It may be seen that 24 clocks of the 33 clocks entered in the TAI computation have decreasing frequencies between September-October 1989 and November-December 1989 (last column). Such a global variation did not occur before in 1989.

At the bottom of Table 2 we report, for each period, the relative weight of the USNO clock ensemble in the TAI computation defined as:

$$P(USNO) = \frac{\sum_i p_i(USNO)}{\sum_i p_i(TAI)} \qquad (7)$$

and the frequency variation given to TAI by the USNO ensemble, defined as:

$$\Delta B(USNO) = \frac{\sum_i p_i \cdot \delta B_i(USNO)}{\sum_i p_i(TAI)} \qquad (8)$$

where pi (clock H_i weight) and δB_i (as defined in eq. 6) are computed over the two month interval ending at date t.

It may be seen that in the last bimester of 1989 the USNO ensemble entered in TAI computation with a high relative weight and gave to TAI a frequency variation much bigger than before. The above correlation appears fugitively and involves almost all the clocks in the laboratory.

To examine this case in detail, we have computed and plotted the monthly frequencies for the period September-December 1989, in Fig. 13 where the decreasing frequencies are underlined. It appears that almost the whole set of clocks reacted to some external event during November 1989 which changed and correlated their frequencies.

Such a phenomenon could have degraded both the stability of the atomic time scale computed by USNO and the stability of TAI.

VI - No detectable correlation

Example 9: TAO clocks

To complete this study, consider Fig.14a which shows the frequencies, relative to TAI, of six clocks kept by TAO in Tokyo. The computed cross-correlation coefficients, μ_{ij} (0), for each pair of clocks, are all less, in absolute value, than 0.35, so they show no detectable correlation.

Any high seasonal variation is evident. In fact, these clocks are kept in particularly good conditions with temperature and humidity control and also electromagnetic shielding between them [6].

The total relative weight of TAO clocks in TAI computation is about 5 % for 1989.

In this data the correlation reinjected by TAI is not detectable because it is occulted by other noise. However if we plot the frequency relative to TAI of the best clock of TAO (142494) and that of PTB Cs(2) (Fig. 14b) an annual fluctuation comes out and the two clocks appear highly correlated (μ_{ij} (0) = 0.7). In this case we again have a very good clock (TAO 142494) whose seasonal noise is due to TAI and not to the clock itself.

Conclusion

Cases of apparent correlation among the frequency changes of the clocks contributing to TAI have been noted. These correlations are real but, because they scarcely reproduce and may occur abruptly (USNO example) or indirectly (IEN example), they cannot usefully be modelized or forecast.

It does not appear that one commercial clock type is more sensitive to correlation than others, or that the reaction to external events is identical for clocks of the same type.

Correlations among frequency changes are often related to environmental variations, in particular to changes of humidity. One way to reduce them is then to control the environmental parameters, another is to control the sensitivity of the clocks by applying techniques such as that suggested by De Marchi [7].

The weight of a given clock in TAI reflects its behavior over the previous twelve months. Seasonal variations are thus recognized and lead to deweighting. Nevertheless, TAI inherits a low (about 4×10^{-14}) but real seasonal noise.

As TAI is definitive each bimester, it does not seem feasible to introduce correlation information into the computation since this can only be established when a long period of data is available. For more precise studies, it could be necessary to issue a different time scale in which seasonal fluctuations are smoothed more completely by post processing; an a posteriori time scale as TT(BIPM) [5] is an example of how this may be done.

Acknoledgements

The authors are grateful to B. Guinot for helpful discussion.

References

[1] B. Guinot and C. Thomas, Annual Report of the Time Section of the BIPM, 1988, Part D, pp. D3-D22.

[2] The Comité Consultatif pour la définition de la seconde declares that "considering that correlation among the frequency changes between contributing clocks can degrade the long-term stability and accuracy of TAI, recommends
- that the BIPM studies and reports on correlation in frequency changes among the clocks contributing to TAI, and
- that the BIPM acts with contibuting laboratories towards improving statistical independence of the clocks", 1989 April 19-20, CCDS report in preparation.

[3] G.P. Box and G.M. Jenkins, Time Series Analysis, 1976, Holden-Day.

[4] A.L. Edwards, An Introduction to Linear Regression and Correlation, 1976, W.H. Freeman and Co.

[5] B. Guinot, Astron. Astrophys. 192, 370-373 (1988).

[6] Private communication from TAO to the BIPM.

[7] A. de Marchi, Frequency Standards and Metrology, 1989, pp. 52-56, Edited by A. de Marchi, Springer-Verlag.

Acronym List

APL Applied Physics Laboratory, Laurel, Maryland, U.S.A.

AUS Consortium of laboratories in Australia.

CH Consortium of laboratories in Switzerland.

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

NAOM National Astronomical Observatory, Mizusawa, Japan.

PTB Physikalisch- Technische Bundesanstalt, Braunschweig, Bundesrepublik Deutschland.

SU Laboratoire d'état de l'étalon de temps et de fréquence, URSS.

TAO National Astronomical Observatory, Tokyo, Japan.

NIST National Institute of Standards and Technology, Boulder, Colorado, U.S.A.

USNO U.S. Naval Observatory, Washington D.C., U.S.A.

Clock Type

12 HEWLETT-PACKARD 5061A
14 HEWLETT-PACKARD 5061A OPT. 4
16 OSCILLOQUARTZ 3200
21 OSCILLOQUARTZ 3210
31 HEWLETT-PACKARD 5061B OPT. 4
4x HYDROGEN MASERS

Clock	δB_i(47639)	p_i	δB_i(47699)	p_i	δB_i(47769)	p_i	δB_i(47829)	p_i	δB_i(47889)	p_i
142482	5.25	37	2.73	38	21.87	0	18.50	3	-15.53	3
142483	0.91	59	4.64	60	7.27	41	2.42	28	-1.98	24
142485	-10.84	8	26.04	8	-3.04	8	-11.21	9	-27.10	4
141117	3.46	27	16.26	16	-6.47	18	-4.41	20	-17.38	10
142314	3.25	15	2.17	18	17.47	6	-13.79	3	16.64	4
142486	4.92	23	3.57	17	-26.25	10	3.69	8	7.97	9
14 116	8.87	9	2.39	10	15.96	8	-17.47	13	-1.53	13
14 583	8.18	3	19.00	3	-60.93	2	14.80	2	30.05	2
142312	15.28	2	-6.41	1	-19.15	2	12.68	7	0.73	13
40 23	4.95	0	-31.83	0	-6.92	0	123.31	0	-42.27	0
141819	3.17	22	11.85	18	2.95	16	-3.52	17	-14.31	15
14 787	17.06	8	12.72	0	-4.41	7	-0.50	7	-24.55	6
43 8	4.01	0	13.18	0	30.28	0	-64.46	0	44.88	0
142481	24.45	11	5.82	9	3.11	7	3.70	5	-3.75	6
31 339	-6.90	7	-2.57	11	10.43	49	19.90	0	-14.63	9
141028	-43.91	2	-25.99	1	-40.44	1	45.06	1	21.09	1
141362	10.79	2	-6.27	3	7.53	3	0.31	8	-7.59	41
141605	4.08	42	7.35	45	2.86	37	2.41	24	-7.43	27
311034	-7.24	12	1.14	14	2.14	13	4.64	55	-4.23	100
31 335	-5.29	11	-15.93	13	10.29	16	2.56	17	-2.45	20
40 22	49.01	0	-25.17	0	-17.82	0	-13.77	0	-14.94	0
141255	3.96	100	-1.07	100	-0.70	100	2.80	100	0.12	100
14 656	-4.51	0	-0.05	73	4.89	100	1.72	100	2.22	88
141710	4.65	0	-3.52	85	-2.33	93	6.19	100	-7.49	100
142488	18.41	0	-4.71	5	-1.15	11	2.49	17	-8.12	23
141305	6.62	0	-6.32	36	-4.15	33	2.67	51	-1.24	69
141846			4.09	0	3.04	44	3.72	35	-4.75	58
40 1					-7.08	0	6.27	0	-35.37	0
141094									-8.70	0
141343									9.42	0
P (USNO) =	8 %		12 %		13 %		13 %		16 %	
ΔB (USNO) =	0,26 ns/day		0,14 ns/day		0,21 ns/day		0,33 ns/day		-0,59 ns/day	

Table 2.- Bimonthly frequencies/TAI of USNO clocks for 1989.

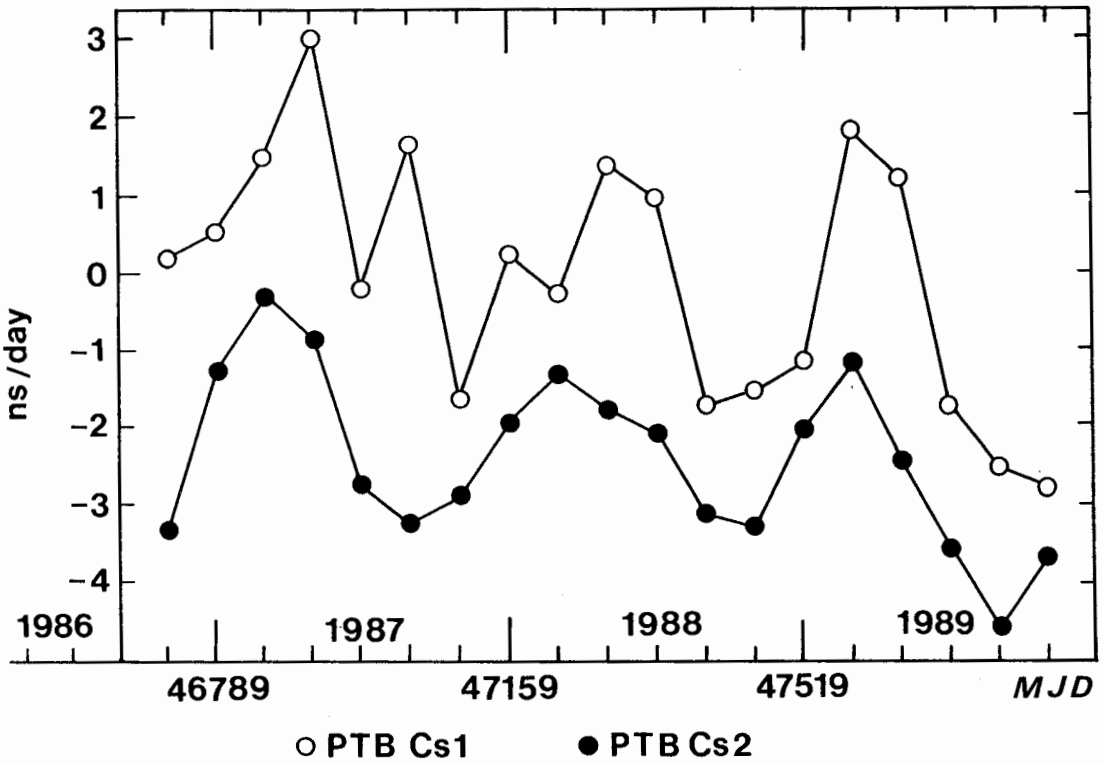


Fig. 1. Frequency of the two primary standards PTB Cs1 and PTB Cs2 relative to TAI versus time.
The basic sample duration is $\tau = 2$ months.

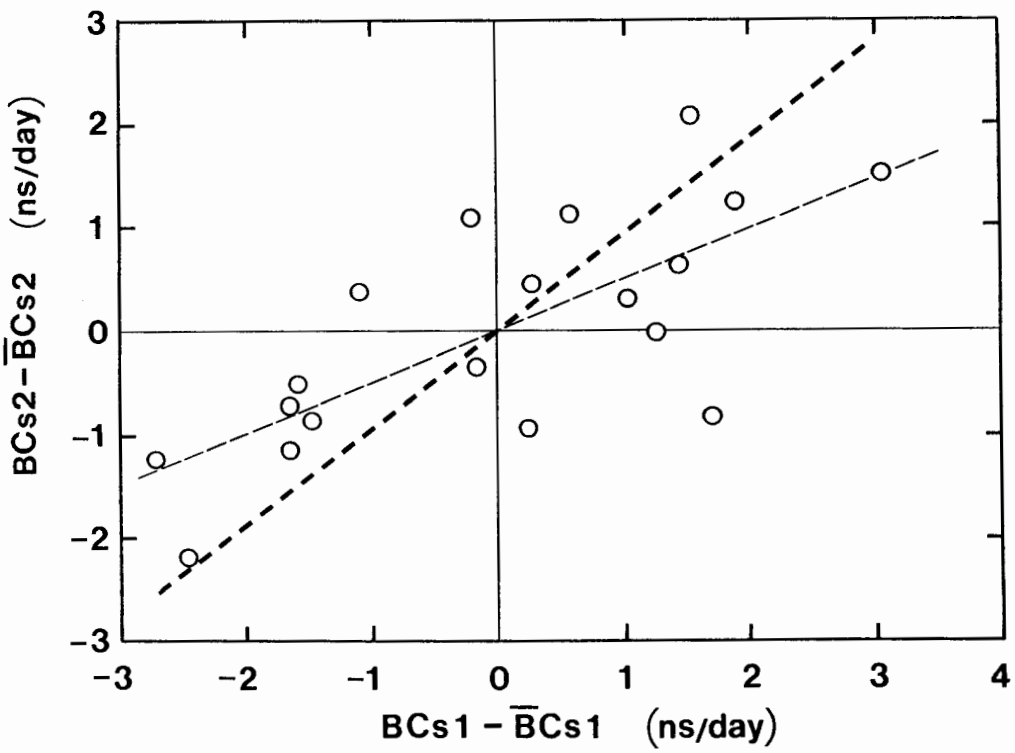


Fig. 2. Frequency of PTB Cs2 relative to TAI versus frequency of PTB Cs1 relative to TAI (the frequencies are reduced to their averages).
– – – – regression line with slope $1/S_i = 1/1.07 = 0.94$
———— regression line with slope $S_j = 0.49$

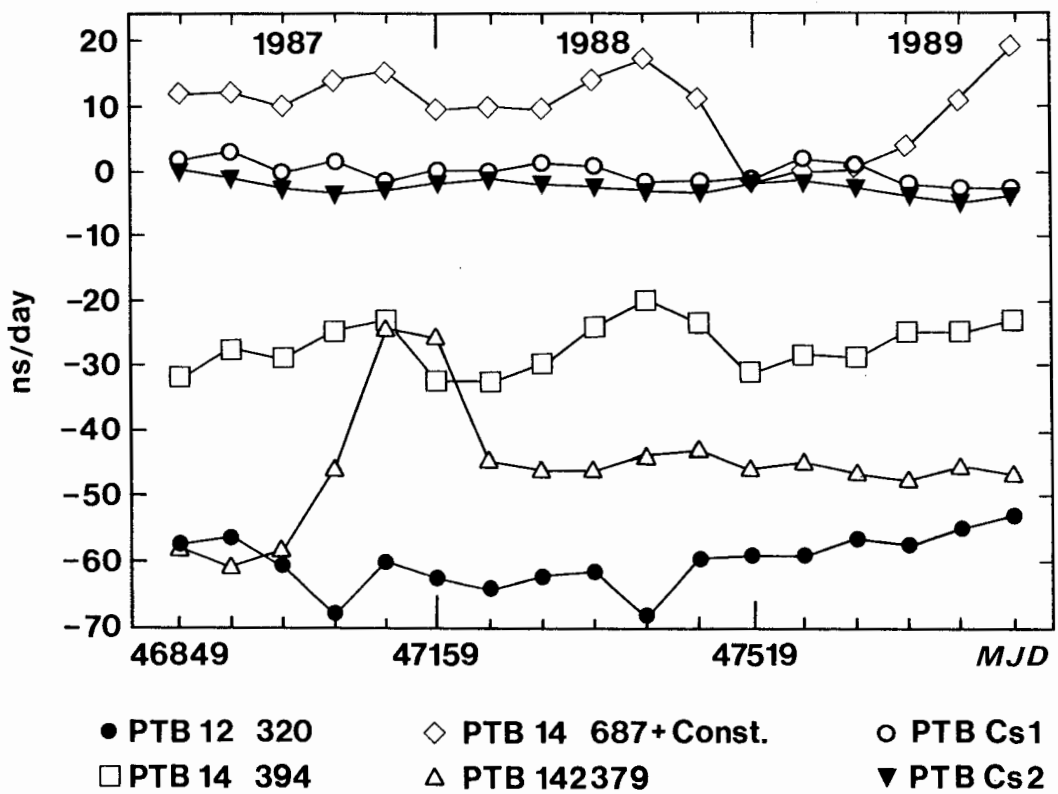


Fig. 3. Frequencies of six PTB clocks relative to TAI (including the primary standards Cs1 and Cs2), PTB 14 867 frequency has been shifted by 180 ns/day.

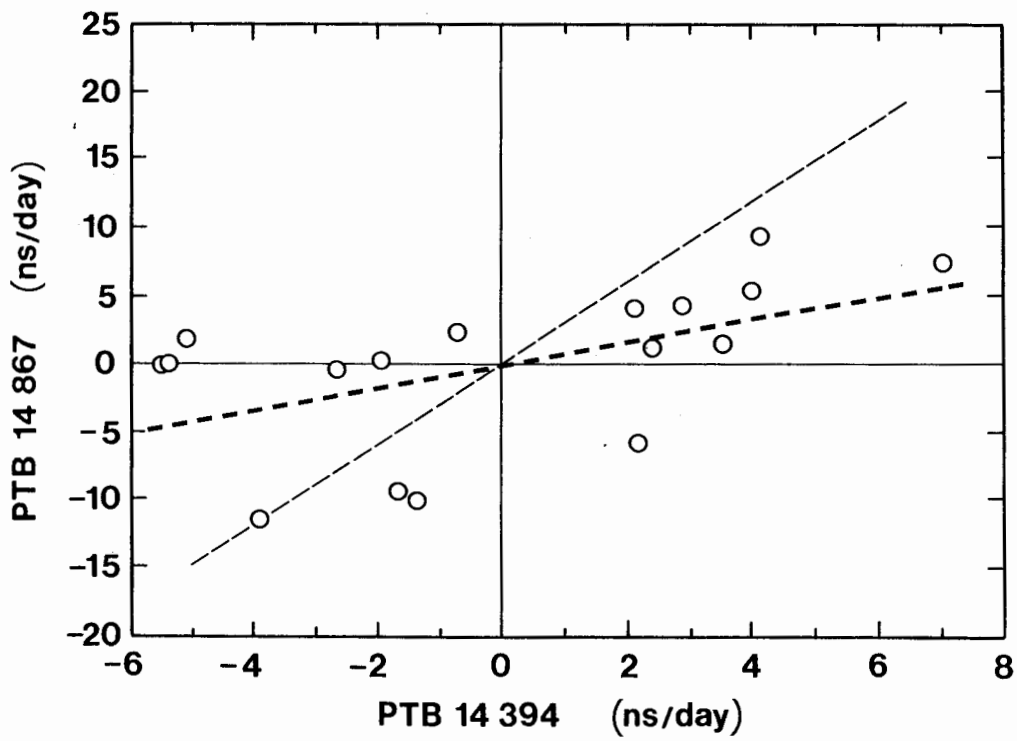


Fig. 4. Frequency of PTB 14 867 versus frequency of PTB 14 394
——— regression line with slope $1/S_j = 1/0.34 = 2.94$
- - - - regression line with slope $S_j = 0.82$

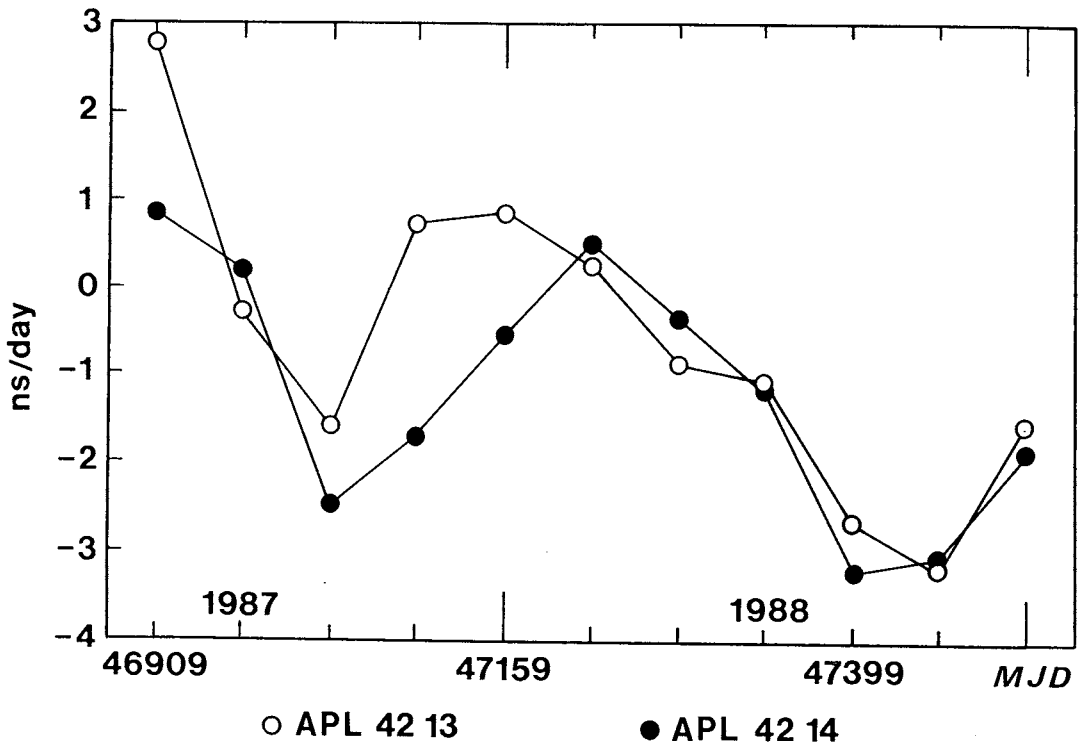


Fig. 5. Frequencies of two APL masers (42 13 and 42 14).

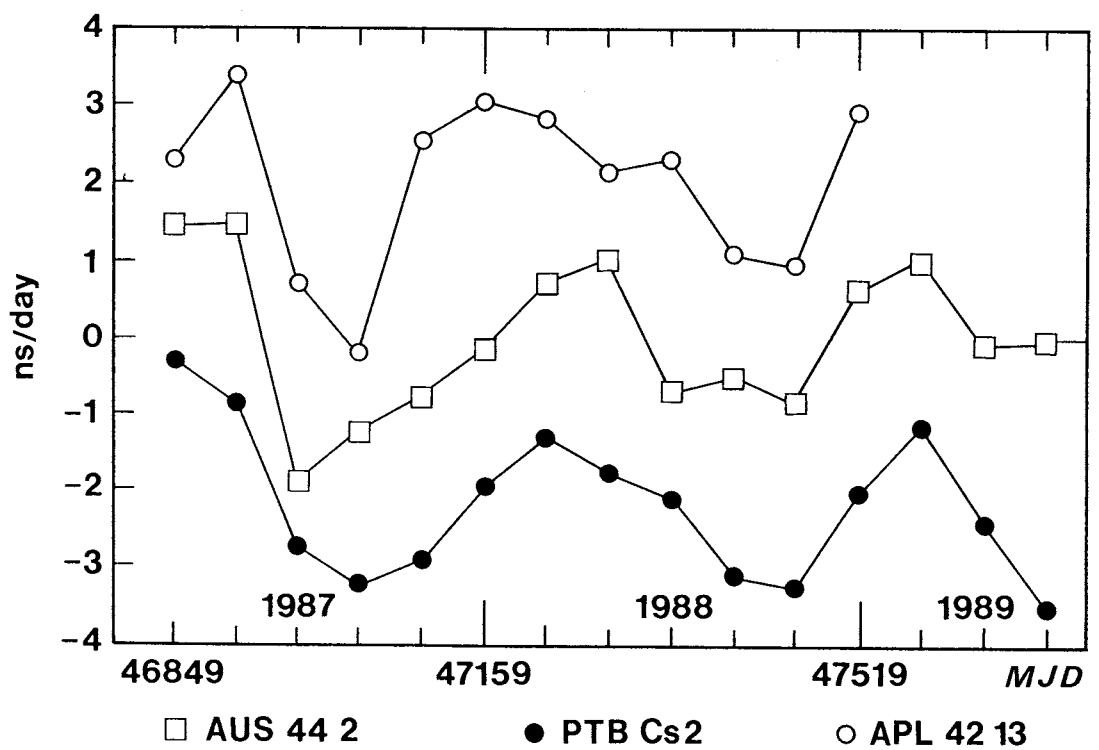


Fig. 6. Frequencies of APL maser 42 13, AUS maser 44 2 and PTB Cs2. For the two H masers, the values are globally shifted and drifts removed.

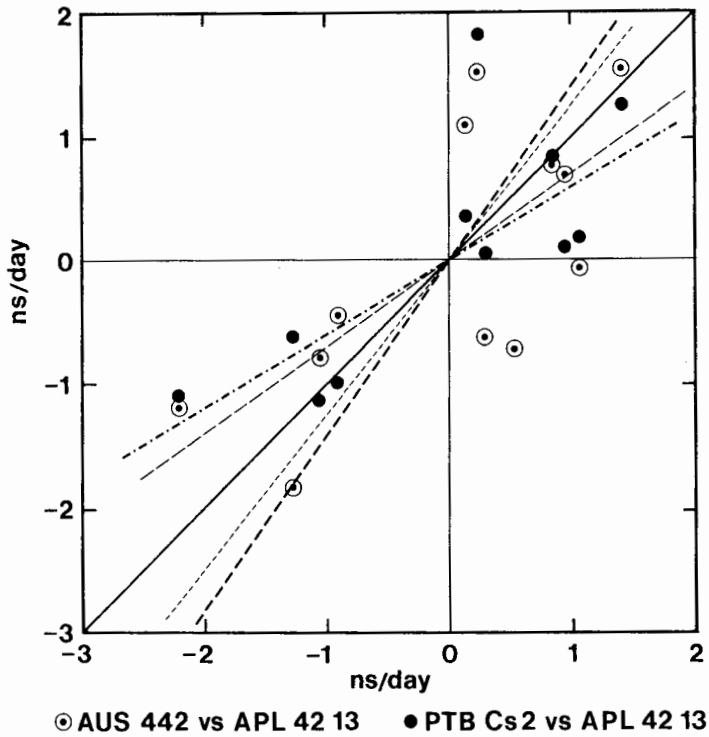


Fig. 7. Frequencies of APL H maser 44 2 and PTB Cs2 versus frequency of APL H maser 42 13 (the frequency values are reduced to their averages)
– – – – regression lines of AUS 44 2 versus APL 42 13
· · · · · regression lines of PTB Cs2 versus APL 42 13
——— straight line of slope + 1.

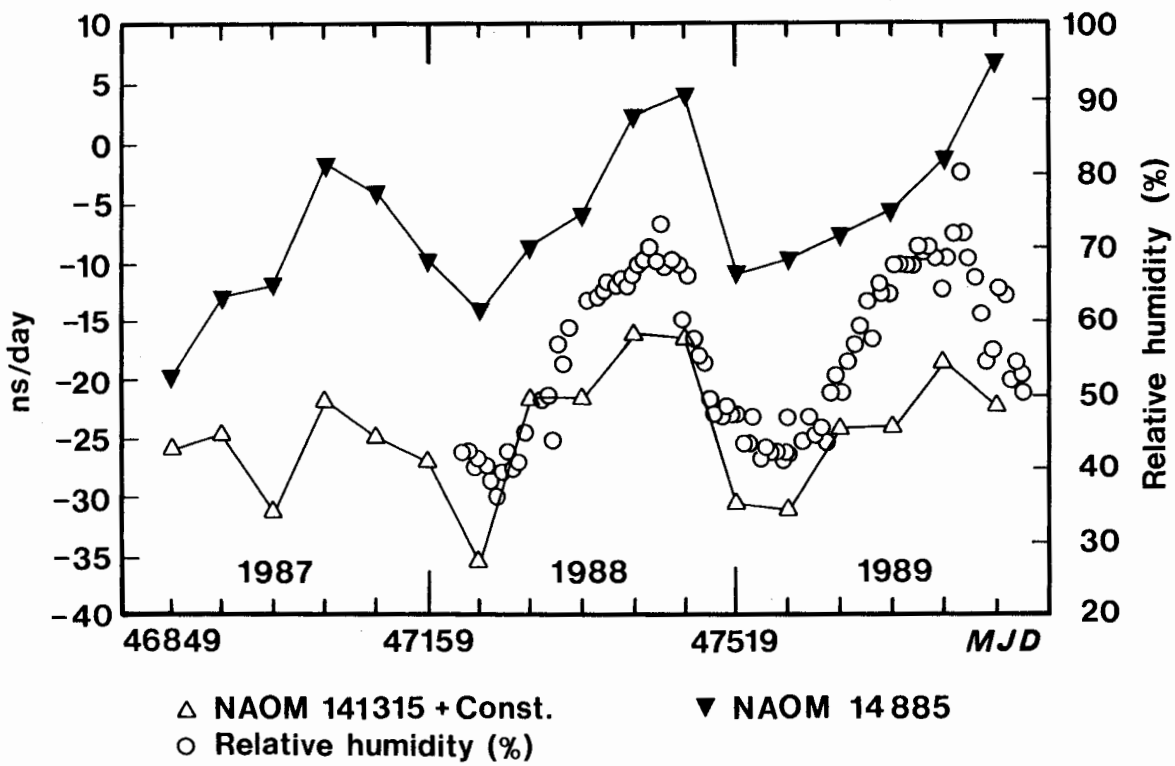


Fig. 8. Frequencies of NAOM 14 885 and of NAOM 141315 (shifted) and relative humidity of clock room.

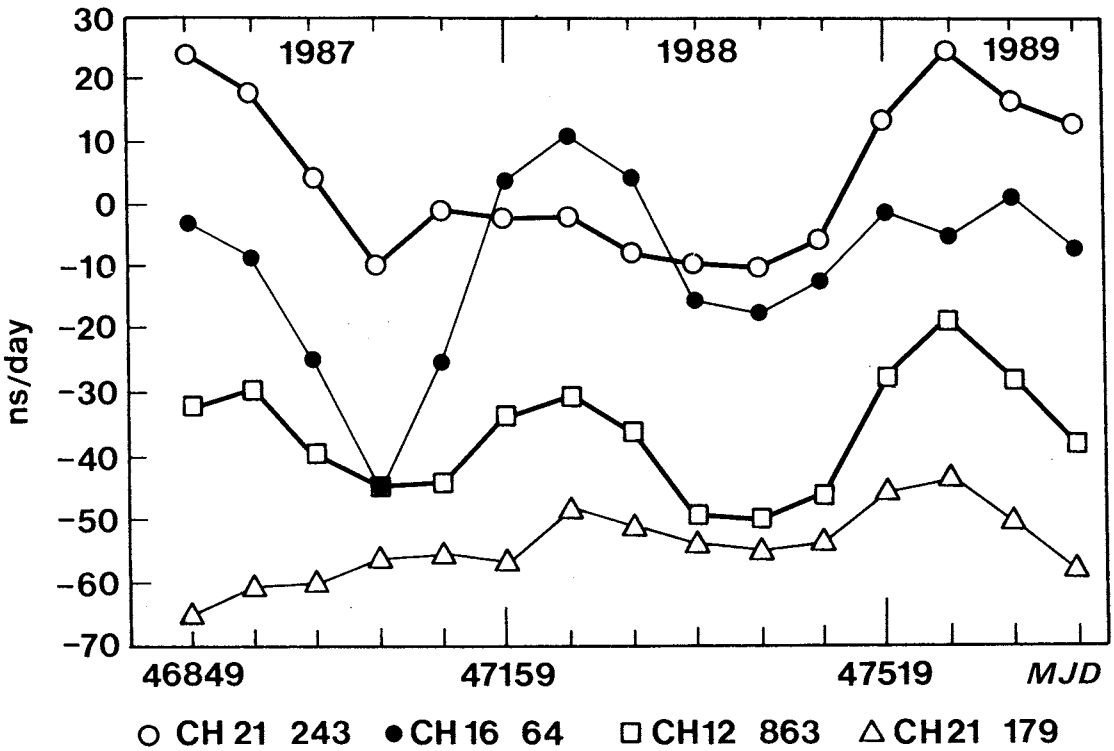


Fig. 9. Frequencies of four Swiss cesium clocks.

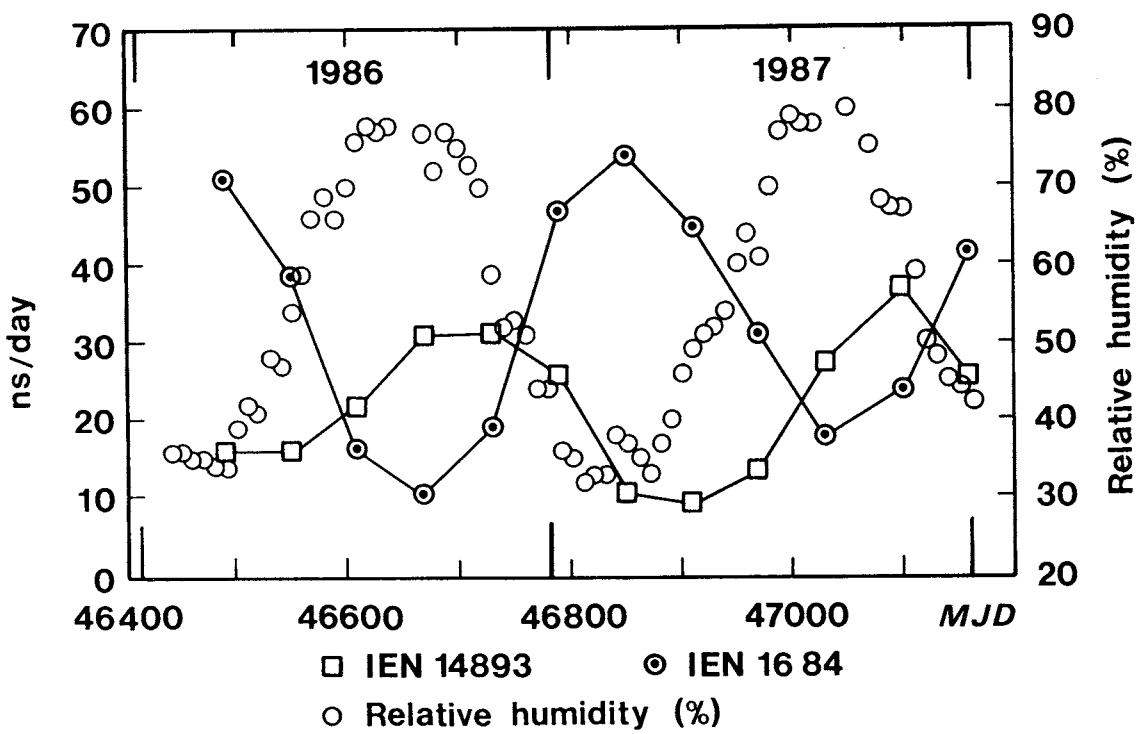


Fig. 10 a. - Frequencies of IEN 14 893 and IEN 16 84 and relative humidity of clock room.

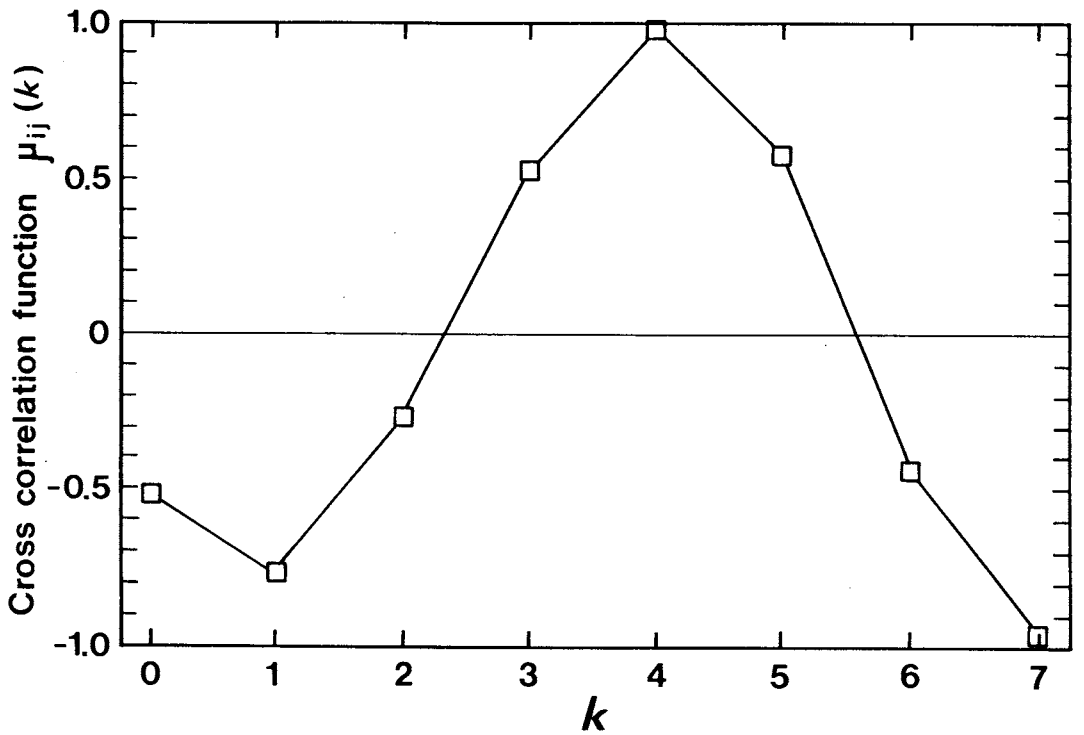


Fig. 10 b. - Cross correlation function $\mu_{ij}(k)$ (as defined in Section 1).

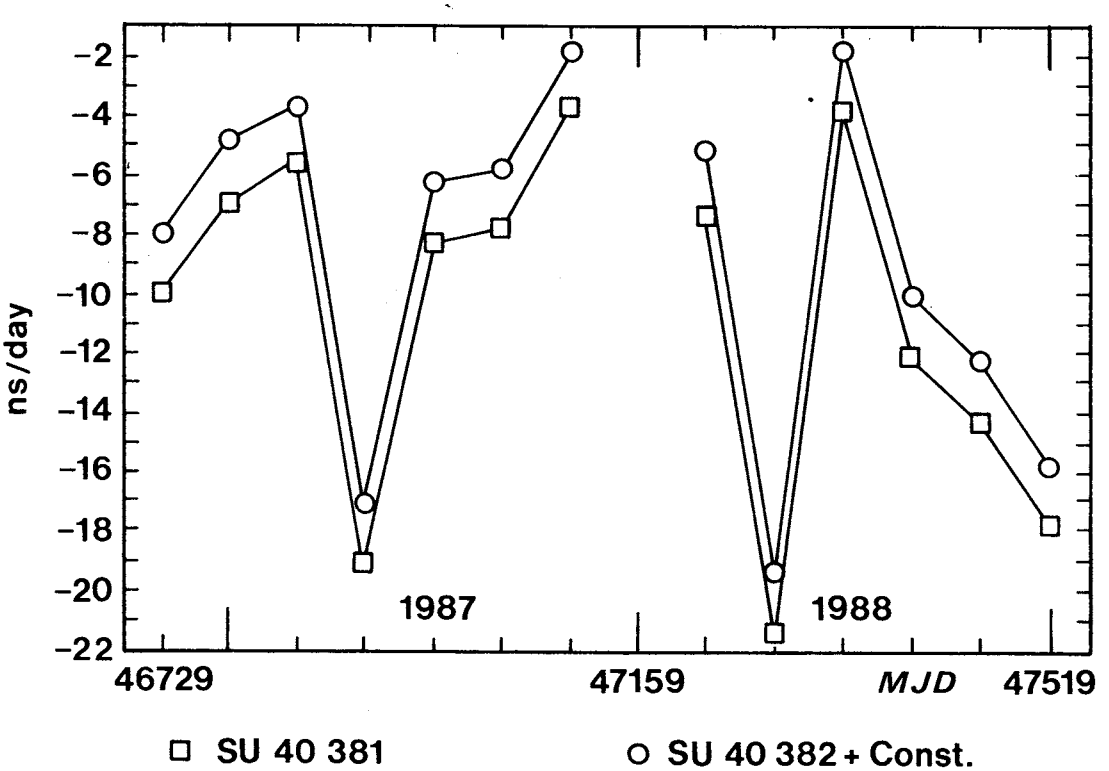


Fig. 11. Frequencies of SU H masers 40 381 and 40 382 (shifted of 2 ns/day).

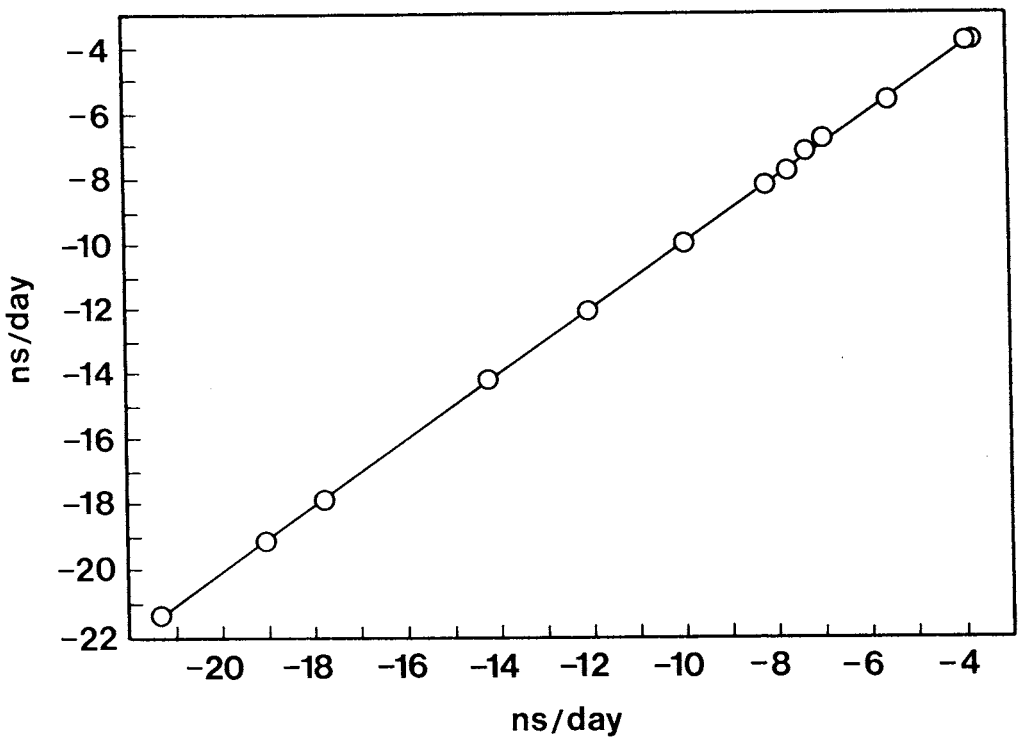


Fig. 12. Frequency of SU H maser 40 382 versus frequency of H maser 40 381.

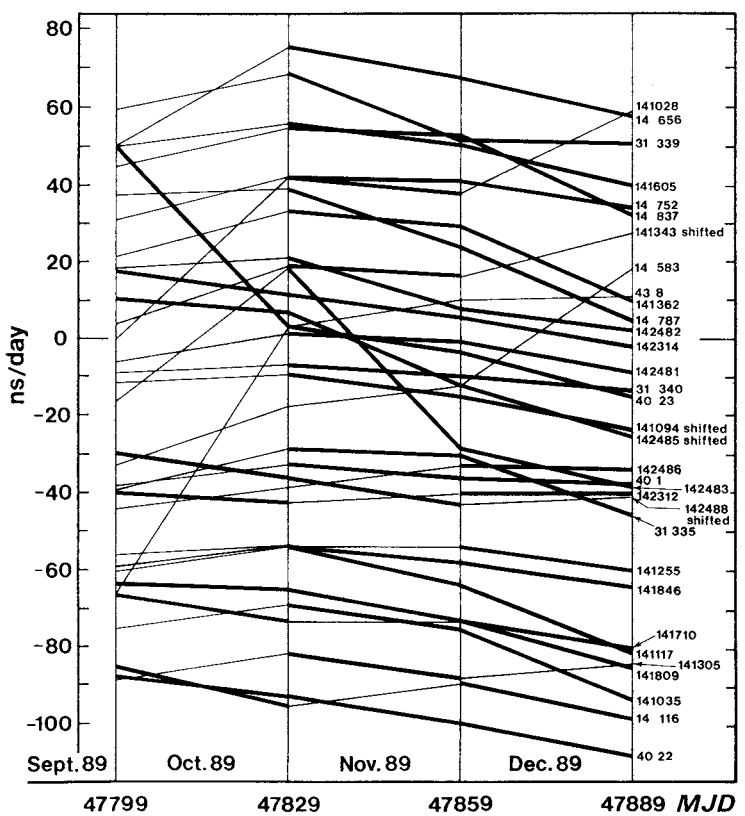


Fig. 13. Monthly frequencies of USNO clocks (from Sept. 89 to Dec. 89)
——— increasing frequencies
▬▬▬ decreasing frequencies.

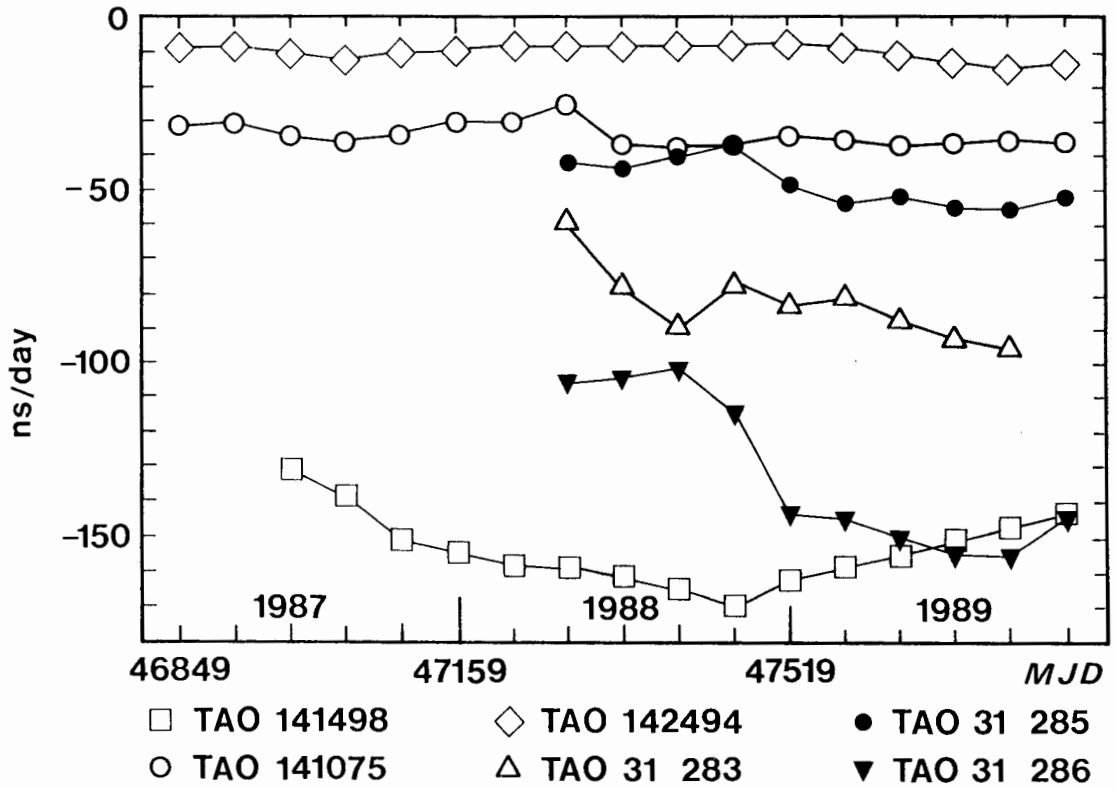


Fig. 14 a. - Frequencies of six TAO cesium clocks.

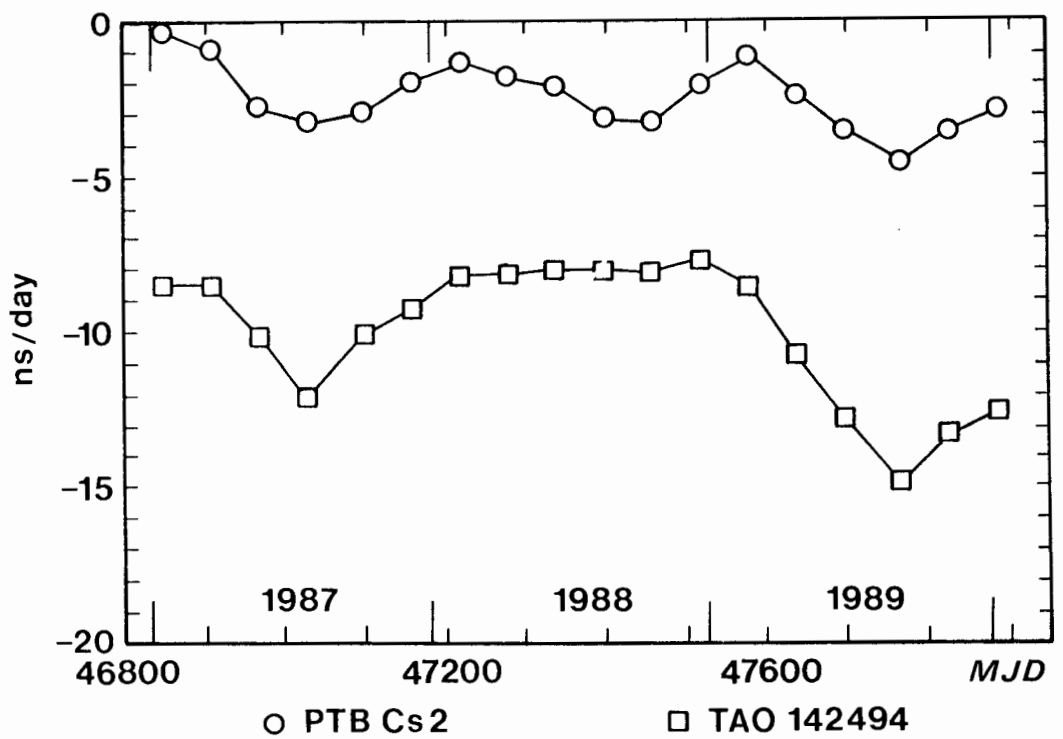


Fig. 14 b. - Frequencies of TAO 142494 and PTB Cs2.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-05 Number 4

DIGITAL SYNCHRONOUS DETECTOR AND FREQUENCY CONTROL LOOP FOR CESIUM BEAM FREQUENCY STANDARD

Lezin TIVON SING, Jacques VIENNET and Claude AUDOIN

Laboratoire de l'Horloge Atomique, Unité Propre du CNRS associée à l'Université Paris Sud,
Bâtiment 221, Université Paris Sud, 91405 Orsay, France

INTRODUCTION

The optically pumped cesium beam tube shows a higher signal to noise ratio and thus a better short term frequency stability than the cesium beam tube with magnet state selectors, as well as promises for an improved long term frequency stability [1,2,3]. In this perspective, it is of prime importance to minimize the sources of frequency offset due to the electronic part of the frequency standard. As long as the processing of both the beam tube response to modulation and of the error signal are concerned, this leads to the choice of digital techniques.

In recently developed cesium beam frequency standards, digital techniques were previously implemented either to control the frequency of the VCXO [4] or to manage the sequence of various tests and commands made [5].The digital set-up described here carries out the synchronous detection and the frequency control of the quartz oscillator.

The digital techniques are well suited to prevent drawbacks inherent to previously used analog methods. For instance, they are able to avoid the occurrence of frequency offsets arising from a possible distortion of the demodulation waveform and from an offset voltage and a finite gain of the operational amplifiers used [6] . Additionnal interests of digital techniques are, for instance, their ability i) to adapt the frequency control loop transfer function to various environmental and operating conditions, ii) to perform other functions such as a measurement and an active control of crucial parameters such as the C-field and the amplitude of the microwave field and iii) to monitor a number of operating parameters.

The digital set-up described here has been designed for future association with optically pumped cesium beam tubes. As a first step of the work, the set-up has been substituted for the analog demodulation and for the analog frequency control loop of a high quality Hewlett Packard 5061 A option 4 cesium beam frequency standard [7] and we have verified that the frequency stability achieved is not modified by the substitution.

EXPERIMENTAL SET-UP

A block diagram representation of the experimental set-up is given in figure 1. The resonance of the Hewlett-Packard 5061A option 4 cesium beam tube at 9192.6 MHz is probed by a microwave signal whose frequency is sine

wave modulated at 137 Hz. The beam tube response to modulation contains a component at the modulation fundamental frequency when the microwave interrogation frequency is not exactly centered at the cesium beam resonance frequency. This component is separated from the other spectral content of the resonator response and its signal to noise ratio is enhanced using a selective analog filter (Datel FLJ-UR 2BA1) having a bandwidth approximately equal to 15 Hz around 137 Hz. Furthermore this filter prevents the occurence of aliasing . Then, the signal is analog to digital converted in a 12 bits ADC (Hybrid System HS 9414), which includes an amplifier whose gain is programmable by softwave, and a sample and hold circuit. The amplifier gain, and therefore the quantization step, depend on the sampled amplitude of the fundamental component of the beam tube response, via control by the microprocessor used. This provides a dynamic range of 126 dB with four possible values of the gain. Saturation of the ADC is then avoided when the frequency of the quartz crystal oscillator is far from its slaved value, whereas the frequency resolution is adequate in steady state when the quartz is locked at his nominal frequency. Moreover, the loop bandwidth is increased during the quartz warm-up time or after a large perturbation of its frequency. This enlarges the practical VCXO frequency acquisition range and speeds the recovery to the steady state regime.

A TMS 32010 16/32 bits digital signal processor [8], clocked at 20 MHz and tied to a 4 K EPROM memory (made of two 27 CX 641 circuits) accomplishes the signal demodulation and executes the algorithm which provides the required loop transfer function. Finally, a 16 bits digital to analog converter (Analog Devices AD 7546), followed by an RC filter, which smoothes the signal, supplies the voltage which controls the frequency of the VCXO. For a DAC ouput change equal to a quantization step, the relative VCXO fractional frequency is shifted by 10^{-14} . Then the capture range of the VCXO is approximatively 6.5 10^{-10}.

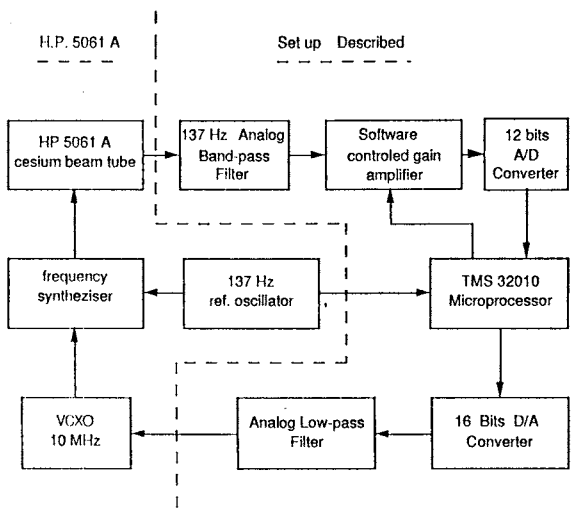


Figure 1 :

Experimental device. The part of the broken line shows the sub-assemblies of the Hewlett-Packard 5061 A option 4 cesium beam frequency standard, which have been used in the work reported. The part to the right of the broken line gives details of the digital set-up which has been substitued for the original H.P. analog demodulator and analog frequency control loop.

SYNCHRONOUS DETECTOR.

An interrupt signal, clocked by the properly delayed 137 Hz reference signal enables the taking of six samples of the filtered beam tube response during each period of modulation, at times $(2\ell+1)T_M/12$, where ℓ is an integer and T_M is the period of modulation (see fig. 2.a). The sampled values are weighted by the value of a sine - function having the period T_M, at times $(2\ell+1)T_M/12$. This means that samples at times $T_M/12$ and 5 $T_M/12$ are multiplied by the factor 0.5. Those at times 7 $T_M/12$ and 11 $T_M/12$ are multiplied by the factor - 0.5. The sample at time 3 $T_M/12$ is kept unchanged and the sign of that at time 9 $T_M/12$ is reversed. Because only division by two and sign change are necessary, the sinusoïdal demodulation is very easy to implement.

A series of samples results, such as shown in fig. 2.b in the case where an idealized noiseless filtered beam tube response is assumed . It is well know that synchronous detection eliminates even harmonics of the input signal. Moreover, it can be shown that the weighting made rejects perfectly the harmonics of rank 3,9,15, etc... of the demodulation input signal. This eliminates greatly the spurious effect of the quantification noise which is present in a large bandwidth, even in the case where the beam tube response is narrow band amplified at frequency $\omega_M=2\pi/T_M$. Furthermore, in the presence of noise it favors samples taken at times where the signal to noise ratio is the largest. The six weighted values which are obtained during one period of modulation are added at times $11n\ T_M/12$, where n is an integer. The sum, called a_n in the following, is the output of the synchronous detector.

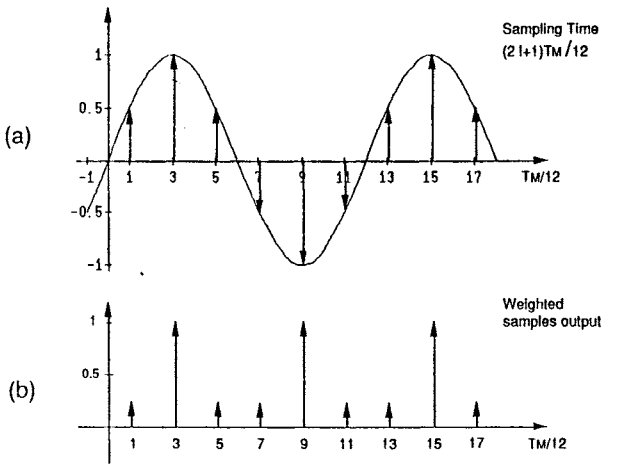


Figure 2 :

a) Sampling of the idealiszed noiseless filtered beam tube response at times $(2l+1)T_M/12$.

b) Weighted sampled values.

DIGITAL FILTER CONTROL LOOP.

In the present configuration, the open loop digital transfer function simulates the analog function:

$$F(p) = \frac{1}{p}\,\frac{1+\tau_1 p}{\tau_2 p}$$

which leads to a second order frequency control loop [6]. The Z transform of the corresponding digital filter which presents the same impulse response is :

$$F(z) = \frac{\tau_1}{\tau_2}\left[\frac{1}{1-z^{-1}}\ \frac{1+(T_m/\tau_1-1)}{(1-z^{-1})}\right]$$

Finally, the digital filtering is obtained by programming the two following recursive equations :

$$b_n = b_{n-1} + a_n$$
$$c_n = c_{n-1} + \left[b_n + (T_m/\tau_1-1)\, b_{n-1}\right] \tau_1/\ \tau_2$$

where c_n is the digital value of the filter output.

It is worth noting that the longest calculation made by the microcomputer occurs at times 11n $T_M/12$. It lasts less than 50 µs, which is small compared to the sampling period, equal to 1.2 ms in the present case. The device is thus able to deal with a larger modulation frequency, to compute a more sophisticated transfer function and to accomplish additional tasks.

EXPERIMENTAL RESULTS

The frequency stability of the VCXO slaved to the cesium beam tube has been measured with respect to a 5 MHz VCXO phase locked to a hydrogen maser [9]. Two sets of data have been obtained, using either the original HP synchronous detector and analog frequency control loop or the digital device with a noise bandwidth of about 0.5 Hz . The two-sample standard deviation $S_y(\tau)$ [10] obtained is given in Table 1 for measuring times from 1 s to 10^4 s. It can be seen that the frequency stabilities achieved are closely identical in both cases.

Time [S]	$S_y(\tau)$ Original device	$S_y(\tau)$ With digital processing
1	$3.8\ 10^{-12}$	$3.9\ 10^{-12}$
10	$1.4\ 10^{-12}$	$1.4\ 10^{-12}$
100	$5.4\ 10^{-13}$	$5.4\ 10^{-13}$
1000	$1.3\ 10^{-13}$	$1.6\ 10^{-13}$
10000	$7.5\ 10^{-14}$	$7.0\ 10^{-14}$

TABLE 1

Measured frequency stability of the cesium beam frequency standard equipped with either its original analog synchronous detector and analog frequency control loop or the set-up described. Each data given is the mean of 30 to 60 measurement samples.

CONCLUSION

A digital synchronous detector and frequency control loop based on a TMS 32010 signal microprocessor has been realized and associated with a high quality commercially available cesium beam tube. A first verification of the properties of the digital device has been made by showing that it enables the achievement of the same level of frequency stability as its manufactured analog counterpart. It is thus expected that the device described is well suited for use with an optically pumped cesium beam tube whose intrinsic frequency stability approaches $10^{-12}/\sqrt{\tau}$. Finally, it is worth noting that the frequency stability obtained is the best reported up to now for a cesium beam tube associated with a digital synchronous detector and a digital frequency control loop.

ACKNOWLEDGEMENTS

The authors are indebted to R. Barillet for its participation to the frequency stability measurement and to DRET for its support.

REFERENCES

1. G. THEOBALD, V. GIORDANO, M. DE LABACHELERIE, A. HAMEL, N. DIMARCQ, P. CEREZ and C. AUDOIN, "Research on the optically pumped cesium beam frequency standards", in Proc. 42nd Annual Symposium on Frequency Control, pp. 496-504, Baltimore, USA, 1988, available from Institute of Electrical and Electronics Engineers, Piscataway, N.J.

2. V. CANDELIER, V. GIORDANO, A. HAMEL, G. THEOBALD, P. CEREZ and C. AUDOIN, "Frequency stability of an optically pumped cesium beam frequency standards" Applied Physics B B49, pp 365-370 1989

3 V. GIORDANO, A. HAMEL, P. PETIT, G. THEOBALD, N. DIMARCQ, P. CEREZ and C. AUDOIN, "New design for an efficient optically pumped cesium beam tube", in Proc. 43rd Annual Symposium on Frequency Control, Denver, pp. 130- 134 USA, 1989, available from Institute of Electrical and Electronics Engineers, Piscataway, N.J.

4. R.M. GARVEY, "A cesium beam frequency standard with microprocessor control", in Proc. 36th Annual Symposium on Frequency Control, pp. 236-239,1982, Philadelphia, USA, available from National Technical Information Service, Springfield VA.

5 J. RABIAN and P. ROCHAT, "Full digital-processing in a new commercial cesium standard", in Proc. 2nd European Frequency and Time Forum, pp. 461-468, 1988, Neuchatel, Switzerland.

6 J. VANIER and C. AUDOIN, The quantum Physics of Atomic Frequency Standards, Bristol UK, Adam Hilger, ch. 5, pp. 747-761, 1989.

7. R.C. HYATT, L.F. MUELLER and T.N. OSTERDOCK, "A high-performance beam tube for cesium beam frequency standards", Hewlett-Packard Journal, vol. 25, pp. 14-24, sept.1973.

8. TMS 320 Product Description. Texas Instruments, 1983.

9. R. BARILLET, P. PETIT, J. VIENNET and C. AUDOIN, "Active H-masers for VLBI Applications", in Proc. 40th Annual Symposium on Frequency Control, pp.410-412, Philadelphia, USA, 1986, available from Institute of Electrical and Electronics Engineer, Piscataway, N.J.

10. J.A. BARNES et al, "Characterization of frequency stability", IEEE Trans. Instrum. Meas. vol. IM-20, pp. 105-120, May 1971.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-05 Number 5

PREVENTIVE MAINTENANCE AND STABILITY OF COMMERCIAL CESIUM CLOCKS

G. FREON

Laboratoire Primaire du Temps et des Fréquences, OBSERVATOIRE DE PARIS
61 avenue de l'Observatoire, 75014 Paris, FRANCE

The use of cesium frequency standards to build time scales of quality requires that these instruments be continuously running, without any interruption in the best environmental conditions and the periodical maintenance operations recommended by the manufacturer be done.

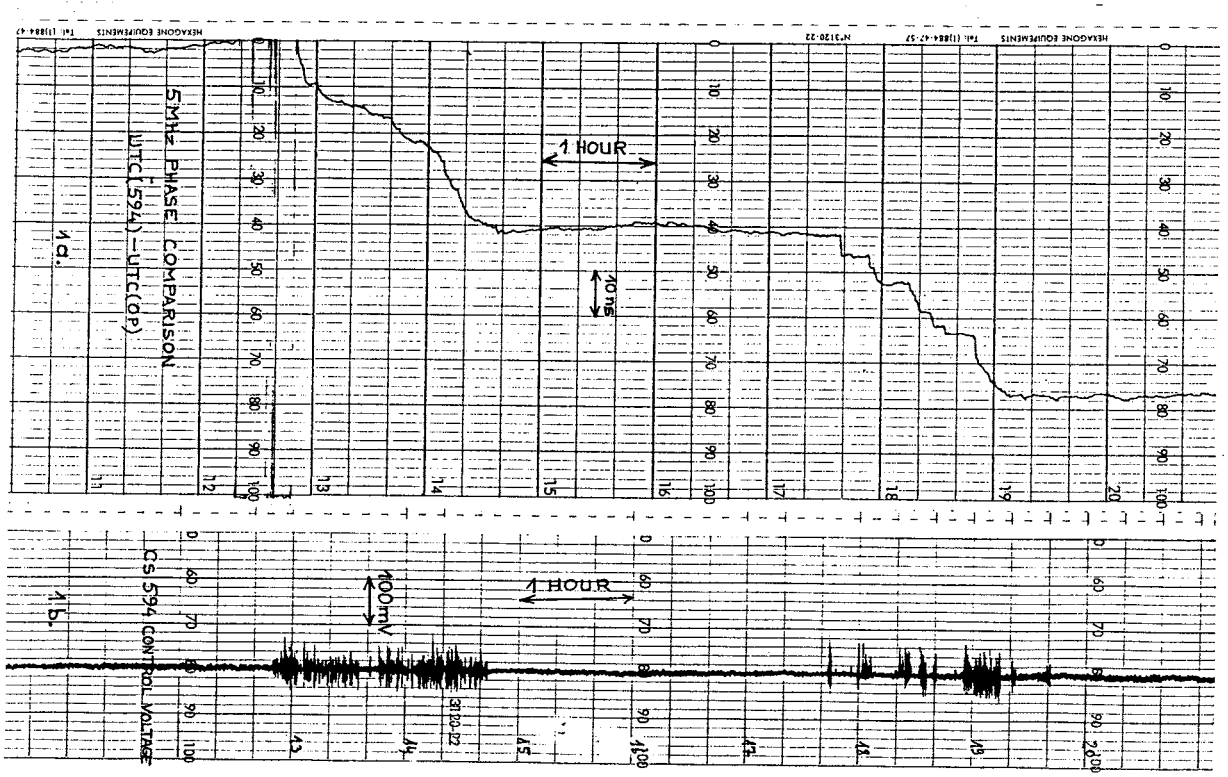
In addition to the frequency instabilities related with the microwave power variations and the C-field setting (1), which are not discussed here, it was observed on LPTF cesium clocks some variations or jumps of the frequency and/or the phase.

Fig. 1a, which is a record of phase comparison of the 5 MHz between two HP5061A option 04 cesium, shows these effects.

I - ENVIRONMENTAL CONDITIONS

At LPTF cesium clocks have been installed in the basement where the temperature is maintained at 22°C ± 0.1°C by an air-conditionning system. The ambiant air is mashed at the rate of 1600 cubic-meters per hour with addition of about 10% of new air taken outside near the ground, so the air is entirely renewed twice an hour. The addition of new air is required by health legislation though nobody works normaly in the room. Humidity is not monitored, temperature is continuously recorded.

It was recorded on one of our standards, namely cesium 594, some jumps of phase such as those of Fig. 1a, although no alarm was set on by the logic of the instrument. The recording of the CONTROL voltage of the crystal oscillator in servo-loop has showed in addition to the normal noise some bursts of great amplitude (Fig. 1b). In



presence of such anomalies any part of the electronics in the servo-loop can be suspected including cables, connectors, solders and power-supply. But in this specific case, the twelve years old crystal oscillator was accused and replaced, and the cesium clock operated properly. Four or five months later, the same symptomes appeared again. Then it was decided to replace one by one all the electronical modules in the servo-loop, which were taken from a dead tube cesium, after to have disconnected all that it was not require to operate. We tried carefully to confirm the result by re-installing each original module. Unfortunately results were not decisive since the noise bursts occured as well as original modules than replacement ones, or never occured, until the time when putting a screwdriver in the sloted axis of the Cs OFF/LOOP OPEN/OPERATE switch the needle of the CONTROL voltage recorder quickly shaked. But that mean we could, or not, produce a lot of phase jumps by a few nanoseconds or several tens. The switch was cleaned with trichloroethane and cesium properly operated... during some weeks only ! In fact, it always had the same failure with the same origin.

When we dismounted the switch (originally silver-plated) all the contacts were covered by a very hard and brown-black-blue coloured film which normally used solvents did not react with. Sliders showed deep strips revealing corroded copper. The silvered parts were coated by silver sulphide acting as insulator.

Inside a few months, the three other cesium standards equiped by the same model of switch successively presented the same failure. At the same time we discovered that the nickel coated steel screws used to fix up our antennas, on the roof of the building, were strangely corroded. The steel was porous and very soft and crisp at the same time.

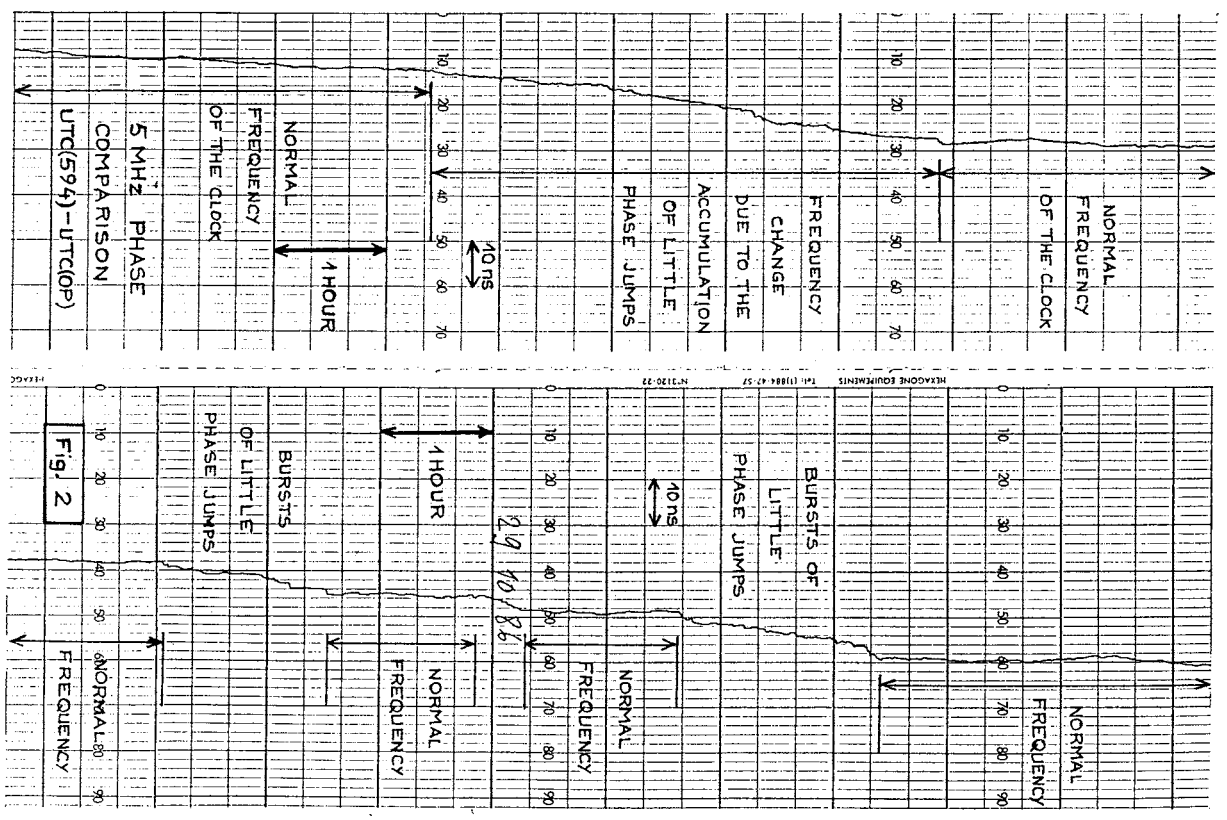
It was suspected that probably aggression came from sulphuretted gases thrown out by the chimney of one building close by the antennas. These heavy chemicals normally fall down, but they were sucked in by the ventilation aperture of the air conditionner of the clock room. Consequently the important rate of air renewal in the room was at the origin of a fast corrosion process.

Finally, we have replaced the silver plated switches of all our cesium clocks by gold-plated ones, we have short-circuited the TIME-CONSTANT switch positions and we have reduced the inlet of new-air in the air-conditionner.

Later an another cesium showed similar phase jumps joined with the extinction of the CONTINUOUS OPERATION light that meant a failure in the synthesizer which was easily reproductible, it was sufficient to slightly put the finger on two of the thumbwheels switch (old model of synthesizer). After opening it was found that the gold-plated sliders were tarnished ; we cleaned them and by security we strapped the appropriated combination.

The same thing was done on other cesium which had the same model of synthesizer. Since the year 1972, these settings have not been changed, that is prejudicial to keep on the quality of the contacts.

After the discovery of the cause of the phase jumps we looked more sharply at the old phase comparison readings -and especially the frequency variations of several 10^{-13} which were observed- and we saw for duration



of several tens of minutes, at intervals from several hours to a few days a lot of succession of very little phase jumps, by one nanosecond or less, such as Fig. 2, which were at first confused with the noise, and the accumulation of which can be interpreted as a change of the frequency of the clock.

II - PROBLEMS IN ELECTRONICS

II-1 - Crystal oscillator

Phase jumps and frequency changes are also observed when the crystal oscillator becomes suddenly unstable at very short term or present a blunt important change of its frequency (more than 10^{-8}). Cesium standards equiped with a single integrator control loop show phase jumps in the order of the nanosecond to some tens of nanoseconds (Fig. 1). Cesium clocks equiped with double integrator control loop do not present any phase jump but a lot of little frequency changes in the order of 10^{-13}. Probably the short term crystal oscillator instabilities also pollute the microwave spectrum near the cesium line.

Display of that kind of anomaly is obtained through the recording of control voltage in the servo-loop during several weeks to verify its stability. The servo-loop control voltage is taken at the "CONTROL" BNC jack at the rear of the instrument, in this way this measurement does not perturb the operation of the instrument. Generally the instabilities are not continuous in the time : it can be seen sudden frequency changes followed by normal operation in a fraction of a second. The changes are completely isolated at intervals of several minutes, hours, days, weeks or months, or on the contrary, some repeated bursts of very fast instabilities. We saw a crystal oscillator having such a failure during three months consecutively and then staying completely quiet during more than six months.

II-2 - Frequency changes in relation with CIRCUIT CHECK switch position

At LPTF we used to verify every week the front panel CIRCUIT CHECK meter indication for each position of the switch. When the standards had the silver-plated switches the checking was always made in the same order, starting from BATTERY and finishing by 100 KHz position on which the switch was left for the next week. Since the installation of gold-plated switches, the contacts of which are naturally fragile, we have got the habit of checking one week starting from BATTERY and the following week starting from the 100 KHz position (last position of the switch). It helps saving the lifetime of the switches. With two cesium clocks it was observed that each clock had two stable frequencies according to the position of the CIRCUIT CHECK switch. The frequency offsets were about 1.5×10^{-13} for one cesium and less than 1×10^{-13} for the other. It is well known by the manufacturer, it seems to be due to the modular designing and the wiring of the instrument ; it is very difficult to cancel out. The easiest mean to release it is, of course, to maintain the switch in the same position.

II-3 - Frequency changes after LOOP GAIN adjustments

According to the manual of the instrument the second harmonic amplitude should be set by adjusting LOOP GAIN control to read 40 on the front panel meter. It can decrease to 25 with no logic alarm.

At LPTF it has been agreed to maintain the second harmonic level between 35 to 45. It has been observed that the cesium standard 347 shows a frequency change of 1 to 2×10^{-13} each time the LOOP GAIN control is adjusted in connection with the second harmonic level (Fig. 3). The manufacturer knows this problem which occurs with certain standards. The cause seems also the modular designing, wiring and unsufficient filtering cautions ; so it is difficult to cure this trouble. It can be tried to reduce the second harmonic range, between 38 to 42 for instance, that needs to adjust LOOP GAIN control more often. It is not sure that the frequency jumps will be completely canceled out because it seems that another condition, which is not identified, is necessary to jump the frequency by adjusting LOOP GAIN.

II-4 - Phase jumps by adjusting the OSCILLATOR FREQUENCY COARSE control

According to the turn-on procedure, after the warm-up period the frequency of the crystal oscillator is adjusted with OSCILLATOR FREQUENCY COARSE control for 0 reading on the front panel meter and the CONTROL position of CHECK SWITCH. Because of the aging of the crystal oscillator this reading is drifting in time, and can reach ± 50 after several months or years. Then the frequency crystal oscillator must be reset before reaching the integrator limit and turning on the ALARM light. After such an adjustment it is observed, if the clock is an old model with a single integrator control loop, a phase jump by about 30 nanoseconds for a change from 50 to 0 of the CONTROL reading, on the other hand the output frequency of the cesium does not change. Such phase jumps must be avoided if cesium is generating a time scale which is designed to be as stable as possible. Consequently care must be taken not to leave the frequency crystal oscillator drifting too much. At LPTF the CONTROL indication is maintained between -2 and +2 on the front panel meter ; adjustments are spaced one to three months apart and phase jumps are only by 2 to 5 nanoseconds maximum.

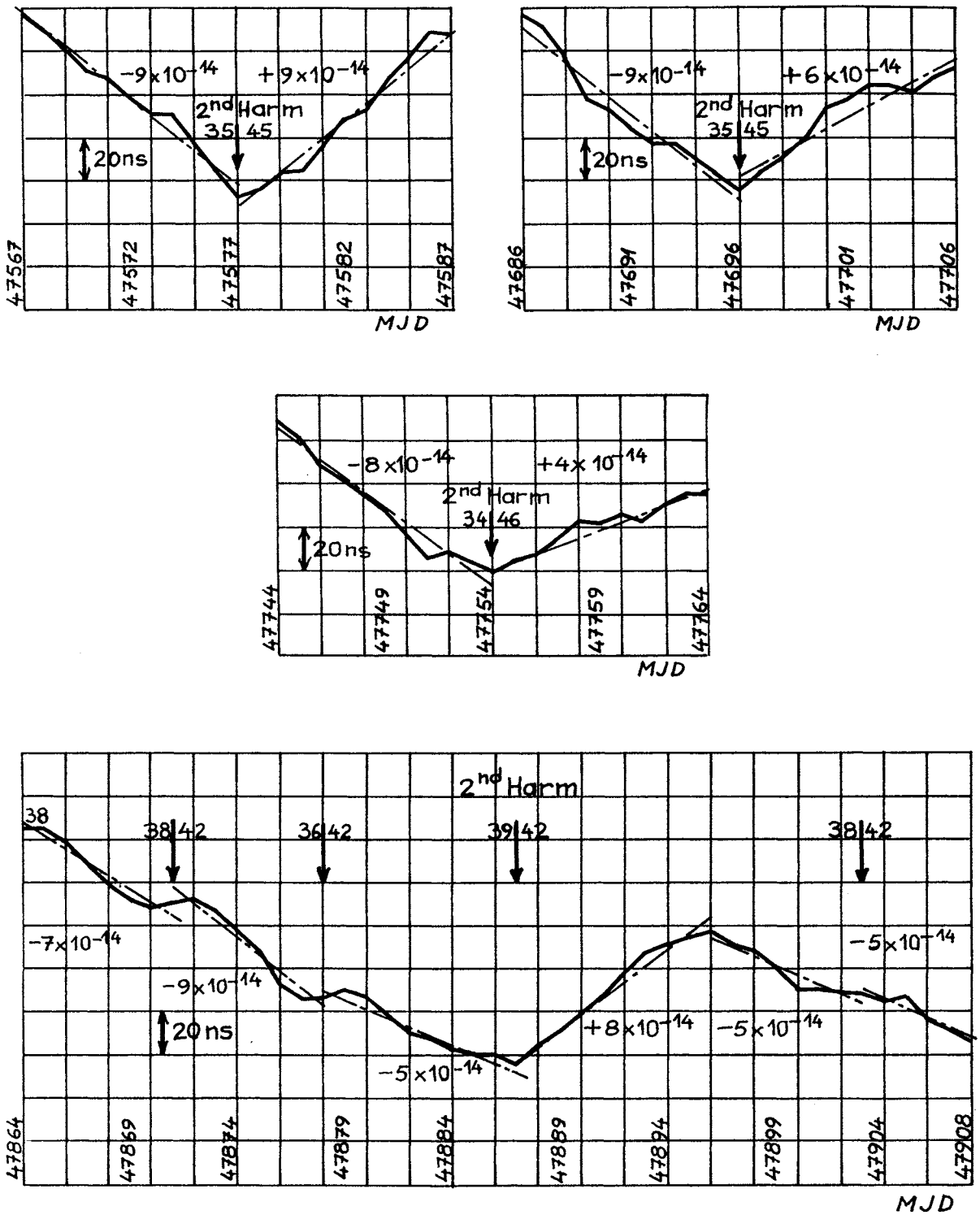


Fig. 3

Time comparisons TA(F)-TA(347) showing frequency changes when adjusting LOOP GAIN control to maintain 2nd harmonic between 35 and 45, and trial to reduce these changes by limitation of the 2nd harmonic range. Dash lines are mean frequencies between adjustments.

For cesium with a double integrator control loop, the crystal oscillator frequency may drift until CONTROL reading reaches ± 50, it should not be seen any phase jump after adjusting OSC. FREQUENCY COARSE control.

III - CONCLUSION

It is clear that apart the problems of atmospheric pollution and short term instabilities of the crystal oscillator, most effects which were identified are negligible for the majority of industrial applications or laboratories which limit the precision of their measurements to the order of 10^{-12}.

Nevertheless, it never been insisted on the need for doing time or frequency comparisons between standards in the same laboratory or with respect an external reference if only one standard is operating. In that last case a good precaution consists in the continuous recording of the control voltage in the servo loop which can reveal some mistakes in electronics before to see undesirable or catastrophic effects on stability or accuracy of the frequency.

It must be taken care of air-conditionning systems which operate with a high rate air renewing and mashing. In industrial or urban sites they can increase chemical agression on the circuitry of the instrument. Following our remarks about silver-plated switches the Hewlett-Packard Company now installs gold-plated switches on their new models of cesium.

For HP5061-A cesium with a single integrator control loop many examples of which are operating yet, the CONTROL front panel meter indication must be maintained close to 0 in the order not to introduce notable phase jump in the time scale generated when adjusting OSCILLATOR FREQUENCY CONTROL.

The front panel CHECK switch will be always maintained on the same position to avoid frequency changes.

Finally, on the occasion of an hard failure which will involve a break in the time scale of the clock it will be recommended to go and back all the switches out of their normal working position ; it is yet the best mean to clean them and keep them in a good condition.

ACKNOWLEDGEMENT

The author wishes to acknowledge Dr. L. Cutler and Mr. F. Lazarus, from Hewlett-Packard, for their regards and helpful discussions to try solve our problems.

Many thanks also to Mrs Monique Prodhomme who did all delicate wiring modifications of LPTF cesium clocks, and to Mrs Catherine Boulland and Mr. Michel Granveaud for their help to prepare this paper.

REFERENCE

(1) A. de Marchi : "New insights into causes and cures of frequency instabilities (drift and long term noise) in cesium beam frequency standards" - 41st Annual Frequency Control Symposium, 1987.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session FS-05 Number 6

An Automated System for Measuring Velocity Distribution in Cesium Beam Tubes

S. K. Karuza, W. A. Johnson, J. P. Hurrell, M. F. Bottjer, and F. J. Voit
Electronics Research Laboratory, The Aerospace Corporation
P. O. Box 92957, Los Angeles, California 90009 U.S.A.

Abstract

This paper describes a completely automated system for measuring the velocity distribution in a cesium (Cs) beam tube. The measurement system uses the short-pulsed microwave excitation method on Cs beam tubes having separated Ramsey-type interaction regions. This pulsed technique allows for the observation of signals that result from very narrow atomic-velocity groups. The brief theoretical operation of this method is discussed, after which the velocity distributions of six Cs beam tubes from four different manufacturers are presented. Data are presented that show the reproducibility of this measurement on the same beam tube, and comparisons among different tubes are made. Also, velocity distributions have been measured on all seven Zeeman transitions from one tube that exhibits small Rabi pulling. The knowledge of this Cs tube's velocity distribution provides considerable insight into the performance and alignment of the beam optics, and thus could be used for the diagnostic testing of Cs beam tubes. We have used this method to characterize a large number of such tubes.

1. Introduction

Since the velocity distribution provides quantitative information about the alignment of the optics in a cesium (Cs) beam tube, and because this alignment in turn affects the final performance of the Cs frequency standard, there is a need to develop a system that measures this velocity distribution accurately. In order to minimize the data-taking time, it was decided to automate the measurement system. The early investigators [1,2,3] developed a method that used the pulsed excitation of atomic beam devices that had Ramsey-type interaction regions; this allowed for the observation of signals that were due to very narrow velocity groups.

Because the duty cycle of the pulsed microwave power is low, the current of the output beam is a small fraction of the current of the CW signal. A typical beam current output for the pulsed condition is about 0.06 nA. Therefore, long integration times are needed to measure the signals used to determine the velocity distributions. Measurements become time-consuming, and the accuracy and calibration of the measurement system itself become critical. Completely automating the measurement process drastically reduces the data-taking time. An additional advantage of automation is that one never has to make and remake the electrical connections.

Figure 1 shows the operation of the Ramsey cavity with pulsed microwave power. The velocity distribution of the beam tube detector cannot be inferred simply from the velocity distribution of the oven, because the beam optics will modify the oven's velocity distribution significantly, as shown in Figure 2. Figure 2 shows a typical beam-tube velocity distribution as measured at the detector, as well as the Maxwell-Boltzmann distribution created by the flux of atoms leaving the Cs oven. The shape of the velocity distribution provides information about the "optical" design and alignment of the tube, namely its reproducibility and its utilization of the Cs flux from the oven. When the velocity distribution is known, it is possible to calculate the second-order Doppler-shift correction for that particular beam tube; it is also possible to determine the cavity phase shift [1,2] from additional experimental data.

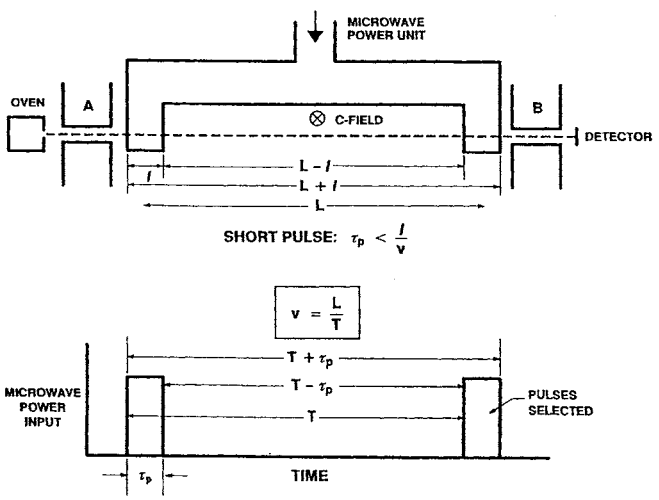


Figure 1. Operation of the Ramsey Cavity with Pulsed Microwave Power. The design of the Ramsey dual-interaction region permits one to select the velocity of the atomic beam by setting the period T of the microwave pulses.

Some of the major factors affecting the width of the Cs tube's velocity distribution are as follows: (1) wider detector ribbons will lead to broader velocity distributions, (2) smaller misalignment of tube components increases the number of fast atoms detected, and (3) higher field state-selecting magnets allow for the improved collection of faster atoms. Changes in the velocity distribution with time may help to determine whether or not unwanted Cs deposits develop on sensitive apertures and thus contribute to background signals that degrade performance. This velocity distribution provides considerable insight into the performance of Cs clocks, because the Cs tube acts as a stable, very high-Q frequency discriminator that controls the long-term performance of the clock. The shape of this discriminator, as determined by the Ramsey resonance curve, is dependent upon the mean velocity and the moments of this velocity distribution.

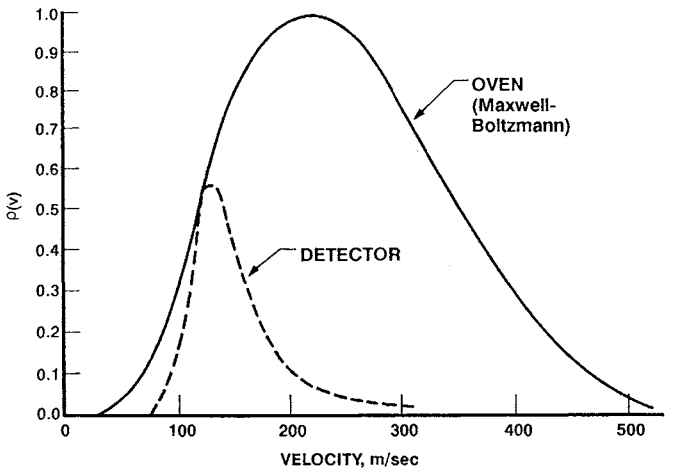


Figure 2. Cesium Beam Velocity Distribution. The Maxwell-Boltzmann velocity distribution out of the Cs oven is shown, along with the modified velocity distribution, selected by the beam optics, as measured at the detector.

A system has been designed and implemented to measure these velocity distributions precisely, using the pulse technique developed by Hellwig et al. [1].

2. Measurement Technique

The interaction between the atomic beam and a pulsed microwave field occurs only for those few Cs atoms that are in the cavity interaction regions when the pulses occur. Since the time when the atoms interact with the microwave field has an uncertainty that is set by the length l of the single cavity region, there exists a velocity window Δv (see Figure 3) that varies with the pulse period T. The smaller the velocity window, the more precise will be the velocity measurement. Since most of the Cs atoms experience the microwave field in only one of the two separated cavity sections, the Ramsey resonance is small and is superimposed on a large Rabi pedestal. Figure 4 shows the spectrum for both CW and pulsed conditions. The widening of the Rabi pedestal is caused by the reduced interaction time τ_p.

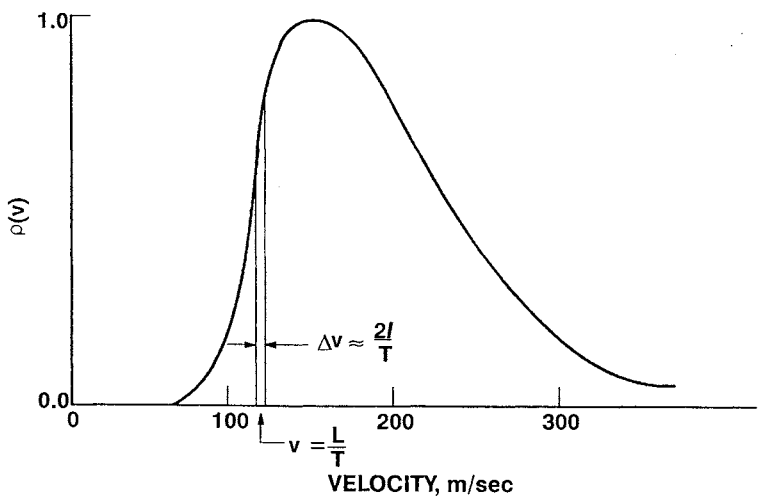


Figure 3. The Velocity Window Δv. The pulsed microwave power technique measures the flux of Cs atoms for a velocity window Δv.

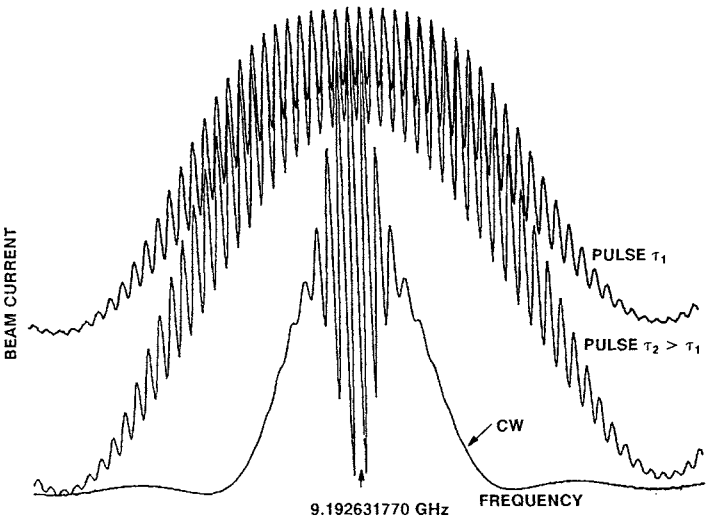


Figure 4. Ramsey Spectrum of the Cs Beam Tube under both Normal CW Operation and the Pulsed Microwave Condition

For a pulse width τ_p and a pulse-repetition period T, there will be a narrow range of atomic velocities for which the atoms will experience resonance in both of the microwave interaction regions. It is clear from Figure 1 that the mean velocity v of these atoms is given by

$$v = \frac{L}{T} \quad . \tag{1}$$

The spectrum of the beam tube shown in Figure 4 is due to this class of atoms. The pulse width τ_p is chosen so that

$$\tau_p < \frac{l}{v} = \frac{l}{L} \times T \quad , \tag{2}$$

as noted in Figure 1.

In general, the fastest atoms in this ensemble will travel a distance $L+l$ in a time of approximately $T - \tau_p$; i.e., they will enter the left side of the left interaction region at the trailing edge of the first pulse and will leave the right side of the right interaction region at the leading edge of the next pulse. Then, if we call this fastest velocity v_f, we obtain

$$v_f(T - \tau_p) = L + l \quad . \tag{3}$$

Similarly, the slowest atoms that will also experience resonance in both of the microwave regions will leave the right side of the left interaction region at the leading edge of the first pulse and will enter the left side of the right interaction region at the trailing edge of the second pulse. Then, if we call this slowest velocity v_s, we obtain

$$v_s(T + \tau_p) = L - l \quad . \tag{4}$$

The range of velocities Δv for which there is interaction in both regions is then

$$\Delta v = v_f - v_s = \frac{L + l}{T - \tau_p} - \frac{L - l}{T + \tau_p} \approx \frac{2l}{T}\left[1 + \left(\frac{L}{l}\right)\left(\frac{\tau_p}{T}\right)\right] \quad . \tag{5}$$

Then, since $l < L$ and $\tau_p \ll T$ for short-pulse conditions,

$$\Delta v \approx \frac{2l}{T} \quad . \tag{6}$$

The values of L and l for the beam tubes investigated were about 13 cm and 1 cm, respectively. Then, for example, a pulse period of 1300 μsec would select atoms with a mean velocity v = 100 m/sec and a velocity spread Δv = 15.4 m/sec. The measurement procedure for this short-pulse experiment is as follows:

1. Choose a pulse period commensurate with the range of velocities that are expected.
2. Choose a pulse width less than l/v. As indicated, pulse widths of 20 to 40 μsec were used in our experiments.
3. Sweep the microwave frequency through the center of the Rabi pedestal and over a frequency range equal to several inverse periods $1/T$.
4. Change the pulse period and repeat the sweep with the pulse width kept constant.

Figure 5 shows the results for three pulse periods. For our computations we used the peak-to-valley amplitude of each sweep's Fourier component at the frequency $1/T$. For the 1300-μsec spectrum, the peak-to-valley Ramsey-resonance current component $I_p - I_v$ (hereafter known as I_{pv}) is very nearly equal to $I_p - I_v$ as marked on Figure 5. However, for the 722-μsec data, I_{pv} at 1/722 μsec is very small, because the spacing of the maxima is close to 1/(2 × 722 μsec). The final data set consists of pairs of pulse periods T (with an associated velocity $v \equiv L/T$) and peak-to-valley currents I_{pv}.

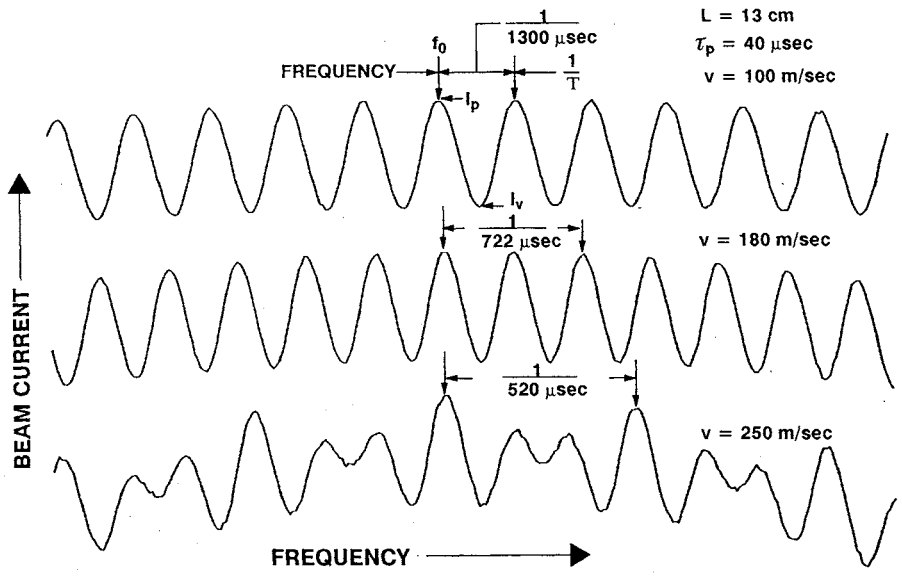


Figure 5. Portion of the Ramsey Spectrum Selected by the Pulsed Microwave Condition. The Ramsey pattern for the three different pulse periods corresponds to three different velocity groups. The fundamental Ramsey frequency is given by 1/T.

During the CW operation of the tube, the number of atoms detected near resonance is a convolution over all velocity components and is given by

$$S(\omega - \omega_o) = \int N(v) \sin^2 2b\tau \cos^2 1/2(\omega - \omega_o)T dv \quad , \tag{7}$$

where
$(\omega - \omega_o)$ is the frequency offset from ω_o;
τ represents the transit time through one cavity ($\tau = l/v$);
T is the time of flight between the cavities ($T = L/v$);
b is proportional to the microwave field in the cavities and represents the interaction between that field and the Cs atom (b is a constant near resonance);
$N(v) \equiv$ the velocity distribution function; and
$N(v)dv = dN$, the number of detected atoms per second moving with a velocity between v and $v + dv$.

Under the short-pulse conditions, τ is replaced by the fixed pulse width τ_p, with $2b\tau = \pi/2$ for maximum signal. The number of atoms captured in a single pulse period is equal to dN times the length of time it takes the atoms to pass through the interaction region $= dNl/v = N(v)dvl/v$. Because the window width is $2l/T$ and $v = L/T$, we finally obtain the number of atoms captured in a single pulse period, which is $N(v)2l^2/L$. The capture is repeated at every interval T, and leads to an average signal I_{pv} that is given by

$$I_{pv} = \frac{1}{T} N(v) \frac{2l^2}{L} \quad . \tag{8}$$

Hence

$$N(v) = \frac{L}{2l^2} T I_{pv} \quad . \tag{9}$$

In the long-pulse approximation in which $\tau_p \gg l/v$, the number of atoms captured per pulse is $\tau_p dN$. Consequently,

$$I_{pv} \approx \left(\frac{\tau_p}{T}\right) dN, \text{ where } dN = N(v) \times \frac{2l}{T} \left[1 + \left(\frac{L}{l}\right)\left(\frac{\tau_p}{T}\right)\right] \quad . \tag{10}$$

The normalized velocity distribution $\rho(v)$ for a tube is defined to be $N(v)/N_{max}(v)$ and can be derived from the short-pulse data according to Eq. (11):

$$\rho(v) = \frac{T I_{pv}(T)}{[T I_{pv}(T)]_{max}} \tag{11}$$

Figure 6 shows the transition from short-pulse to long-pulse operation. The peak signal (about 0.14 nA) evident in the short-pulse regime occurs when $2b\tau_p = \pi/2$, where b is proportional to the square root of microwave power. The long-pulse regime occurs when $\tau_p > l/v$ (≈ 100 μsec) and leads to a signal proportional to $\tau_p \left[1 + \left(\frac{L}{l}\right)\left(\frac{\tau_p}{T}\right)\right]$ [see Eq. (10)]. When the velocity window exceeds the width of the velocity distribution itself, however, the signal becomes proportional to τ_p and tends to the normal CW value.

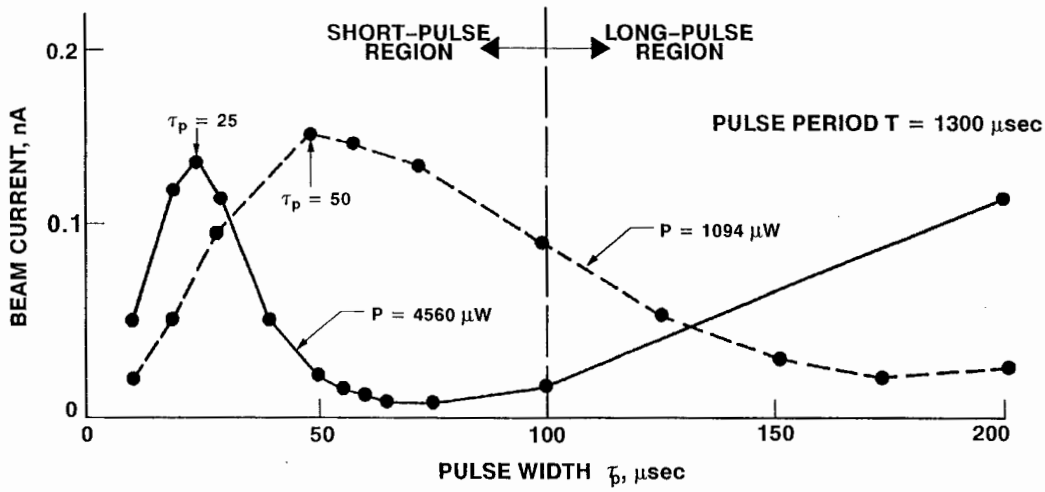


Figure 6. A Plot of the Output of the Cs Beam Tube as a Function of the Pulse Width τ_p for Two Different Microwave Powers

To determine the average velocity from the velocity distribution curve, we calculate as follows:

$$\text{average velocity} = \bar{v} = \frac{\int_0^\infty v\varrho(v)dv}{\int_0^\infty \varrho(v)dv} \tag{12}$$

Figure 7 shows the block diagram of our pulsed microwave measurement system, interfaced to a Hewlett-Packard model 9825 digital computer that is used as a controller and data-acquisition system. The microwave sweep and the setting of the pulse period are both computer controlled. All of the data necessary to compute the velocity distribution function for a tube are taken automatically and stored on disk. The final data reduction is accomplished by transferring the data to an IBM model 3090 computer.

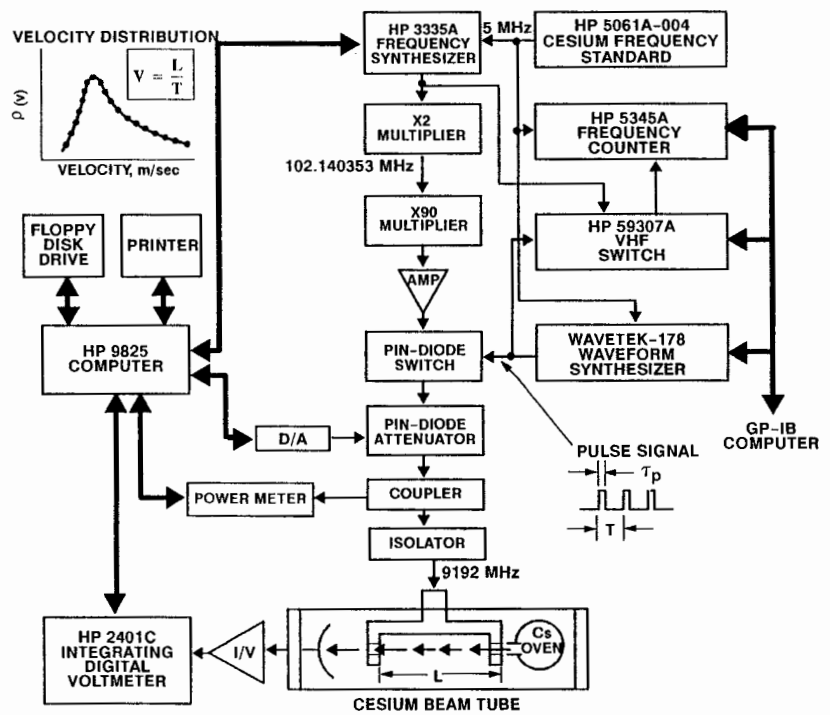


Figure 7. Block Diagram of the System for Measuring the Velocity Distribution of the Cs Beam Tube under Pulsed Microwave Conditions

The peak-to-valley currents that are measured in the pulsed microwave tests are much smaller than the I_{pv} for the CW signal, for two reasons. First, there is a low duty cycle during pulsed operation; second, the measured currents result from only those Cs ions that are within the selected velocity window. These small signals require long integration times to generate acceptable signal-to-noise ratios. Thus, a velocity window of ten percent of the most probable velocity will lead to a signal current of approximately one percent

of the CW signal. For a particular Cs beam tube (in this case tube 1), we can see from Figure 6 that when we pulse the microwaves with a pulse width τ_p of 25 μsec and a peak microwave power of about 4560 μW, we then obtain a maximum signal current of 0.14 nA. For each data point the integration time was about two seconds. In this case the CW peak beam current was about 40 nA. For tubes with much lower CW peak currents, the integration time could be as long as five seconds per data point.

For a typical measurement that spans a velocity range of 70 to 240 m/sec, 80 data points per curve are taken, with 101 frequency samples per data point. An integration time of two seconds is used for each data point and a one-second counter-gate time for each frequency measurement. To collect all the raw data required for the velocity distribution curve takes about 6.7 hours; the data are then processed by the IBM 3090 computer and are plotted.

3. Results of the Velocity Distribution Measurements

Velocity distribution measurements were performed on six Cs beam tubes from four different manufacturers. Multiple measurements were made on some tubes to determine the reproducibility of the measurements.

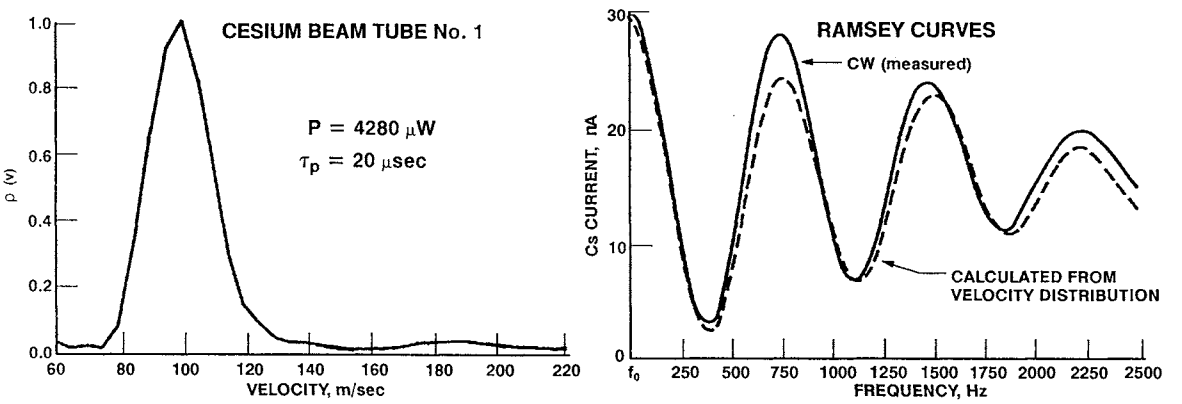


Figure 8. Comparison of the Experimentally Measured Ramsey Pattern under Normal CW Operation with a Pattern Calculated Using the Measured Velocity Distribution of Cs Beam Tube No. 1 (left)

To verify that the velocity distribution that was measured was correct, we calculated the Ramsey resonance curve for tube 1 by using Eq. (7) and the measured velocity distribution; then we compared this calculated curve to the measured Ramsey resonance curve for the tube. Figure 8 shows the measured velocity curve and a comparison between the calculated and measured Ramsey resonance curves. In general, the two Ramsey curves are in close agreement.

Figure 9 shows the plots of the velocity distributions of four other Cs beam tubes. Notice that some of these velocity distribution widths (for tubes 1 and 2) are much narrower than they are for others, indicating that there is a substantial amount of velocity rejection by the beam optics. The wider velocity distributions have Ramsey resonance curves that are broader and more

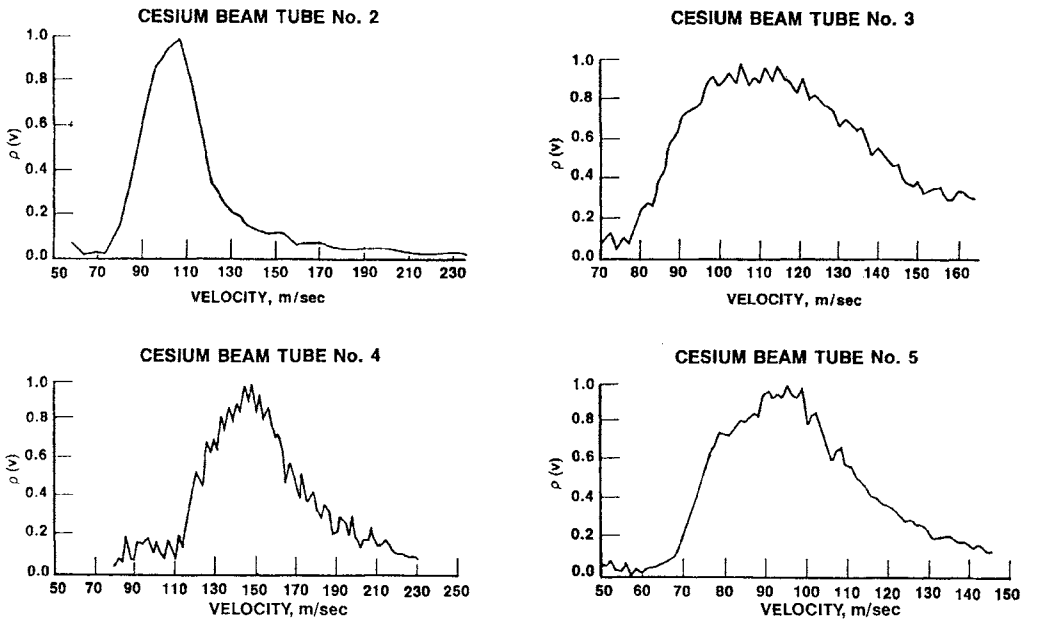


Figure 9. Measured Velocity Distributions for Four Cs Beam Tubes from Four Different Manufacturers

damped. The width of the Ramsey discriminator decreases as the average velocity decreases, and for a fixed average velocity the Ramsey signal increases as the width of the velocity curve increases. Both effects improve the frequency stability of the clock. Note that in some velocity distribution data (e.g. for tube 4 in Figure 9) there appears to be a fine structure in the data. This structure is not really in the velocity distribution, but is caused by contributions from nearby Zeeman transitions [1].

Table 1 summarizes the velocity measurement parameters for five Cs beam tubes. Tubes 1 and 2 are from the same manufacturer and demonstrate the effect of manufacturing tolerances on velocity distributions. The other tubes are from different

manufacturers and show the variability of short tube-velocity distributions among different designs. Tubes 1 and 2 were measured more than once, to show the reproducibility of the measurement. The time interval between these measurements was about three days. The table shows the average velocity $\bar{v}$ and the 50% velocity distribution width $\Delta v_{50\%}$; these are tabulated along with estimates of the average velocity from the measured CW Ramsey resonance curve. The CW Ramsey velocity estimate is calculated by averaging the frequency separations of the Ramsey peaks and then multiplying this average frequency separation by the cavity length L to obtain the estimated average velocity for the Cs beam. The results show that the two different methods give answers that are in close agreement. Other parameters are also tabulated, such as the measurement pulse width τ_p, the microwave power, and so on.

Table 1. Measurement Data for Five Cesium Beam Tubes

TUBE No.	CW DATA		PULSED DATA					
	I_p, nA	I_v, nA	τ_p, sec	P_{PEAK}, μW	$\bar{v}$, m/sec	$\Delta v_{50\%}$, m/sec	$\bar{v}_{EST}$, m/sec	v_{PEAK}, m/sec
	40	3.2	40	1000	103.5	23	102	99.9
1	40.0	3.2	20	1000	103.1	23	102	99.9
	40.0	3.2	20	4280	108.6	24	102	99.9
2	4.6	1.6	20	1010	116.1	29	113	107.5
	4.6	1.6	20	4050	115.6	30	113	107.5
3	15.7	3.5	35	500	119.0	54	107	108.9
4	5.7	2.1	40	2000	153.0	46	144	149.7
5	207.6	47.9	20	4000	98.4	37	92	95.7

In order to investigate the origin of the Rabi pulling in a sixth Cs beam tube, we measured the individual velocity distributions for each of the Ramsey curves of the seven Cs transitions. This Cs beam tube exhibited Ramsey curves with high signal-to-noise ratios for all of the seven transitions, as shown in Figure 10. Ramsey curves are usually highly attenuated on the outer Zeeman transitions, because C-field inhomogeneities broaden the spectrum. Consequently, this particular tube has a very uniform C-field and permits the measurement to be made.

Figure 11 shows the plots of all seven velocity distributions. The velocity distribution of the main Ramsey transition is shown by itself, but the velocity distributions of the upper and lower Ramsey transitions are superimposed; for example, the velocity distribution of the first higher-frequency transition is compared with that of the first lower-frequency transition, and so on. Note that

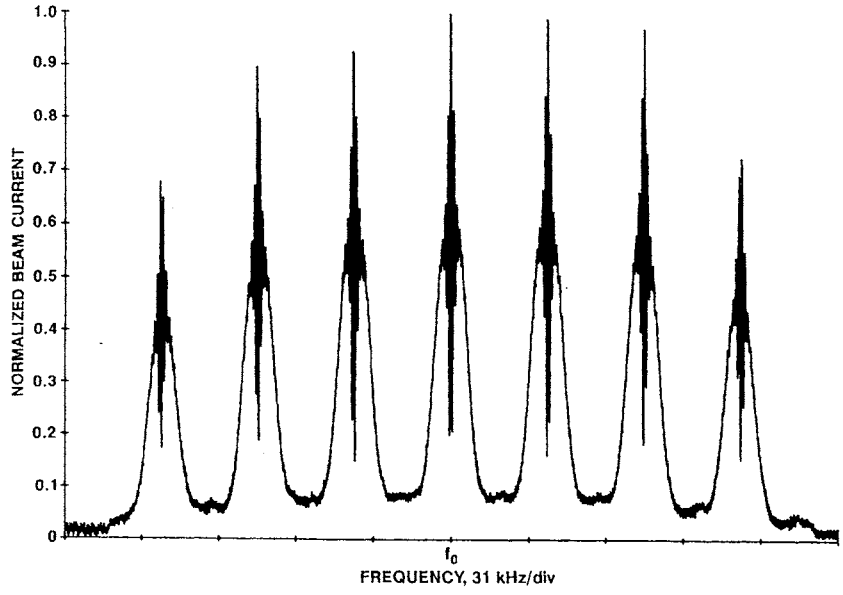


Figure 10. The CW Ramsey Patterns of the Seven Zeeman Transitions in a Cs Beam Tube

the higher-frequency velocity distribution is shifted above that of the lower by about 3 to 4 m/sec for each of the pairs. Apart from this shift, we can see that there is very little difference between these paired velocity distributions. This implies that the effects of Rabi pulling should be small. This was confirmed in our earlier study [4], which showed the effects of microwave power variation on the output of the standard using this tube, as shown in Figure 12. The maximum frequency offset was only 3×10^{-12} for a +3-dB change in microwave power, and the C-field could be changed over a wide range without affecting the standard's dependence on microwave power level.

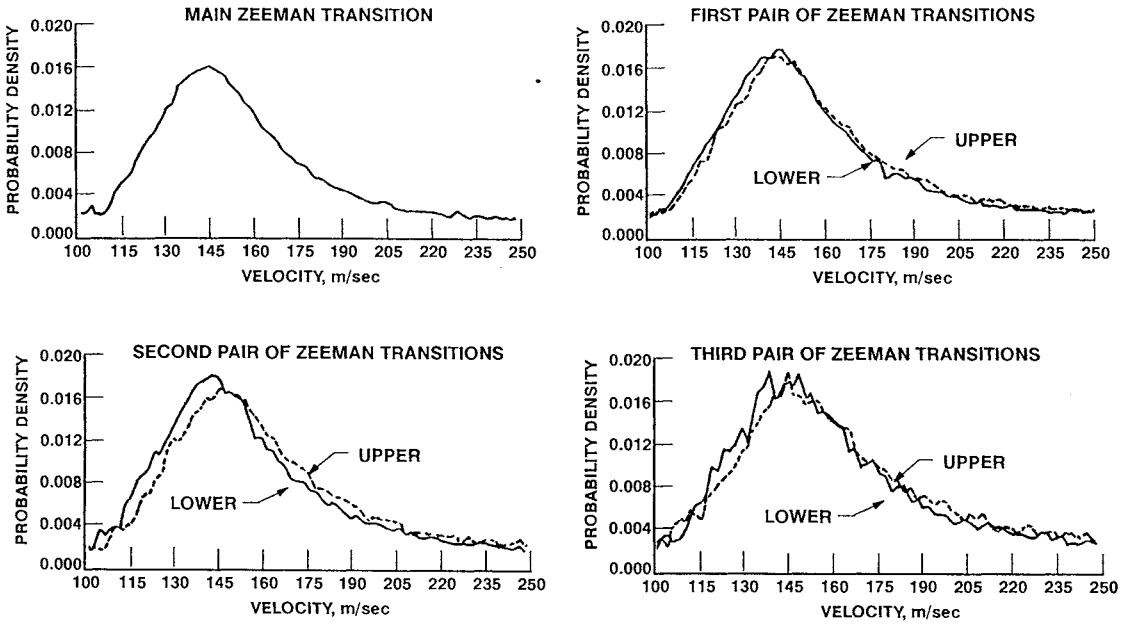


Figure 11. A Plot of All Seven Zeeman-Transition Velocity Distributions, Comparing the Higher-Frequency and Lower-Frequency Velocity Distributions

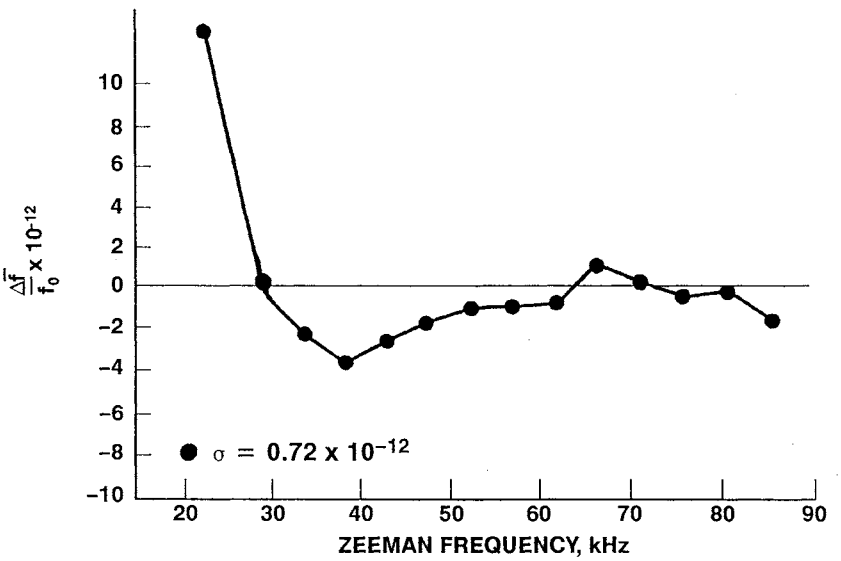


Figure 12. A Plot of the Change in Frequency of Cs Standard No. 6 as a Function of its C-Field for a CW Microwave Power Change of +3 dB above the Optimum Power Level (the Power Level that Results in Maximum Beam Current)

4. Summary

We have constructed an automated pulsed microwave measurement system to measure the velocity distributions of Cs atoms in Cs beam tubes. Measurement data on six Cs beam tubes from four different manufacturers are presented. Computer control enables the data to be collected efficiently and turns this technique into a simple tool for beam tube evaluation. We have illustrated the usefulness of this technique with three examples: (1) a comparison of two tubes taken from the same manufacturer to demonstrate the effect of manufacturing tolerances, (2) a simple comparison of the velocity distributions of tubes from different manufacturers, and (3) a measurement of velocity distributions for different Zeeman transitions and the correlation of those distributions with Rabi pulling. The technique augments other methods for characterizing beam tubes, and can help in the selection of criteria for acceptable velocity distributions.

References

[1] H. Hellwig, S. Jarvis, D. Halford, and H. E. Bell, "Evaluation and operation of atomic beam tube frequency standards using time domain velocity selection modulation," *Metrologia*, vol. 9, pp. 107–112, 1973.

[2] H. Hellwig, S. Jarvis, D. J. Glaze, D. Halford, and H. E. Ball, "Time domain velocity selection modulation as a tool to evaluate cesium beam tubes," in *Proceedings of the 27th Annual Symposium on Frequency Control* (Ft. Monmouth, N. J.), pp. 356–362, June 1973.

[3] D. A. Howe, H. E. Bell, H. Hellwig, and A. DeMarchi, "Preliminary research and development of cesium beam tube accuracy evaluation system," in *28th Annual Symposium on Frequency Control*, pp. 362–372, 1974.

[4] S. K. Karuza, W. A. Johnson, J. P. Hurrell, and F. J. Voit, "Determining optimum C-field settings that minimize output frequency variations in cesium atomic frequency standards," in *21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting* (Redondo Beach, Calif.), November 1989.

Resonators, Oscillators and Materials

ROM-04 Materials

Chairman: A. Haston

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-04 Number 1

PROGRESS ON ROTATED $GaPO_4$ Y-CUT BAW RESONATORS

J. Stadler, P. Krempl, G. Engel
AVL LIST Ges.m.b.H., Physics Division, Graz, Austria

Abstract

Recently, quartzhomeotypic $GaPO_4$ was introduced both as a promising piezoelectric material for high temperature sensor applications as well as a new material with temperature compensated cuts and electromechanical coupling larger than α-quartz and berlinite.

The present paper concerns the performance of rotated $GaPO_4$ Y-cut resonators. In particular, the influence of quality of raw material, design and geometry on the equivalent circuit parameters of thickness shear resonators is discussed.

It is shown how the classical tools for preparing crystal resonators, such as the use of the energy trapping concept, yield increasing performance on $GaPO_4$ resonators.

1. INTRODUCTION

Several metal orthophosphates $M^{3+}P^{5+}O_4$, with M = Al, Ga, Fe, crystallize in a quartz homeotypic structure[1]. Among these, the gallium orthophosphate crystal α-$GaPO_4$ is particularly interesting. There are two reasons for this: the first is the relatively large piezoelectric coefficient d_{11} which shows a low temperature dependence over a very large temperature interval[2], and the second is the existence of crystal orientations with temperature-independent bulk-wave resonance frequencies (the so-called AT cut)[3]. Growth[4-5], crystal quality[6] and low frequency dielectric properties[7] were reported recently.
However, the research on α-$GaPO_4$ is far from being completed. Crystal quality, piezoelectric properties, or thermophysical properties have not yet been sufficiently assessed to allow a thorough evaluation of all potential applications. Of current interest for resonator applications is a more complete material characterization (e.g. location and specifications of the AT cut and other compensated cuts). The present paper intends to report on progress towards this goal.

2. EXPERIMENTAL

2.1 Measurement of Equivalent Circuit Parameters

According to the IEC 122[8] or DIN 45100[9] publications, the equivalent circuit parameters of quartz resonators are equally well applicable to $GaPO_4$. The usual method of measurement (IEC 444[10] or DIN 45105[11]) relies on pi-network systems for the determination of the zero phase resonance frequency and the motional resistance R_1.
In principle, the same methods can be used for $GaPO_4$.

Another possibility, used here, is the transmission measurement according to Fig. 1a, which allows the recording of admittance ($Y = G + jB$, Fig. 1b), from which the parameters L_1, C_0, C_1, R_1 are determined according to Ref. 12:
The measured components B and G of admittance Y as a function of frequency are related to the equivalent circuit parameters by

$$G = \frac{R_1}{R_1^2 + (\omega L_1 - 1/\omega C_1)^2} \tag{1}$$

$$B = \frac{-(\omega L_1 - 1/\omega C_1)}{R_1^2 + (\omega L_1 - 1/\omega C_1)^2} + \omega C_0 \tag{2}$$

and satisfy the equation

$$(G - 1/2R_1)^2 + (B - \omega C_0)^2 = 1/4R_1^2 \tag{3}$$

which describes a circle centered at $x_0 = 1/2R_1$, $y_0 = \omega C_0$ with radius $r = x_0$. After computation of x_0, y_0 and r by means of the procedure described in Ref. 12, the admittance circle is rewritten in the form

$$\tan\Theta = \frac{B(\omega) - y_0}{G(\omega)} = \frac{-(\omega L_1 - 1/\omega C_1)}{R_1} \tag{4}$$

or

$$\tan\Theta = Q \frac{\omega_s^2 - \omega^2}{\omega_s \omega} \tag{5}$$

where

$$Q = 1/(\omega_s R_1 C_1) \tag{6}$$

Q and ω_s are determined from the linearization of Eq. 5 around a suitably chosen reference frequency ω_{ref}:

$$\tan\Theta = a(\omega_{ref} - \omega) + b \tag{7}$$

from which a and b are obtained from a least squares fit.
Now, the series resonance frequency, Q, and C_1 are obtained by

$$\omega_s = b/a + \omega_{ref} \tag{8}$$

$$Q = a\,\omega_s/2 \tag{9}$$

$$C_1 = 1/(\omega_s R_1 Q) \tag{10}$$

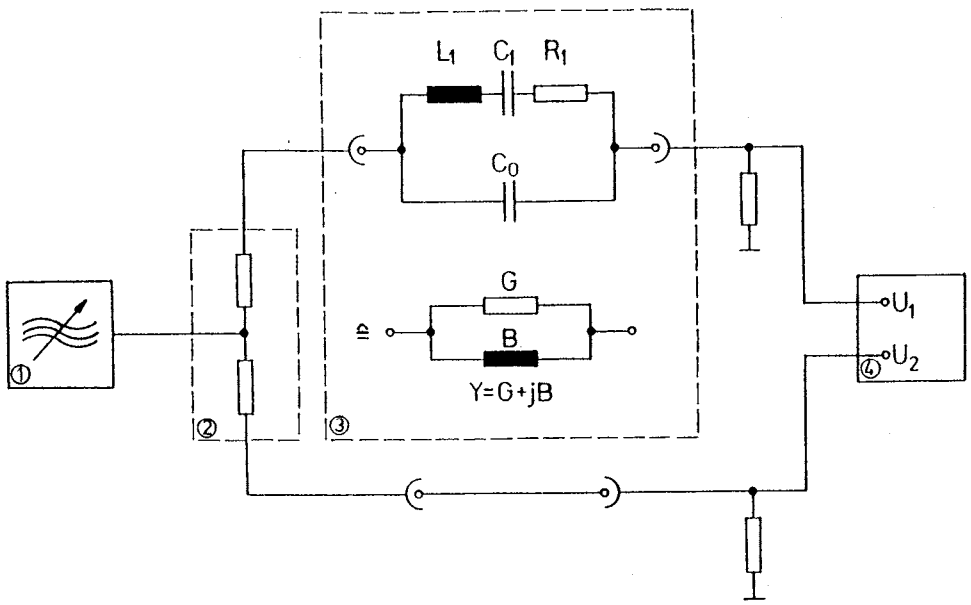


Fig. 1a: Measurement set up for determination of equivalent circuit parameters.
1 Signal generator
2 Power splitter
3 Equivalent circuit of device under test
4 Vector analyzer

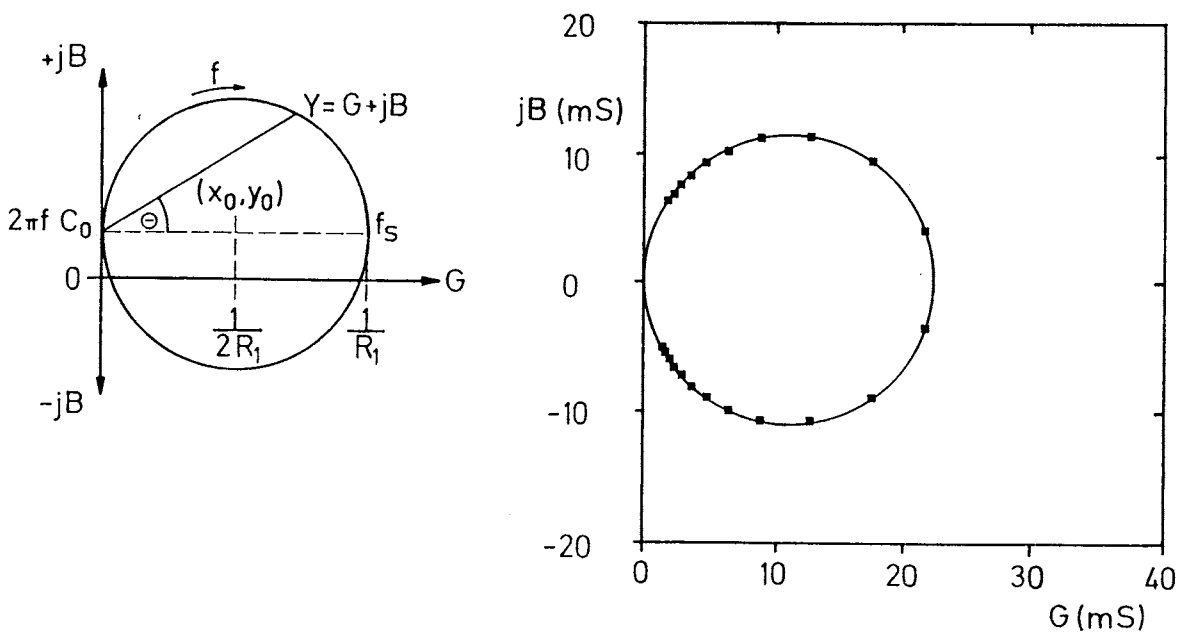


Fig. 1b: Admittance circle to illustrate the determination of f_s and equivalent circuit parameters and an example of a fitted admittance circle of a $GaPO_4$ resonator.

2.2 Blank Preparation

The $GaPO_4$ crystals were grown at AVL by a hydrothermal method. Starting from Y-cut seeds, a rhombohedral habit develops. Since according to Ref. 3 the AT cut is expected near the orientation (302), relatively large crystal plates were obtained from crystals with dimensions with at least 10 mm along the optic axis, enabling resonator diameters of 4.9 mm and 7.4 mm. A schematic outline of the relation of blank dimensions and rotated Y-cut geometry is shown in Fig. 2. Water traces were estimated by means of the usual IR-procedure described in Ref. 13 which has already been used previously[3]. The crystal plates were subjected to quality control by an etching method. RF sputtering was used for the deposition of gold electrodes. The amount of plating was measured by a quartz microbalance. The electrode tabs were aligned in parallel to the Z' axis.

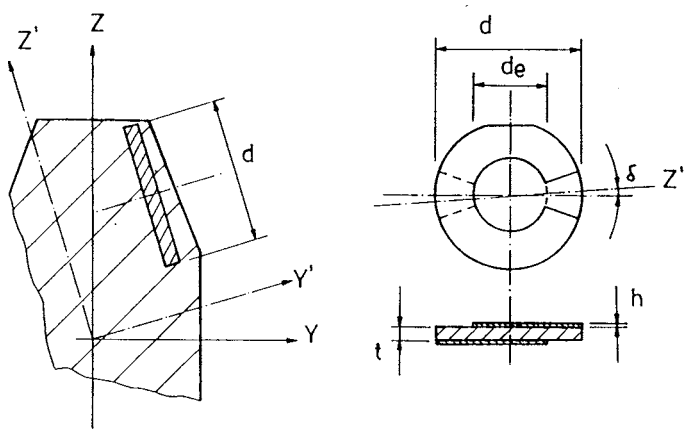


Fig. 2: Resonator design geometry.

2.3 Energy Trapping

In general, each harmonic mode of order n is followed by unwanted inharmonic modes. For the suppression of these modes, the energy trapping concept[14] was developed. Its application to thickness-shear quartz resonators led to a reduction of the motional resistance, R_1 and to an increase of the motional capacitance C_1. According to Ref. 14, the mode of order m and all higher inharmonics can be suppressed by letting the mass loading $R = 2\rho'h/\rho t$ (ρ' and ρ are the densities of the electrode material and the crystal) not exceed the amount

$$M_m^2/n^2 \; (t/d_e)^2, \tag{11}$$

where the constants M_m depend on the elastic constants. From the above mentioned geometry of available crystal plates, a ratio of $t/d_e \sim 0.08$ can be achieved. Starting from this ratio, values of the maximum mass loading for $GaPO_4$ resonators for inharmonics following the first harmonic mode are shown in Table 1, where the elastic data given in Ref. 3 are used to calculate M_m. The corresponding M_m's for an AT-cut of quartz with the elastic constants according to Ref. 15 are also given in Table 1.

Table 1:

Constants M_m of Eq. (11) for different symmetric (s) and asymmetric (as) thickness twist (TT)-modes.
Computed values of M_1 for quartz[15] (AT-cut) and $GaPO_4$[3] (cut in the vicinity of (302) plane).
Mass loadings for $t/d_e = 0{,}08$ are shown.

Mode	M_m	Quartz M_1	Quartz mass loading (%)	$GaPO_4$ M_1	$GaPO_4$ mass loading(%)
TT_1/as	$(2m-1)\left(\frac{C_{55}'}{2C_{44}'}\right)^{1/2}$	0,944	0,6	0,914	0,5
TT_1/s	$m\left(\frac{2C_{55}'}{C_{44}'}\right)^{1/2}$	1,888	2,3	1,827	2,1
TT_3/as	$(2m-1)\left(\frac{C_{55}'}{2C_{66}'}\right)^{1/2}$	1,09	0,9	1,04	0,7
TT_3/s	$m\left(\frac{2C_{55}'}{C_{66}'}\right)^{1/2}$	2,17	3,04	2,08	2,8

3. RESULTS AND DISCUSSION

Because of the lower frequency constant of $GaPO_4$ resonators, the electrode diameter can be smaller in comparison with a quartz resonator of the same frequency and mass loading (Table 1 and Eq. (11)). The electrode thickness of this resonator can be larger in $GaPO_4$ by about 10%. This is quite different from quartz and $AlPO_4$ because of the different crystal densities (2.65 g/cm³ in quartz and $AlPO_4$ but 3.59 in $GaPO_4$).
Fig. 3 shows the frequency responses of a $GaPO_4$ crystal plate with different mass loadings which illustrate the qualitative verification of the energy trapping concept. The reducion of the mass loading decreased the activity of all modes except one, which is supposed to be antisymmetric. This mode is shown by the frequency response with the lowest mass loading. Asymmetric modes are possible under e.g. electrode misalignement and slight deviations from parallelism. According to Glowinsky et al[16], the activity of these modes is increased if the mass loading is reduced. We believe that this case is seen in the frequency responses of lowest mass loading. The mode which occurs at a $\Delta f/f$ of 0.018 is supposed to be symmetric. This mode is lost ("untrapped") for the resonator with the lowest mass loading.
Another aspect is the influence of the water content on the activity of inharmonic modes. The frequency response of two different $GaPO_4$ resonators with the same plating, about the same orientation and with a relatively large difference of water content, is compared in Fig. 4.
The inharmonic modes of the resonator with the higher insertion loss show a lower activity (dashed line). We believe that this behaviour is associated with the high water content which has already been observed by Jumas et al[17] in connection with berlinite.
In Fig. 5 we see a comparison between achieved R_1 values of some $GaPO_4$ resonators and the commercial specifications for R_1 of quartz resonators of corresponding enclosures.
As is apparent from the diagram that we are approaching the quartz specification for the

motional resistance R_1. However, one of our $GaPO_4$ resonators achieves a R_1 value of about 40 Ω. This relatively low value of R_1 which corresponds to an insertion loss of 3 dB is well comparable to quartz.
Another well-known aspect of the resistance R_1 is the influence of the water content of the crystal on this value. Fig. 6 shows the achieved resistance R_1 of some $GaPO_4$ resonators in relation to their water content. No clear correlation between the values of R_1 and the water content is obtained, suggesting alternative explanations for high R_1 values: an insufficient design geometry, micro twinning or other crystal defects. At the present moment we do not exactly know which one of these is the major culprit.
The equivalent circuit parameters of C_0 and C_1 are shown in Fig. 7 as compared to quartz. The higher C_1 values of the $GaPO_4$ resonators correspond to the higher coupling coefficient which is about twice that of quartz.

In our opinion, further work concerning the following subjects is necessary:

- Determination of material constants on crystals with low water content.
- X-ray topography is required for the mode identification of resonators.
- Further improvement of design should demonstrate that $GaPO_4$ is a suitable material for high performance AT-resonators.

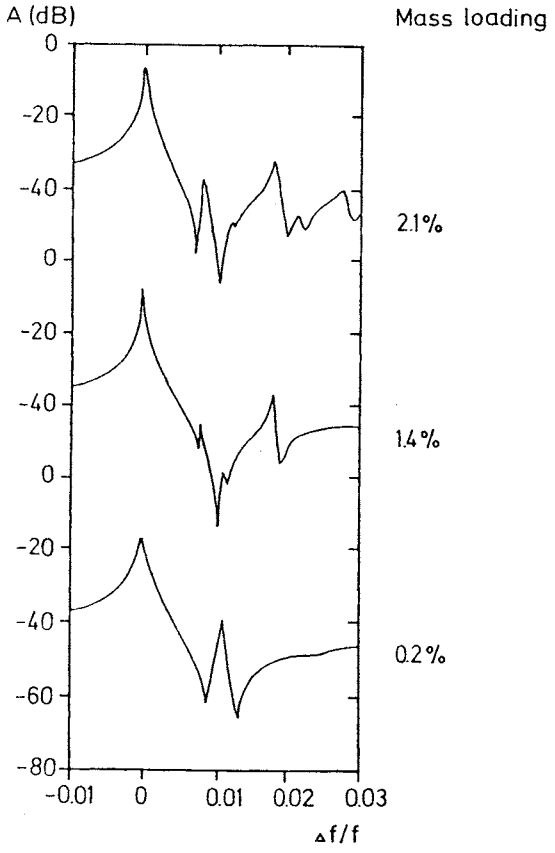


Fig. 3: Frequency responses of a $GaPO_4$ resonator with different mass loading. The resonator plate is oriented in the vicinity of the (302) plane.

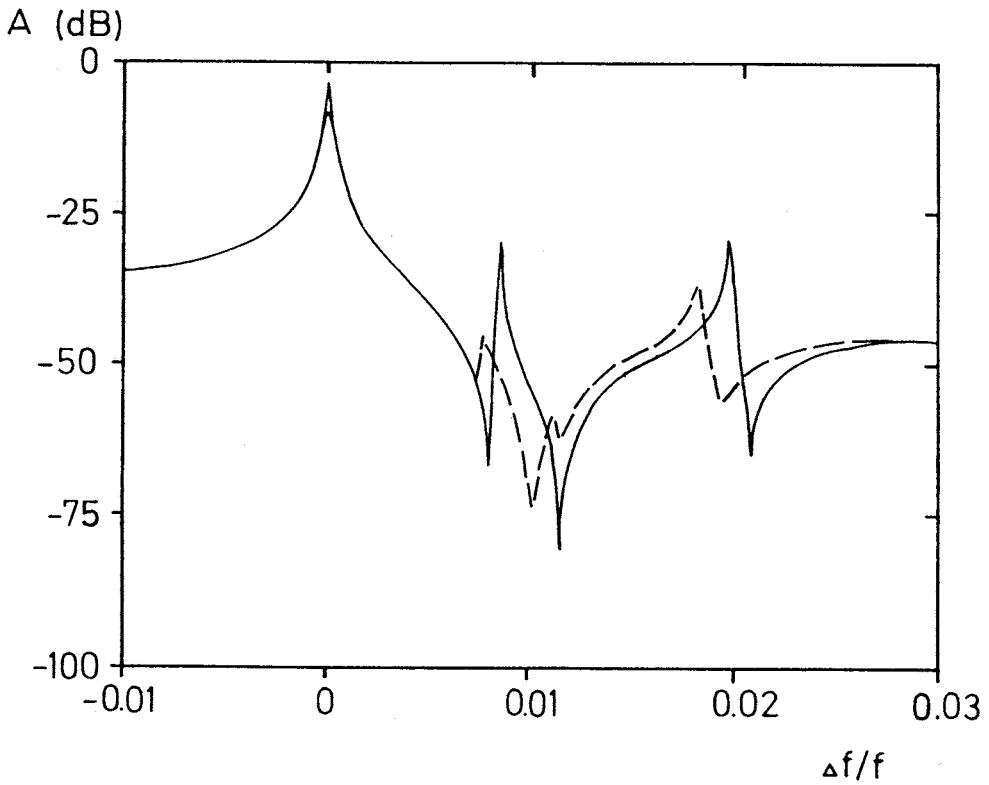


Fig. 4: Frequency responses of two $GaPO_4$-resonators with an orientation in the vicinity of the (302) plane.

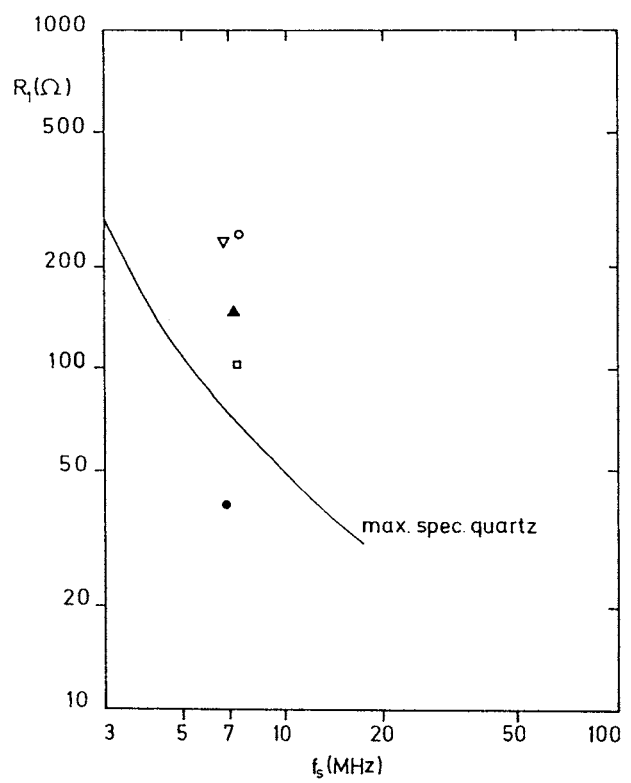


Fig. 5: Achieved R_1 values of some $GaPO_4$-resonators in comparison to commercial quartz specifications.

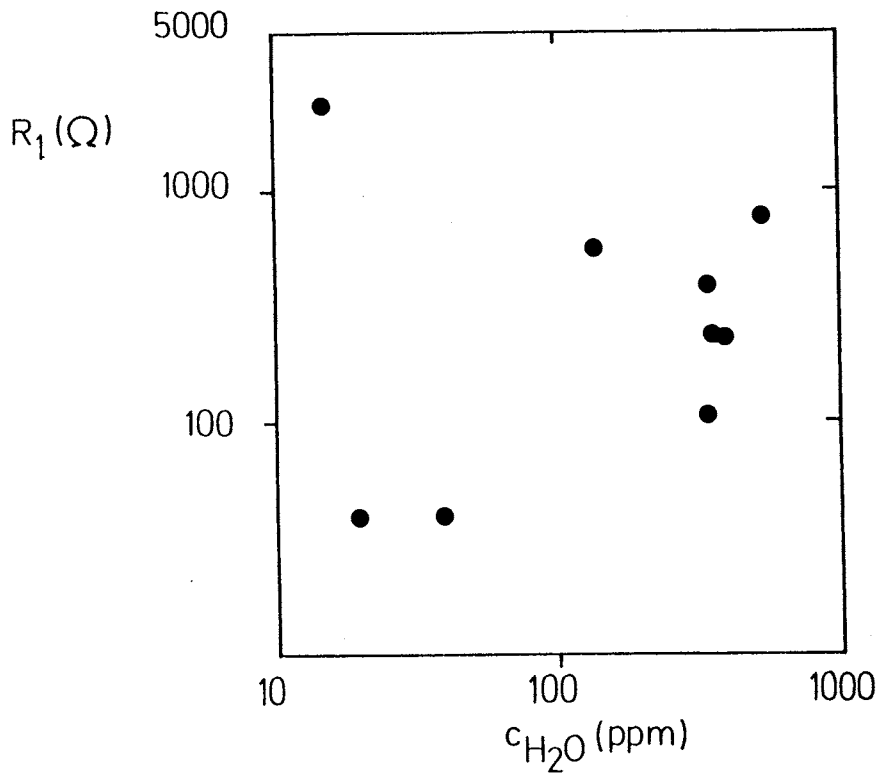


Fig. 6: Water content of $GaPO_4$ plates estimated from IR spectra and achieved values of motional resistance R_1.

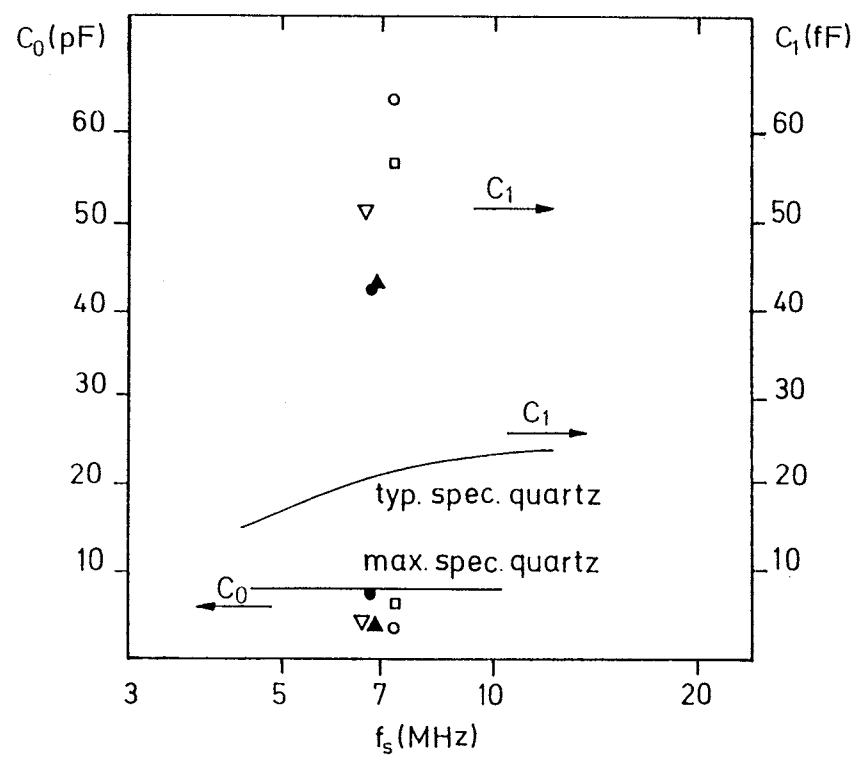


Fig. 7: Achieved values of equivalent circuit parameters C_0 and C_1 in comparison to commercial quartz specification.

4. CONCLUSION

The latest state-of-the-art computational methods have been used to evaluate progress on rotated Y-cut resonators.
The use of high quality (low water content) $GaPO_4$ crystals and improved design (energy trapping) have led to low motional resistance (as low as 40 Ω) and attractive motional capacitance.

ACKNOWLEDGEMENT

This work was supported by the Austrian Forschungsförderungsfonds für die Gewerbliche Wirtschaft (Project-No. 3/6882), which is gratefully acknowledged.

REFERENCES:

1) K. Kosten and H. Arnold,
Z. Krist. 152 (1980), p. 119
2) G.F. Engel and P.W. Krempl,
Proc. 2nd Europ. Frequency & Time Forum (Neuchatel, 1988), p. 827
3) G. Engel, P. Krempl and J. Stadler,
Proc. 3rd Europ. Frequency & Time Forum (Besancon, 1989), p. 50
4) S. Hirano, K. Miwa and S. Naka,
J. Crystal Growth 79 (1986), p. 215
5) B.N. Livitin, V.I. Popolitov, M.A. Simonov, O.V. Yakubovich and I.M. Yaroslavski,
Sov.Phys.Crystallogr. 32 (1987), p. 284
6) G. Engel, H. Klapper, P. Krempl and H. Mang,
J.Crystal Growth (1989)
7) H. Ribes, J.C. Giuntini, A. Goiffon and E. Philippot,
Eur.J.Solid State Inorg.Chem., t.25 (1988), p. 201
8) IEC-Publication 122-1, International Electrotechnical Commission (1976)
9) DIN 45100, Deutsches Institut für Normung, Berlin, Beuth Verlag Berlin, 1981
10) IEC-Publication 444-1 (1986)
11) DIN-45105, Teil 3, (1978)
12) R.J. Willamson, IEEE-Transactions on Ultrasonics and Frequ. Control,
Vol. UFFC-34, N.6, Nov. 87, p. 681
13) R.D. Aines and G.R. Rossmann,
J. Geophys.Res. 89 (1984), p. 4059
14) W. Shockley, D.R. Curron and D.J. Koneval,
J. Acoust. Soc. Amer., 41 (1967), p. 981
15) R. Bechmann, Phys. Rev., 110(5), 1958, p. 1060
16) A. Glowinski, R. Lancon and R. Lefevre,
Proc. 27th Ann. Frequency Contr. Symp. (1973), p. 233
17) J.C. Jumas, A. Goiffon, B. Capelle, A. Zarka, J.C. Doukhan, J. Schwartzel,
J. Detaint and E. Philippot,
J. Crystal Growth 80 (1987), p. 133

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-04 Number 2

RECENT ADVANCES IN BERLINITE GROWTH AND CHARACTERIZATION

E. PHILIPPOT, A. IBANEZ, A. GOIFFON*

X. BUISSON, R. ARNAUD**

B. CAPELLE, A. ZARKA***

Y. TOUDIC, J. SCHWARTZEL, J. DETAINT****

* URA DO407, USTL, 34095, MONTPELLIER, FRANCE

** SICN, ANNECY, FRANCE

*** UNIVERSITE DE PARIS VI, 75252 PARIS, FRANCE

**** CNET/LAB/OCM LANNION ET CNET/PAB/BAG/MCT BAGNEUX FRANCE

INTRODUCTION

In the recent years, berlinite crystals with a high Q factor were obtained by high temperature growth in H_3PO_4 [1] then using the H_2SO_4 solvent [2]. Since the beginning of 1989, the growth experiments were mostly devoted to increase further the quality and the size of the crystals in view of an important industrial application (I.F. Filters for the new paneuropean numerical radiotelephone) that will use a large quantity of this material in the nineties. We present here the most important results recently obtained.

First, we report and discuss the present state of crystal growth conditions of berlinite in sulphuric acid medium correlated with infrared spectroscopy and X-ray topography investigations. In the last part, we discuss the evolution of the angular position of the AT cut with the crystal perfection. Then, we present the results of experiments made to obtain an accurate determination of the angular position of the AT cut of the high purity material now available and we indicate the properties of berlinite devices using this orientation.

I - CRYSTAL GROWTH STUDIES

Recently, we have demonstrated [3] that the production of high Q berlinite crystals was no longer an utopia. Nevertheless, in the recent times, it has appeared that some of the growth parameters could be further optimized. Among the remaining questions, one was to find what could be the most appropriate sulphuric acid concentration for berlinite crystal growth. That it to say, what are the best crystal growth conditions to have an OH concentration as low as possible, and hence a Q factor as high as possible. The figure 1 summarizes the problem : all other crystal growth parameters being constant, the [OH] content decreases in crystals when sulphuric acid concentration is reduced from 9 to 4.5 M.

Then, we have started studying more diluted solutions, 3M in particular. Many trials have been carried out, at a growth temperature, $T_c \sim 240°C$, by the vertical temperature gradient method with $2° < \Delta T_c < 6°C$. With theses conditions we have never registered suitable growths, i.e. we have observed either seed dissolution, or zero, or very weak crystal growth. Technical problems being avoided, an explanation can be given from two previous Raman spectral studies made on sulphuric solutions. Indeed, we can schematize the sulphuric acid dissociation in water as:

$$H_2SO_4 + nH_2O \rightarrow HSO_4^- + H^+(nH_2O)$$

$$HSO_4^- + nH_2O \rightarrow SO_4^{=} + H^+(nH_2O)$$

From Robertson et all. [4], "the maximum predicted for the average hydratation number of the proton, n, at 2 ± 1 M sulphuric acid is a real fact, it indicates that drastic changes in the solution are occurring in this region, an observation supported by a wide variety of evidence". This consideration is confirmed by Chen et all. [5]. Their results can be illustrated by the figure 2 where the curve shown is closely related to the variation of ultrafast proton transfer from one water molecule to another one, versus sulphuric acid concentration.

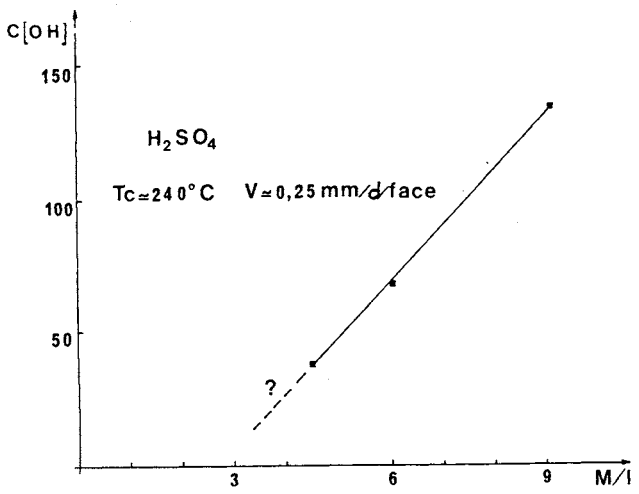


Figure 1: OH content as a function of the sulphuric acid concentration (at constant growth rate and crystallization temperature).

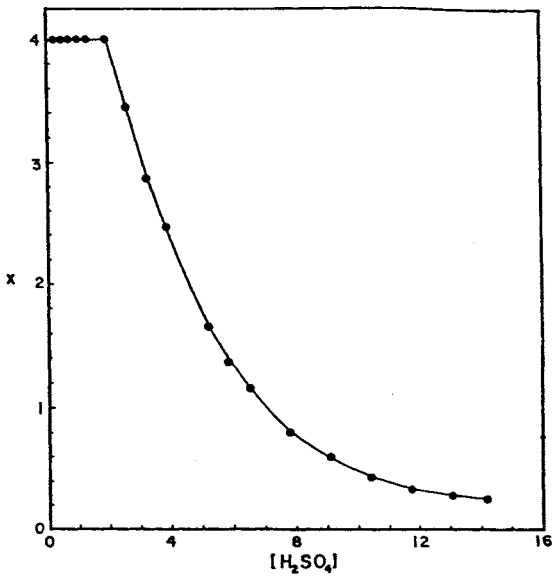


Figure 2: Variation of the partial reaction order (x) with respect to water in ultra-fast proton transfert versus sulphuric acid molarity (from ref. [4]).

If the possibility of lower sulphuric acid concentrations must be forsaken to improve quality of berlinite crystals, another parameter can be considered : the increase of the crystal growth temperature, $T_c > 240°C$, [6]. Unfortunately again, there are two difficulties:

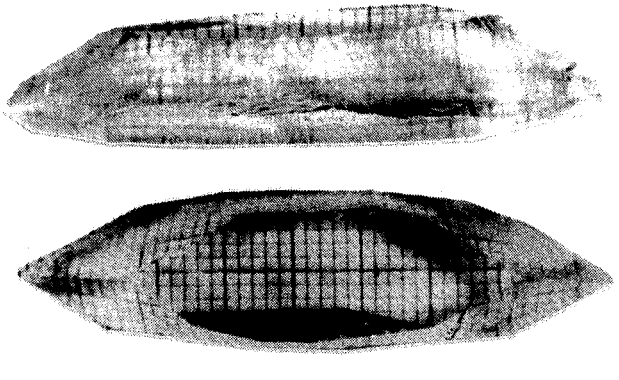
First, from a chemical point of view, we observed a beginning of partial sulphuric acid decomposition [7] since $T_c = 240°C$, (except for the acid of Prolabo source) :

$$H_2SO_4 \rightarrow H_2O + SO_3$$

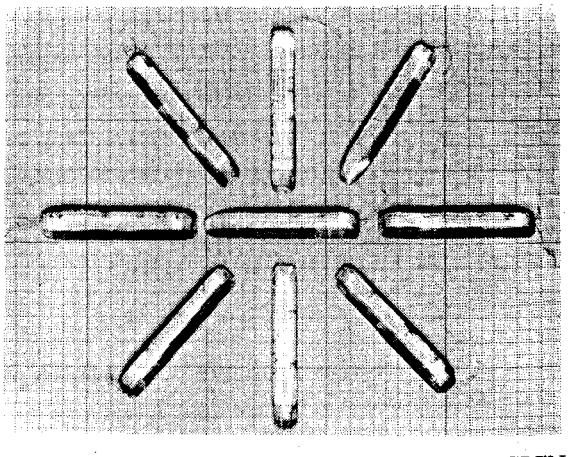
with yellowing solutions and characteristic SO_3 smelling.

Secondly, if we consider the crystal growth rates in terms of T_c, all other parameters being constant, the V_x/V_z ratio increases too much with T_c and becomes unsuitable for industrial crystal growth purposes.

Thus, from all our present results, the best conditions for a good crystal growth seem to be $T_c \sim 240°C$ and 4.5 M sulphuric acid medium, giving berlinite crystals with [0H] content of about 20 ppm (photo 1). These conditions have been transferred at an industrial scale to SICN in Annecy (photography 2).



Photography 1: Crystal grown in 4.5M sulphuric acid at T =240°C.



Photography 2: Crystals for seeds, grown at SICN (Tc=210°C, 6M H_2SO_4

II - INFRARED CHARACTERIZATION AT DIFFERENT STEPS OF CRYSTAL GROWTH

This study lead us to measure the "water" content following each step of the crystal growth. This "water" can be included both under the form of very tiny fluid inclusions (molecular water) and/or of point defects (-0H groups).

It can be detected by near infrared spectroscopy (N.I.R.S.). As is shown in figure 3, from Boulogne et all [8], it is possible to distinguish molecular water (near 5200 cm^{-1}) and -0H groups (near 4300-4400 cm^{-1}).

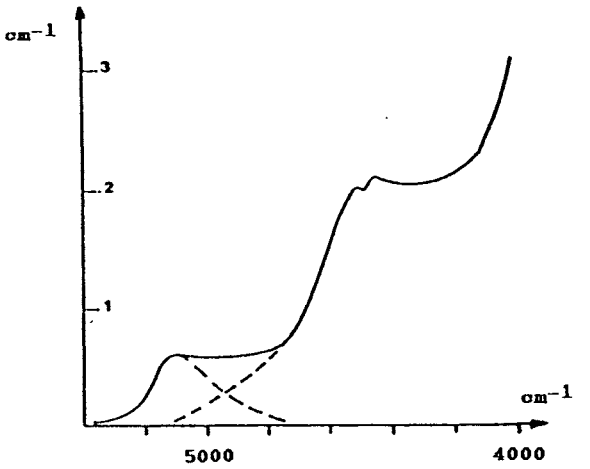


Figure 3: NIR absorbance spectrum (referred to 1cm thickness) of berlinite. The peak near 5200 cm-1 is attributed to molecular water while this near 4300cm-1 is due to OH point defects (from ref. [8]).

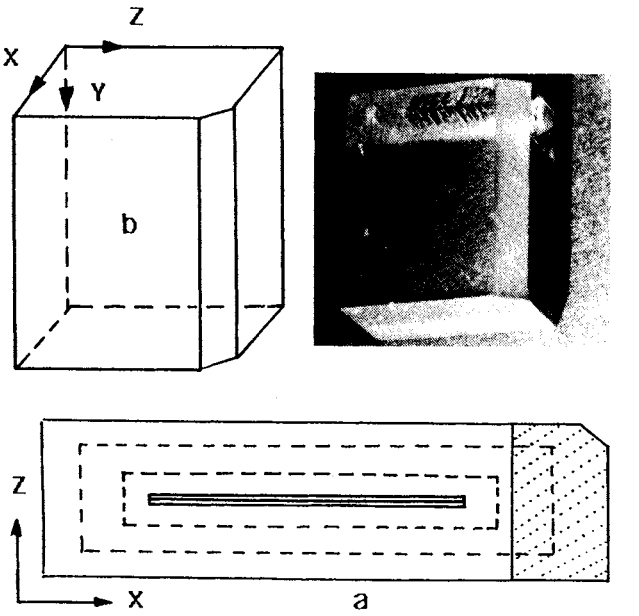


Figure 4 and Photography 3:
4a:crystal grown in 6M sulphuric acid, Tc=210°C with Vx=0.34 mm/d/face and Vz=0.11 mm/d/face .
4b: +X zone part of the crystal chosen for the IR study.

For this purpose, we have chosen a crystal with an high enough total water content, so to be easily measured ($50 < C_{[OH]} < 100$ ppm) [synthesized with growth conditions : H_2S0_4 6M, $T_c = 210°C$, $V_x = 0.34$ mm/d/face, $V_z = 0.11$ mm/d/face]. In this crystal we have cut a small parallelepiped (X = 6.0 mm, Y = 12.21 mm and Z = 9.97 mm) out of seed, (figure 4 and photo 3). First, we have investigated all crystal directions with a 5 mm cylindrical beam. All spectra are approximately identical and figure 5 gives one of them with incident beam parallel to Z axis in the range 6000 - 2500 cm^{-1}. In these spectra ,we can only observe -0H groups between 4300 and 4400 cm^{-1} and no molecular water near 5200 cm^{-1}.

After that, we have carried out a step-scanning of Y-face parallel to Z-axis with a slit of 1 x 2 mm. The results, given in figure 6, demonstrate that the central part of the crystal, close to the seed, shows a maximum -0H content. Using the quantitative formula suggested by Steinberg et all. [9] to measure the water concentration :

$$\alpha = 1/d(cm)(A_{3300} - A_{3800})$$

and $c_{OH}(ppm) = 48.7\alpha$

we find $C_{[OH]} \sim 70$ ppm (d is the sample thickness, A_{3300} and A_{3800} are the absorbance values at 3300 and 3800 cm^{-1}). This C value decrease slowly to $C_{[OH]} \sim 50$ ppm after about 2 mm growth. We can also observe the same evolution for point defect concentration between 4300 and 4400 cm^{-1}, figure 7a. In this case, no quantitative formula is proposed for the -0H content but the qualitative values given correspond to the maximum height of the peak of absorbance measured as is shown in figure 8. Then, a step-scanning of the Y face made in a direction parallel to the X-axis has evidenced a continuous decrease of the point defect concentration when one moves away from the seed, figure 7b.

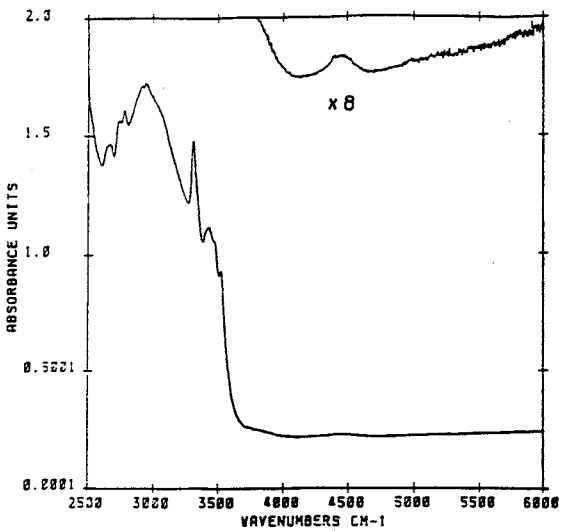


Figure 5: IR absorbance spectrum of the berlinite crystal shown in figure 4 . Incident beam parallel to the z axis . The upper NIRS has a magnified scale (X8).

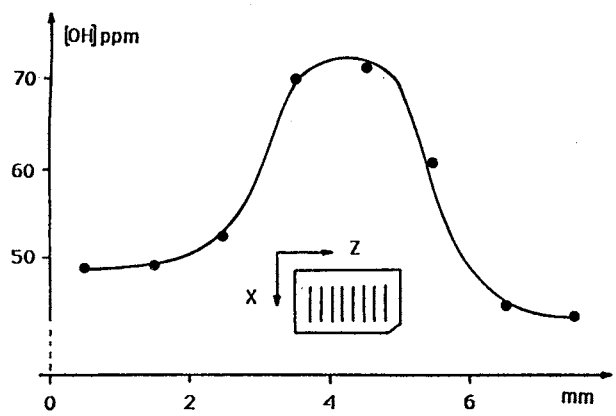
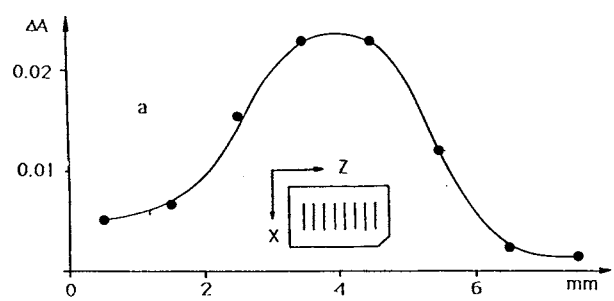
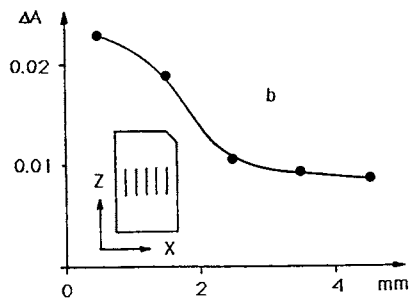


Figure 6: Step-scanning of the OH content from left to right, in the Y face, in a direction parallel to the z axis.

In both cases, the maximum value of the peak height ΔA (measured in absorbance units) is of about 0.023, but, it must be noticed that the decrease of the OH concentration is most important when the step-scanning is parallel to the Z axis. This difference can be explained by the fact that the V_z rate (0.11 mm/d/face) is weaker than the V_x one (0.34 mm/d/face), knowing that $C_{[OH]}$ increases with the growth rate [9].



7a: step-scanning parallel to the z axis .



7b: step-scanning parallel to the x axis.

Figure 7: Variations of ΔA(maximum value of the peak near 4300cm-1) in absorbance units (see figure 8) for a beam parallel to the y axis.

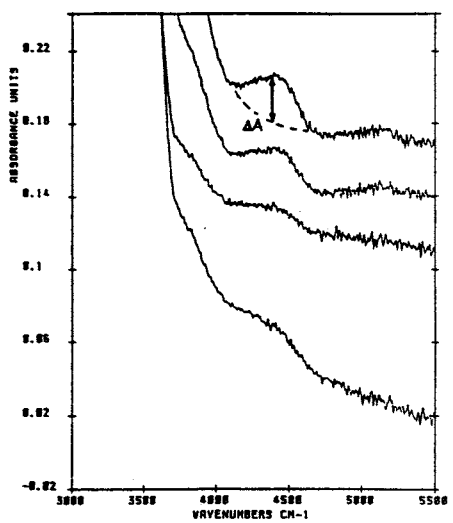


Figure 8: NIR absorbance spectra of a berlinite cystal showing ΔA values decreasing from the upper curve (near the seed) to the lowest one .

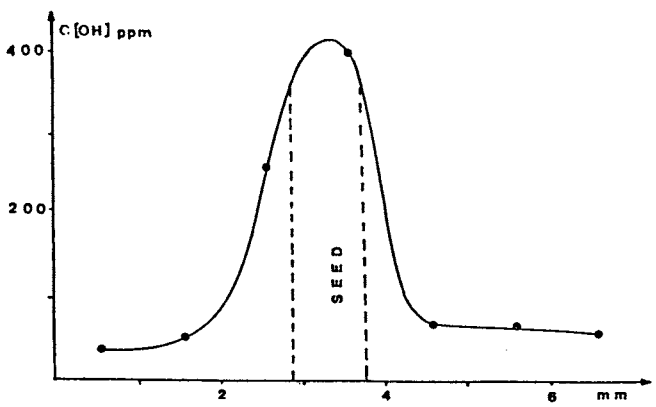


Figure 9: Variations of the OH concentration in a crystal grown with an OH rich seed when the start up and growth conditions are good (Tc=240°C,H_2SO_4 4.5M).

On the other hand, as shown last year [3], the berlinite crystal growth carried out at higher temperature (T_c = 240°C) in lower H_2SO_4 concentration (4.5 M) from "wet" seed lead to a drastic reduction in -OH content as shown in figure 9.

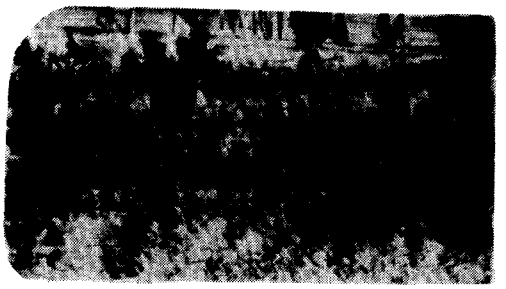
In conclusion, all theses NIRS results confirm that T_c = 240°C and 4.5 M H_2SO_4 are presently the best growth conditions. For the crystal grown at relatively low temperature (T_c ~ 210°C in 6M sulphuric acid) the slow decrease from the seed of -OH content may explain some contrasts observed near the seed by X ray topography.

III - CHARACTERIZATION BY X-RAY TOPOGRAPHY

III.1 - Improvement of crystalline quality

Using growth techniques based on the method suggested by Zarka et all. [10] for quartz samples, the quality of berlinite crystals has been improved in taking a Z seed in the X zone.The photography 4 shows a topograph of a Z plate obtained from a Z seed of standard quality. A method was derived from the previous one permits to improve the quality of crystals that can be obtained with seeds of modest quality. Photography 5 is a topograph of an X-cut seed of poor quality which presents a high density of dislocations.The topographs of two Z plates obtained from this X seed are shown in photography 6. The crystal quality, namely in the X zones, were sensibly improved in comparison with the seed one and the dislocation density has been much lowered. Some growth bands are visible in the +X zone. An important difference of quality between the +X and -X zones can be noticed.

Photography 4: Topograph of a Z plate obtained from a Z seed .

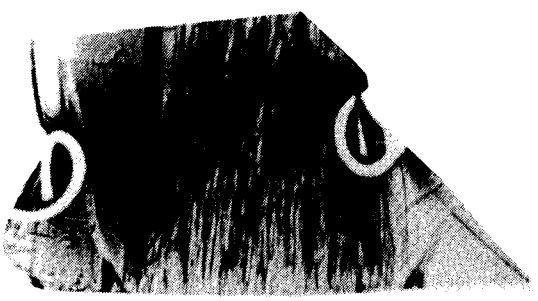


Photography 5: Topograph of an X cut seed .

Photography 6: Topograph of two Z plates grown with the X seed shown in photography 5.

Another example is shown in Photography 7. This crystal was grown using the same method with a standard seed. The quality is excellent in both X and X⁻ zones.The variation of OH content in the crystal, evidenced by the NIRS measurements shown in this paper, induces also X-ray contrasts in the vicinity of the seed as it can be observed on photography 7.

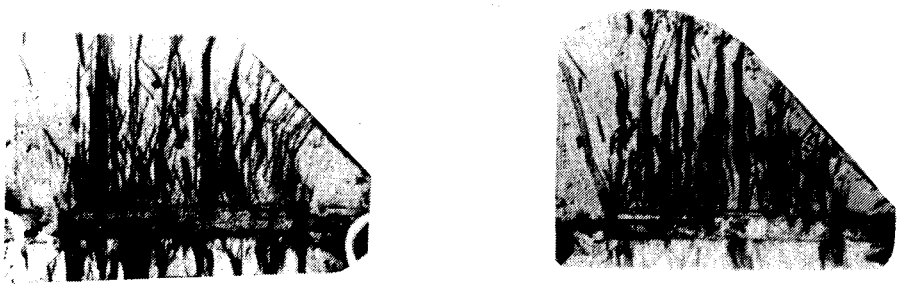
Photography 7 : topograph of another sample obtained using the same method as preceedingly.



Photography 8: Topograph of a recently obtained AT cut plate.

III.2 - Quality of recently obtained crystals

The improvement of the crystalline quality and the increase of crystal sizes allow to get AT-cut sample having an useful surface exceeding one square centimetre. These samples were cut in crystals grown with X plate seeds. The external contour of the AT plate shown here results of the shape of the crystals (end plates near one extremity). The three samples presented in the photographies 8, 9 and 10 have different dislocation densities. The last two ones are of excellent quality and they are characteristic of the present state of the crystalline quality. The measurements of the thermal behaviour of plate near AT (shown latter in this paper) were made using resonators cut in similar plates. The external contour of the plates shown in these topograph results mostly of the shape of the crystals (plate cut at one end of the crystals).



Photography 9 , 10: Topograph of recently obtained AT plates.

IV - PIEZO-ELECTRIC CHARACTERIZATION OF RECENTLY OBTAINED CRYSTALS: properties of the AT cut of berlinite

The Y-rotated cuts are presently the most useful ones in berlinite, since, as for quartz, only one thickness mode is excited by an electric field normal to the thickness (this is important for filter applications) and also since the two most important compensated cuts (AT and BT) are Y rotated cuts.

IV.1 - Evolution of the angular position of the AT cut with the crystal perfection

The measurements of the constants of berlinite by Chang and Barsch [11] and the calculation of A.Ballato [12] have permitted in 1976 to determine that a compensated cut with properties similar to those of the AT cut of quartz exists for berlinite near Y-30°. Due to this analogy, this cut was also called the AT cut of berlinite. The cubical law of variation of the resonance frequency as a function of temperature was first observed experimentally in 1979 for a Y-29°45' cut [13]. Since that time, the crystalline perfection and the purity of the material were greatly improved. During this improvement which has mainly consisted in removing the very important quantity of OH impurity (up to several thousands ppm) and the extended defects existing in the crystals, a large variation of the values of the temperature coefficients of the elastic constants was observed [14] . Correlativelly, an important shift of the angular position of the AT cut was also observed . With crystals having an OH concentration of a few hundred ppm the AT cut was still situated near Y-30° figure 1 0 [14] , The angular situation of this cut , computed using the constants and their temperature coefficients measured by Bailey et Al. [15] in 1984 , is also situated near Y-30° figure 11 . More recently, it was possible using high temperature growth in phosphoric acid [1] [16] , and then in sulphuric acid [2] [17] to obtain crystals having an OH concentration as low as about 20 ppm. With these crystals having at 5 MHz an intrinsic Q factor greater than 800000, the AT cut is situated near Y-33°(see below).

The effect of the reduction of the OH concentration on other quantities characteristic of the piezoelectric modes appears to be less important. For the phase velocity (or for the frequency constants) of the slow transverse mode of the Y rotated cuts an increase of a few % was observed . For the coupling coefficient, only very small difference were measured between samples with different OH concentration and also with the computed values (using the constants of reference [15]). This fact is probably due to a compensation between an increase of the values of the elastic constants and a decrease of the dielectric constants. An example typical of the effect of the OH impurity on several parameters of the resonators is given in table I for the Y cut.

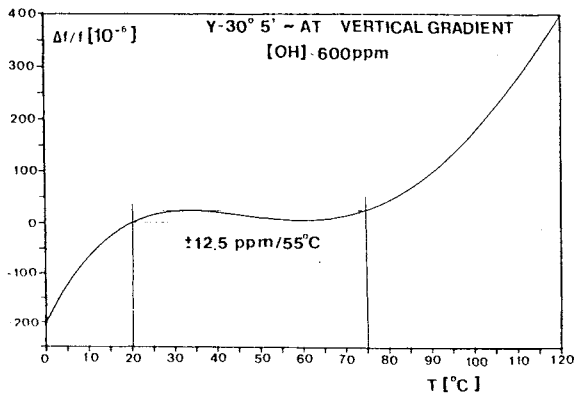


Figure 10: Thermal behaviour of the Y-30°05' cut of OH rich crystals.

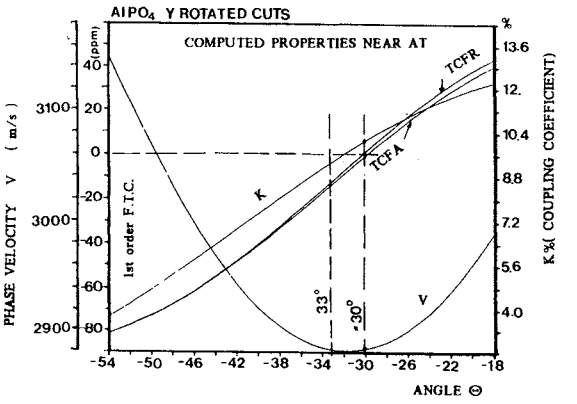


Figure 11: Computed properties of resonators with near AT orientations.

Table I effect of the OH concentration

Solvent growth T°C	[OH] (ppm)	Q (5MHz)	N fr kHz.mm	CTF1 $10\text{-}6°C^{-1}$	CTF2 $10\text{-}9\ °C^{-2}$
H_3PO_4 180	~600	3500-15000	1700 (±7)	+55 to +65	-60 to +180
H_2SO_4 230	20-30	600000-800000	1712 (±2)	81. (±1)	47.(±1)
computed*			1688	63.5	

In table I it can be noticed that a dispersion of the frequency constants and principally of the temperature coefficients that is much larger than the experimental incertitude is observed for the [OH] rich samples ; this dispersion is reduced to the order of the experimental errors for the high purity samples. The computed values were obtained with the constants of reference [15]. From the measurements made by IR spectrometry and also by electron microscopy [18], it can be inferred that the large dispersion observed in [OH] rich crystals results of that of the OH concentration within the crystals. A some how similar effect, but of much reduced importance was still observed, only on the 2nd and 3rd order temperature coefficients, for OH concentrations as low as 80 to 50 ppm. This latter observation is probably to be related to the non-uniform distribution of the OH concentration noticed in figures 6, 7 and 9 , and also, to the contrasts observed near the seed by X-ray topography (Photography 7). It corresponds also to facts already observed in synthetic quartz, where a higher OH concentration can be present near the seed or in the different growth zones of the crystals.

IV.2 - Properties of resonators near AT

A systematic research of the AT cut has been made using the crystals obtained in H_2SO_4 solvent. The principal results are given in table II for cuts in the vicinity of the AT cut.

Table II : Properties of Y rotated resonators near AT

CUT	CRYSTAL	Q (5 MHz)	CTF1 $10\text{-}6\ C^{-1}$ at 20°C	CTF2 $10\text{-}9\ C^{-2}$ at 20°C
Y + 21°(r)	220°/Y	75000	+ 51	+ 21
Y + 21°(z)	210°/Y	50000	+ 47	+ 17
Y - 30° 0'	230°/Z	17000	+ 10	-
Y - 32°05'	230°/Z	22000	+ 5.4	+ 2.6
Y - 32°30'*	240°/X	130000	+ 2.9	- 9.8
Y - 32°45'*	240°/X	107000	+ 0.87	- 11
Y - 33°02'*	240°/X	123000	0.5 ± 0.1	- 14 ± 3
Y - 32°12	240°/X	37000	- 1.2	- 19
Y - 33°20'	240°/X	37000	- 1.7	- 25
Y - 33°40'*	240°/X	135000	- 3.5	-
Y - 38°13'	230°/Z	15000	- 21	-

* measurements made on third overtone

** In accordance with our previous remark concerning the possible influence of OH concentrations on the 2nd order FTC, these values are not given for crystals with [OH] >50ppm.

In table II we can observed that the AT cut of berlinite crystals with low OH concentrations is situated very near Y-33°.This can be better observed in table III where the temperature coefficients of the resonance frequency of cuts near AT are given; they are referred to temperatures close to the inflexion temperature (60° or 62.5°C). The values measured with an AT quartz are also given for a reference purpose.

Table III higher order F.T.C. near the AT cut of berlinite

Growth cycle	Angle theta	RFTC1 $10^{-6}°C^{-1}$	RFTC2 $10^{-9}°C^{-2}$	RFTC3 $10^{-12}°C^{-3}$	Tref. °C
SICN 320	32°45'	+ 0.46	+ 1.7	102	62.5
SICN 319	33°02'	- 0.14	+ 0.59	116	62.5
SICN 248	33°12'	- 1.9	- 3.6	120	60.
QUARTZ SICN	~35°16'	-0.119	-0.28	104	27.0

It can be observed in table III that the third order temperature coefficient near AT is of the same order of magnitude as that for AT quartz.

On figure 12 is represented the frequency-temperature variations observed for the Y-33° 02' (+-1') plates operating on the 3rd overtone mode. it can be noticed that the inflexion temperature is higher than for AT quartz . The relative peak to peak frequency variations on a 80 °C interval is comprised between -4 and +4 10-6. This is a demonstration of the possibility to achieve very high temperature stability with berlinite. On figure 13 are represented the relative frequency temperature variations for cuts around the AT cut at nearly a 15 minute interval, it can be observed that a behaviour very similar to that of quartz is obtained and also that the Y-33° 02' is very near (a few minute) the "zero angle cut" i.e. the cut with an horizontal inflexion tangent for third overtone resonators. Since we

have observed that the temperature coefficients of the antiresonance frequency are for any overtone rank extremely close to those of the corresponding resonance frequency, the temperature coefficient of the coupling coefficient is very small and the difference between the "zero angle" cuts for different overtones is believed to be small and of an order of magnitude similar to that observed in quartz. This point is currently under investigation.

As for quartz, for the AT cut a simultaneous cancellation of the 1st and 2nd order temperature coefficients of the resonance frequencies occurs for a given temperature (the inflexion temperature of the "zero angle cut") .It can be observed that, for berlinite as for AT quartz, the inflexion temperature remains nearly constant as the angle is varied around the zero angle cut. It can be seen in figure 12 and 13 and in table III that this temperature is very near 65°C.

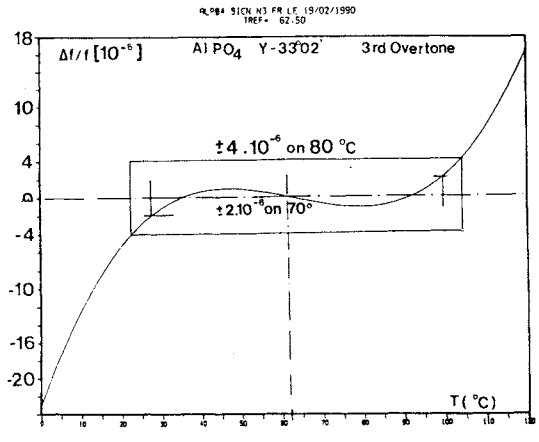


Figure 12: Relative thermal variations of the resonance frequency for Y-33°02'(+-1') resonators.

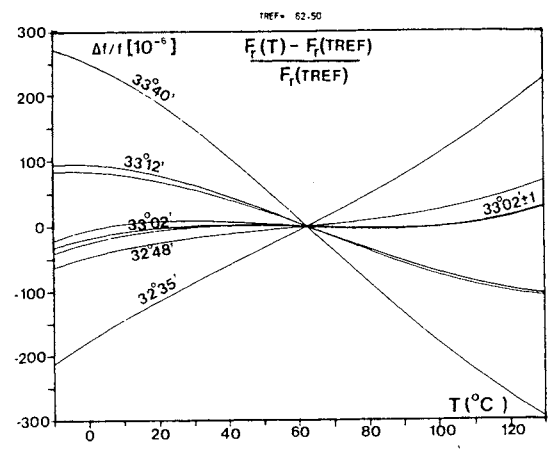


Figure 13: Thermal behaviour of berlinite resonators with orientations near Y-33°.

In table IV are given the observed frequency constant of the fundamental resonance (at zero mass loading) and the measured coupling coefficient for the Y-33°02' cut, they are compared to results computed for the same cut. The experimental value of the coupling coefficient was obtained from the resonance-antiresonance distance with no corrections for the parasitic capacitances.

Table IV - Frequency constant and coupling coefficient for the Y-33°02' cut

	Frequency constant (kHz.mm)	Coupling coefficient (%)
Experimental	1477	10.1
Computed	1436	9.35

In figures 14 and 15 are the represented the electrical responses of AT resonators, operating on the fundamental and the third overtone mode.

On the whole, the AT cut of berlinite appear to have very attractive properties including a very favourable thermal behaviour, a high Q factor and a value of k^2 much higher than that obtained with AT quartz. These properties are very interesting for applications to filters having bandwidth larger than what is achievable with quartz and for VCXO. The value of the inflexion temperature may also be very useful for ovenized devices. The energy trapping properties of AT berlinite are similar to that of AT quartz and the vibration modes have a quite moderate lateral anisotropy. The ratio of the electrode diameter to the thickness of the plate to trap only the 1st anharmonic of each overtone is smaller than for AT quartz, this may have a major interest in view of the obtention of miniaturized devices.

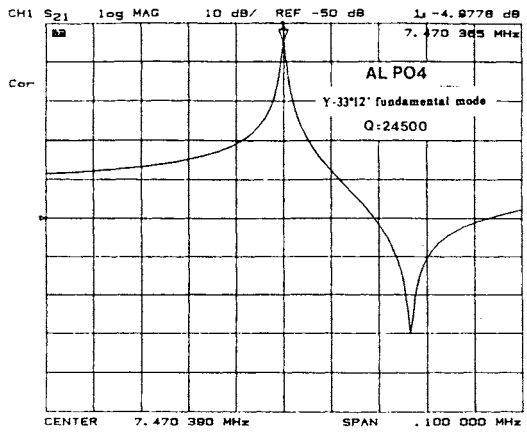


Figure 14: Electrical response of a fundamental mode near AT resonator.

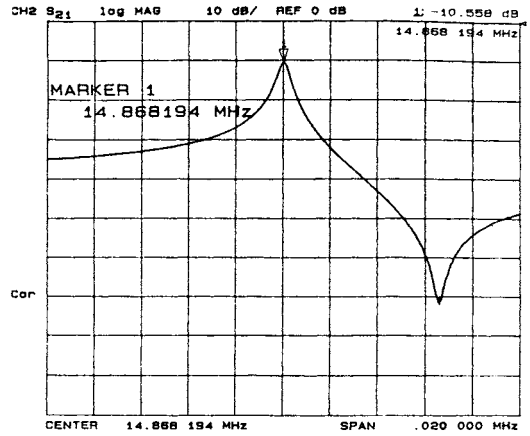


Figure 15: Electrical response of the 3rd overtone of an AT resonator.

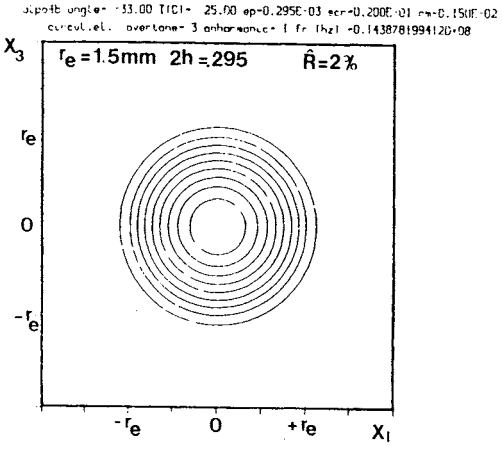


Figure 16: Computed vibration mode of the 3rd overtone of an AT resonator.

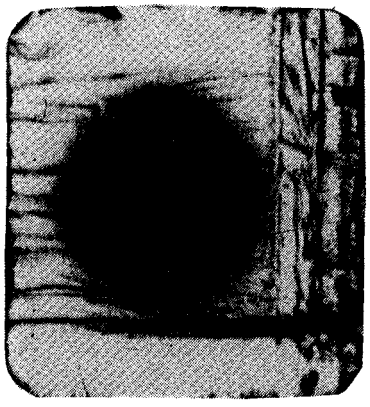


Figure 17: Experimental mode of the 3rd overtone of an AT resonator.

On Figure 16 and 17 are represented the computed and the observed 3rd overtone mode of a plane AT resonator.

V - Conclusion

The experimental results presented in this communication indicate that , the "driest" berlinite crystal ,at present, are obtained in a 4.5 M sulphuric acid medium for a crystallization temperature T =240°C. The infrared measurements have shown that the OH concentration is decreasing from the seed to the external boundaries of the crystals. The OH concentration decrease to much lower values with these new growth conditions than when a lower Tc and/or higher acid concentrations were used . For all cases, the NIRS measurements have shown that the OH impurity is present only as point defects and not as molecular water (fluid inclusions).

X-ray topography experiments were made to evaluate several methods to improve the crystalline quality. They have permitted to observe that an important improvement of this quality can be obtained with growth conditions leading also to a very reduced OH concentration.

The angular position of the AT cut has been accurately determined. The thermal behaviour of AT berlinite devices is very favourable, this joined to a much higher value of k^2 than for quartz, makes berlinite a very attractive material for filter and VCXO applications.

Acknowledgments:

The authors acknowledge the CNRS and the CNET for financial support. They thank Professor M.Galtier and Mr. J.M. Durand for NIRS measurements ,MM Douis and D.Excoffon for cutting and polishing the many samples used in this study and Professor M.Maurin for useful comments and suggestions. They thank particularly Mrs M.P.Louis for the preparation of this manuscript.

VI - REFERENCES

[1] J.C.Jumas, A.Goiffon, Y.Toudic, J.Schwartzel, J.Détaint, E.Philippot, Mat. Res . bull. 22 , 1063-1071 (1987)

[2] E. Philippot, A.Goiffon, M.Maurin, J.Détaint, J.Schwartzel, Y.Toudic, B.Capelle, A.Zarka, to appear in J. Cryst. Growth 1990

[3] E.Philippot,A.Goiffon,C.Avinens, X.Buisson,R.Arnaud, O.Bignon,A.Zarka, B.Capelle,M.T. Sebastian, Proc. 3rd European Time and Frequency Forum. 57 (1989)

[4] E.B.Robertson,H.B. Dunford, J. Amer. Chem. Soc. 86, 5080 (1964)

[5] H.Chen, D.E. Irish , J.Phys. Chem. 75 (17), 2672 (1971)

[6] J.C. Doukhan, B.Boulogne, P.Cordier, E.Philippot, J.C.Jumas, Y.Toudic, J. Cryst. Growth 84, 167 (1987).

[7] Bodenstein, Katayama, Z.Elektrochemie 15,244(1919)

[8] B.Boulogne,P.Cordier, J.C.Doukhan, Proc. 1st European Time and Freqency Forum (1987)

[9] R.F.Steinberg, M.K.Roy, A.K. Estes ,B.H.T. Chai,R.C. Morris, Proc. 1984 IEEE Ultrasonics Symposium. 279 (1984)

[10] A.Zarka, Liu Lin, X.Buisson, J. Cryst. Growth 53,394, (1981) & 57,466 (1982)

[11] Z.P. Chang, G.R. Barsch, IEEE Trans. Sonics UltraSon. SU23, 127, (1976)

[12] A. Ballato , G.J. Iafrate, Proc. 30th Ann. Frequency. Control Symposium 141 (1976)

[13] J.Détaint, M.Feldmann, J.Henaff, H.Poignant, Y.Toudic, Proc. 33rd Annual Frequency Control Symposium. p70 1979

[14] J. Detaint, E. Philippot, J.C. Jumas, J. Schwartzel, A. Zarka, B. Capelle, J.C. Doukhan, Proc. 39th Annual Frequency Control Symposium p 234 (1985)

[15] D.S. Bailey, J.C. Andle, D.L. Lee, W. Soluch, J.F. Vetelino, B.H.T. Chai, Proc. 1983 IEEE Ultrasonics Symposium 335 (1983)

[16] J.P. Hou, B.H.T. Chai, Proc. 1987 IEEE Ultrasonics Symposium. p419 (1987)

[17] E. Philippot, A. Goiffon, J.C. Jumas, C. Avinens, J. Detaint, J. Schwartzel, A. Zarka, Proc. 42nd Annual Frequency Control Symposium p. 138 (1988)

[18] J. Detaint, E. Philippot, B. Capelle, Y. Toudic, B. Boulogne, J.C. Jumas, A. Zarka, J.C. Doukhan, Proc. Symp. on Freq. Control and Synthesis p 314 (1987)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-04 Number 3

NEW DESIGNS FOR RESONATORS AND FILTERS USING LITHIUM TANTALATE

J. Détaint, J. Schwartzel, H.Carru, R. Lefèvre, C. Joly
CNET dept PAB/BAG/MCT/CMM 196, Av. Henri Ravera Bagneux France
And
B. Capelle, Y. Zheng, A. Zarka
L.M.C. Université P. & M. Curie Place Jussieu Paris France

I - INTRODUCTION

Lithium tantalate is a material which has for the thickness mode a very large electromechanical coupling factor (k reaching 45 % for the fast shear mode) together with compensated cuts (1) (2). These properties are very attractive to obtain filters with large bandwidth and oscillators with important shifts. Moreover, despite a rather poor crystalline quality (non stoechiometry, dislocations etc...) this material presents of one the smallest acoustic dissipation known (3).

However, excepted for some doubly rotated cuts (4) (5), only the fundamental fast thickness shear mode of cuts near the X one is thermally compensated, and, with a conventional design this mode has a rather modest Q factor. During the course of a study presently made in view to research an optimal solution for the I.F. filters of the new pan-european numerical radiotelephone system several conventional and new designs were investigated to enhance the Q factors of devices using the fundamental mode and to obtain compensated overtones with a sufficient capacitances ratio to have the required bandwith (150 to 250 kHz at center frequencies in the range 45-75 MHz).

Also, using the sophisticated technique of X ray strosboscopic topography, a new insight on the properties of the vibration modes in this material was obtained which explains the low Q factor observed with the fundamental mode and several features of the different new designs for overtone resonators.

II - THICKNESS SHEAR MODES IN LITHIUM TANTALATE

Several loci of zero first order temperature coefficients exist for the two basic families of shear modes in lithium tantalate (2). They are represented as a function of the plate orientation, for the fast and the slow shear modes, in Figure 1 and 2. In these figures the plate orientation is located by the two polar angles α and β defining the orientation of the normal to the plate.

For the fast shear modes (B mode) (Figure 1) the most important zero F.T.C. locus is that situated around the X cut. These orientations provide a very high coupling coefficient. In this case the thermal compensation for the resonance frequency is obtained due to a large negative value of the first order T.C. of the coupling coefficient (6). This compensation disappears for the overtones (Figure 1). The second locus of zero F.T.C. is situated around the Y rotated cuts, it exists for the fundamental mode and the overtones, but, requires the use of doubly rotated plates. For the slow shear mode (C mode) two loci of zero first order temperature exist, one is around the Z cut, the second is close to the Y-10° cut (Figure 2). For a conventional excitation by a field in the direction of the normal to the plate, doubly rotated cuts are required to have a non zero coupling coefficient.

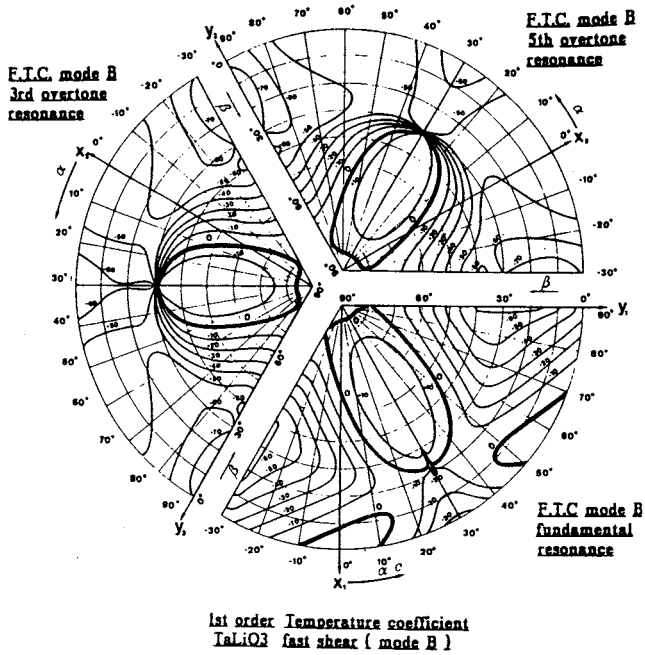


Figure 1: 1st Order F.T.C. (B mode).

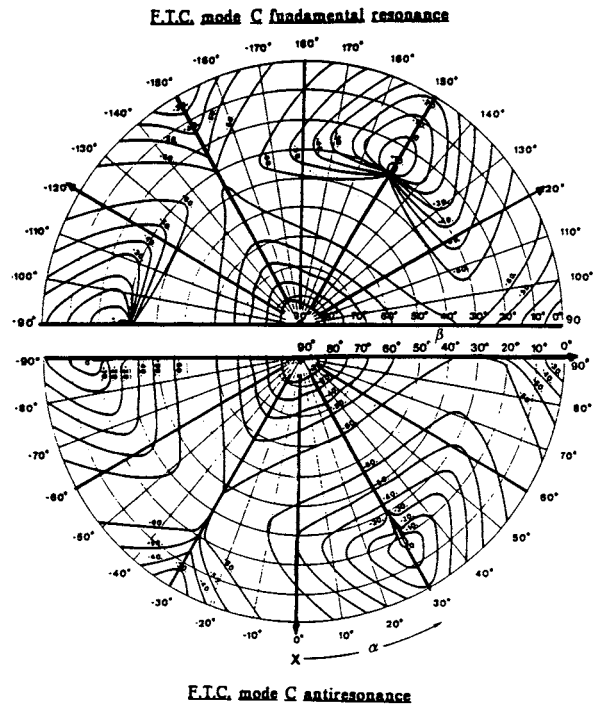


Figure 2: 1st Order F.T.C. (C mode).

Using, the same representation, level curves for the electromechanical coupling coefficient are displayed for the fast shear (B mode) and the slow shear (C mode) in Figures 3 and 4 (2).

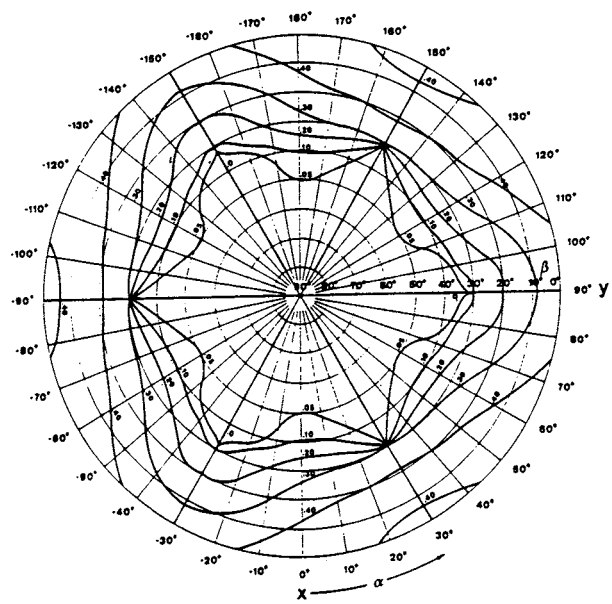


Figure 3: Coupling coefficient (B mode).

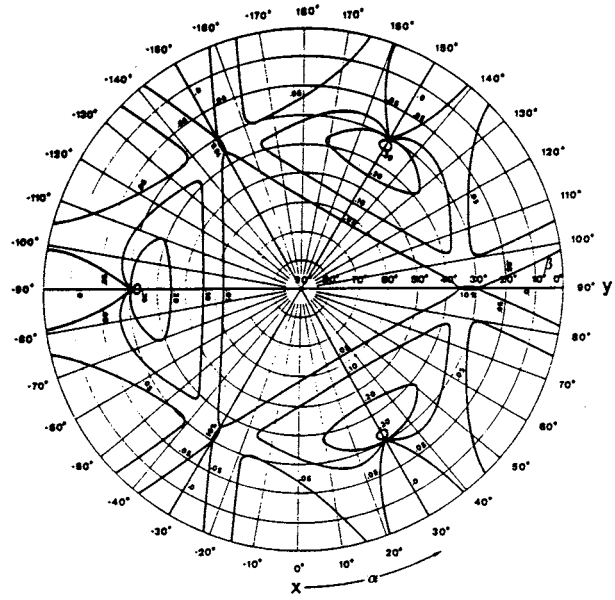


Figure 4: Coupling coefficient (C mode).

III - DEVICES USING THE FUNDAMENTAL MODE

The most interesting cuts are those near the X one. For these cuts the coupling coefficient is about 45 % and the frequency constants, without mass loading, are near 2000kHz.mm.

On Figure 5 are represented the most important properties of the B mode of X rotated cuts (around the y axis). A typical frequency spectrum of an X cut resonator is given on Figure 6. It can be noticed that the C mode is also excited but has a rather low coupling coefficient. The properties of the B and C modes of fundamental X cut resonators are given in table I.

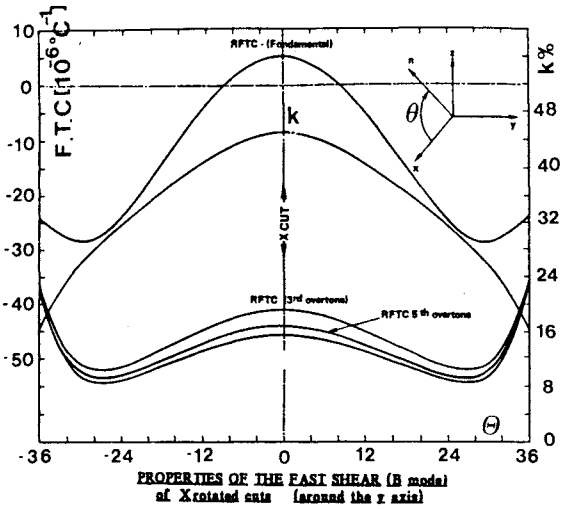


Figure 5: Fast Shear mode of X rotated cuts

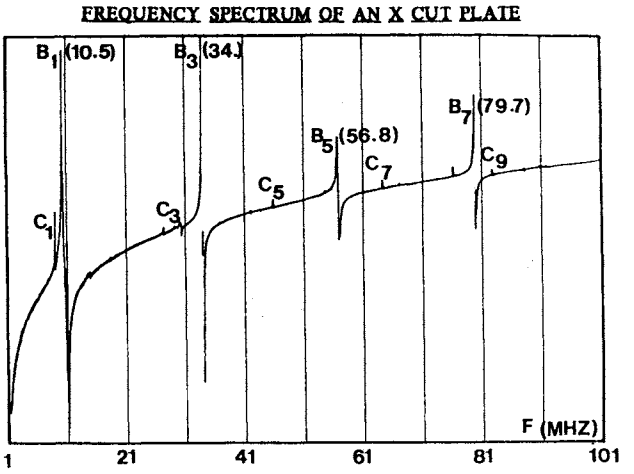


Figure 6: Frequency spectrum of an X cut plate.

TABLE I

MODE	$\vec{u}$	VELOCITY	$\bar{C}$(MKSA)	k	T.C1Fr	TC1Fa	TC2Fr.1	TC2Fa	Fr_1*2h
B	u_1=0.0 u_2=0.5749 u_3=-0.8181	4212 m/s	1.323 10^{11}	45,1 %	+4,0 10^{-6}	-44.10^{-6}	114.10^{-9}	-53 10^{-9}	1916 (kHz.mm)
C	u_1=0.0 u_2=0.8181 u_3=0.5749	3368 m/s	0,845 10^{11}	3,5 %	-63.10^{-6}	-63.10^{-6}	-30.10^{-9}	-30 10^{-9}	1683 (kHz.mm)

III.1 - Conventional X cut resonators

The Q factors of X cut resonators using the fundamental mode are very low, in most cases in the range of a very few thousands. This is one of the main concern with this type of resonator (particularly for applications to filters with bandwith lower than 1%); one other concern being the possibility to have unwanted responses near the resonance or the antiresonance not due to anharmonic modes (as for other materials, the anharmonic modes can be suppressed by the choice of proper dimensions for the electrodes , a proper mass loading and the fulfilment of certain conditions of symmetry).

An examination of the vibration mode of such resonators was made by stroboscopic X ray. topography using synchrotron radiation (7) (8). This method allows to determine the components of the displacement u_k (x_1, x_2, x_3, t) provided they have an $e^{j\omega t}$ time dependence where ω is the angular recurrence frequency of the synchrotron radiation. These experiments were made with X cut fundamental resonators having a resonance frequency equal or close to 12667120 Hz or 15846400 Hz, using the white radiation of D.C.I at the L.U.R.E (ORSAY-FRANCE). The time resolution obtained is 1 ns and the spatial resolution in the order of some ten microns depending mostly, in this case, on the absorbtion and the perfection of the material. Some of the topographs obtained are displayed in Figure 7 and 8.

On figure 7a is displayed mostly the main component of the thickness shear displacement (component u_1' in the direction of the $\vec{u}$ vector given in table 1 for the B unidimensional mode). This topograph was obtained using the stroboscopic technique ; it can be observed that the conventional trapped shear mode observable at the center of the plate is coupled to a plate mode extending up to the edge of the plate. On Figure 7b is represented a conventional topograph obtained in the same electrical and diffraction conditions (with integration over the time of $\vec{u}$ (x_1, x_2, x_3, t). The coupled plate mode can no longer be observed in these conditions. This means that this plate mode is not a standing wave but a progressive one. It was directly verified in similar cases, by a video imaging technique using the stroboscopic effect provided by a slow shift of the relative phase between the X ray pulses and the excitation current, that the plate waves were travelling from the electroded region towards the mounting clips of the resonator.

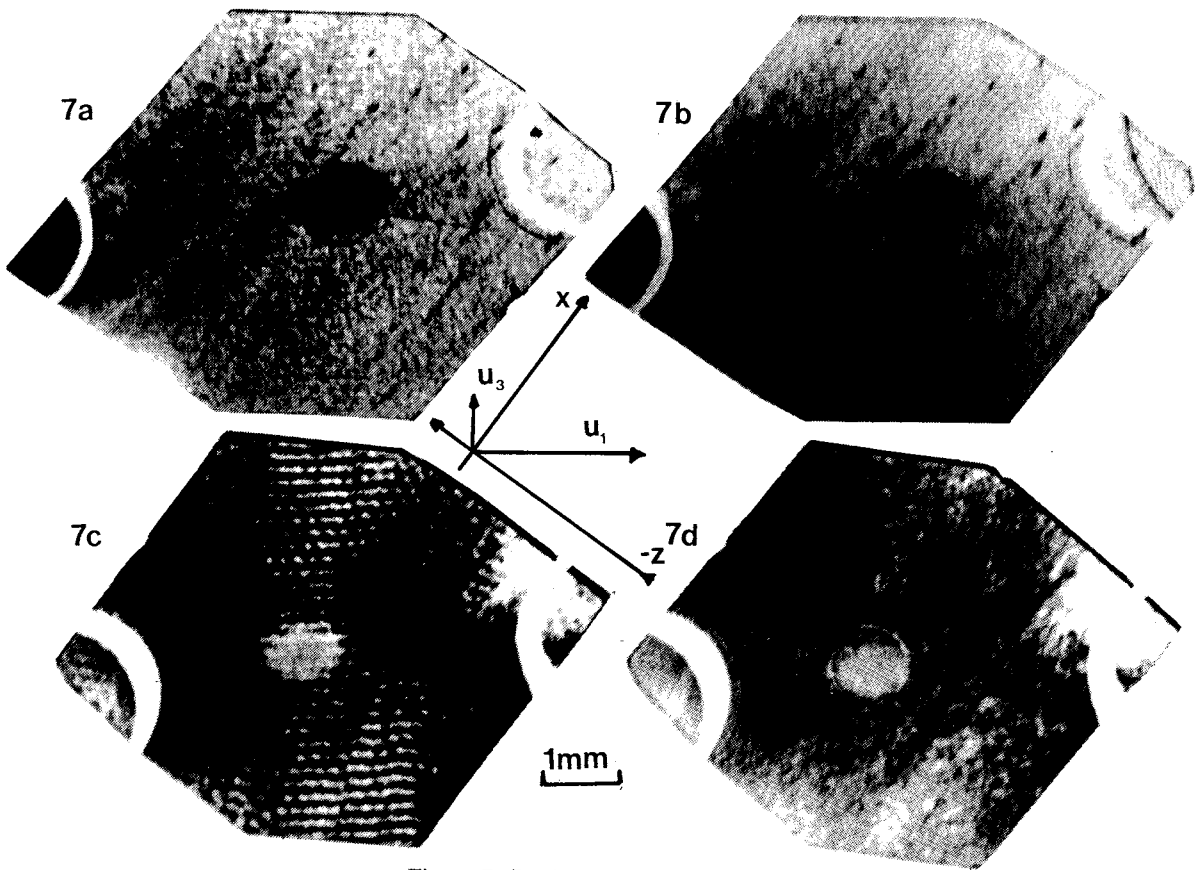


Figure 7: Stroboscopic topographs.

On figures 7c and 7d are displayed respectively a stroboscopic and a time integrated topograph representing a linear combination of the two shear components. Again it appears that the plate coupled mode is a progressive one. In this case (corresponding to the same resonator as in figures 7a and 7b), it appears that the second progressive component has a very large amplitude.

On figure 8 are represented 3 stroboscopic topographs that display the following components of the vibration mode

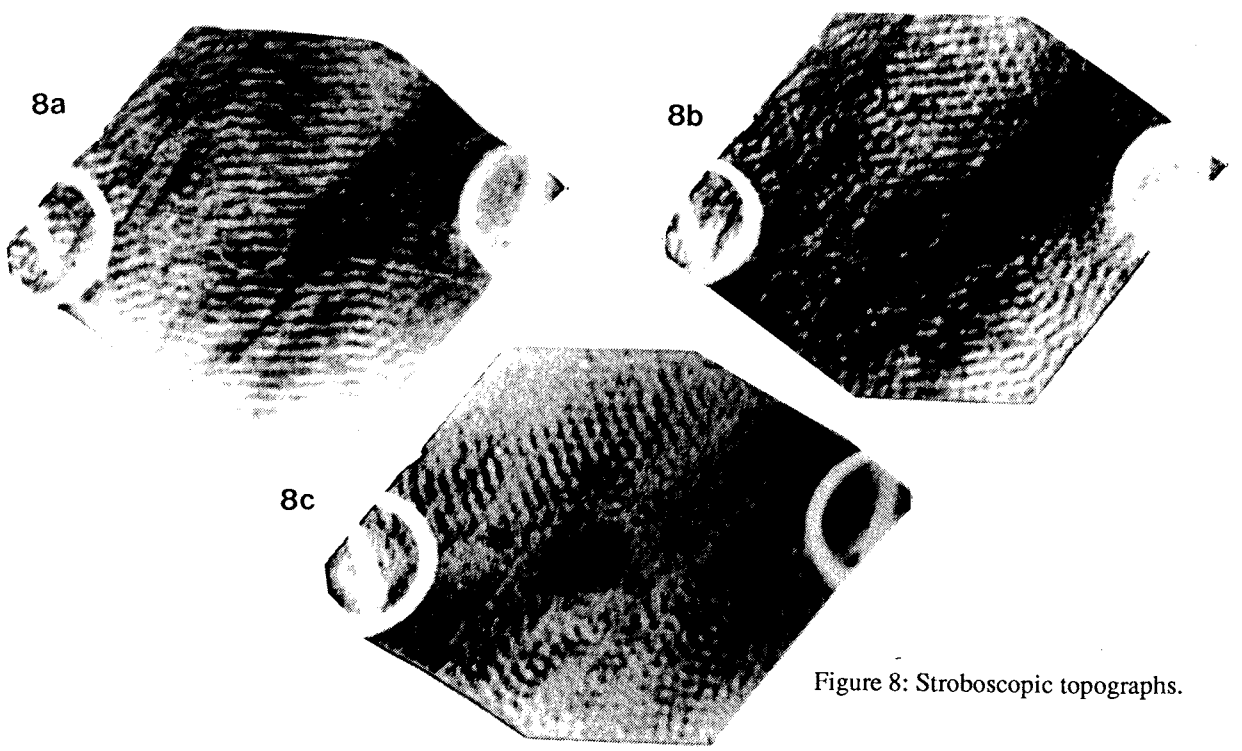


Figure 8: Stroboscopic topographs.

- On figure 8a is displayed a linear combination of the u_2' and u_3' components of the mode (mostly u_3').
- On figure 8b is displayed a linear combination of the u_1' and u_2' (mostly u_1').
- On figure 8c we have a linear combination of the three components of $\vec{u}$.

[Where the u_j' are the projections of the observed mode on the three (orthogonal) eigen vectors representing the displacement (polarization) of the three one dimensional thickness modes of an X plate (the axis x_1', x_2', x_3' corresponding respectively to the displacements of the B, A, and C modes].

From these observations it can be inferred that the coupled plate mode(s) have in this case 3 non negligible components. Other observations made on about 10 resonators with different designs, have indicated that the u_2' and u_3' components are, in all the observed cases, progressive waves. Also it is possible to observe that the lateral anisotropy of the stationary part of u_1' (which corresponds to the usually considered vibration mode of a trapped energy resonator using the B mode) is non negligible. The trapped part of the mode extends more on the direction x_1' of its polarization than on x_3'. As indicated by the theories (9) (10) (11), this component displays an elliptical shape.

One important thing to notice is the very large amount of elastic energy relative to the progressive (but with a large standing wave ratio) components. From the plate theories established by Professors Mindlin, Tiersten and Lee, it is believed that these progressive components can become resonant (standing waves) at some frequencies but probably in a system including more that the plate (plate + mounting clips ?+ holder ? + ...?). We believe also that, when they are resonant, they are the cause of the activity dips and of the parasitic responses observed in many cases near the resonance of the B mode. To get further insight on this question, an absorbing compound was put on the periphery of an X cut resonator to absorb at least in part the plate waves. On figures 9a and 9b are represented the electrical responses of the resonator without absorbant and with absorbant. On figure 9c, the difference between these

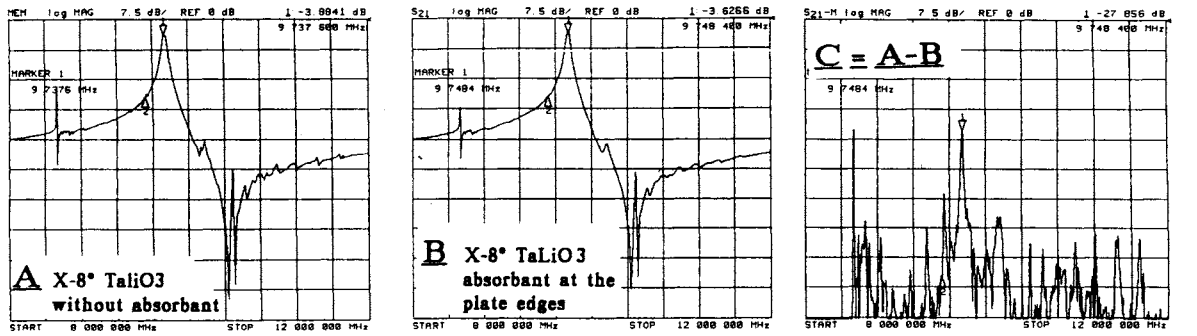


Figure 9: Effect of an acoustic absorbant.

two responses is displayed. The obtained curve, which is a representation of the absorbed energy as a function of frequency (to be corrected proportionally to S_{21} of figure 9a) has a discrete structure. In this Figure each peak of energy absorption corresponds much probably to a resonant coupled plate mode.

III.2 - Miniature bar shaped thickness shear resonators

In the present state of the theory it seems very difficult to find a mean to avoid any coupling to plate modes for the X cut resonators of conventional design; principally because the very large coupling coefficient leads to have a large amplitude of vibration over a very large frequency range.

If one of the lateral dimensions of the plate is greatly reduced, the spacing of many of the plate modes can be enlarged so that the probability of coupling can be reduced. An experimental investigation of such bar resonators was made. Our experiments have confirmed (12) that it is possible by a choice of the lateral dimensions of the bars to achieve greater Q factors than with conventional X cut resonators. A gain of a factor about four was achieved. The shape of such resonators is indicated in figure 10. We have chosen to put the length along the x_1' axis previously defined to have rectangular electrodes with a lateral anisotropy similar to that observed experimentally. This choice was also dictated by the observation in plates of coupled modes looking like flexure modes with their nodal lines in the x_1' direction.

During this investigation it was observed that mode coupling can occur near the resonance of the main mode for improper width (Figures 11a and 11b), that, above a certain value, the length is not an important parameter, and

also that the anharmonic spectrum of the resonator remains mainly dependant of the electrode dimensions (Figure 11c). The maximum observed Q factors were in the range of 6000. The experimental observations have suggested that much better values of the Q factor could be obtained using some modification of the geometry. A theoretical analysis of the vibration modes of such devices is currently being made using the latest developments of the theories of piezoelectric plates. On the whole, this type of resonators have the interest to give enhanced Q factors and also to use much less material by resonator function than conventional resonators.

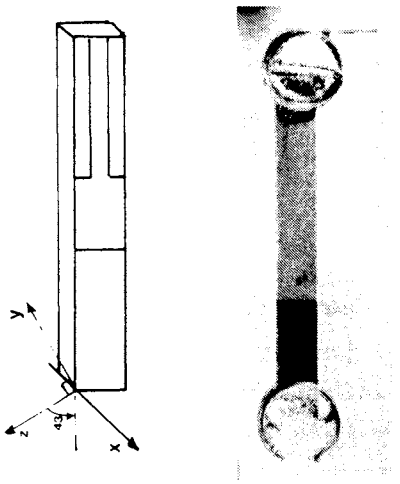


Figure 10: Bar shapped resonator.

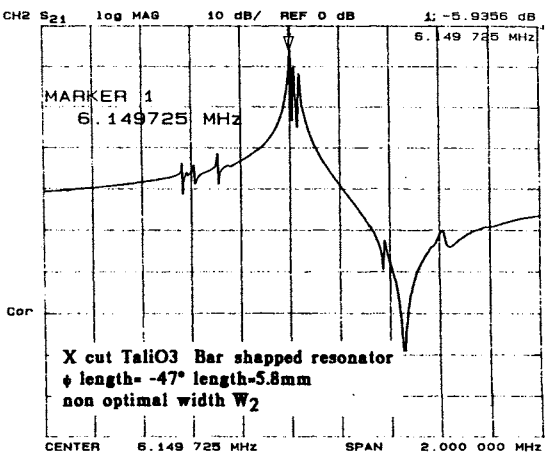


Figure 11a

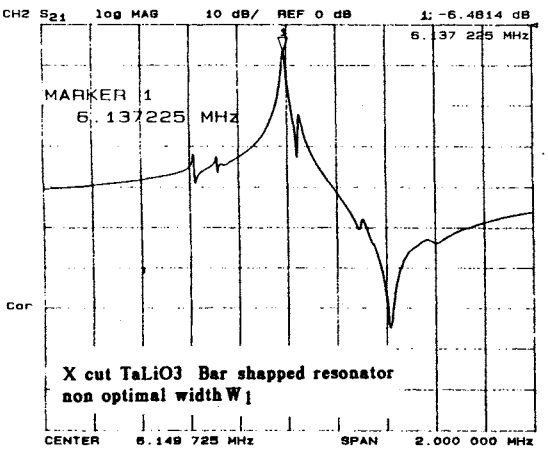


Figure 11b

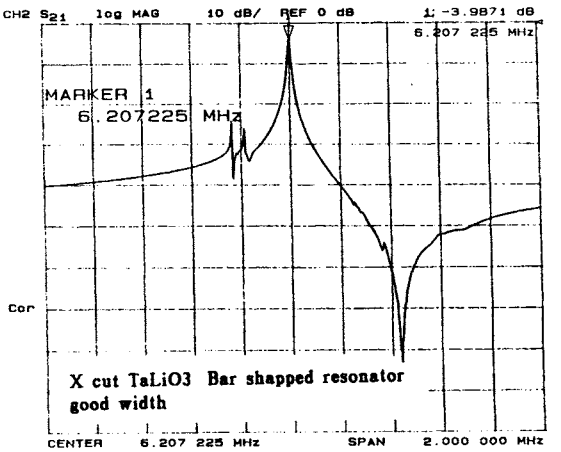


Figure 11c

IV - DEVICE USING OVERTONE MODES

The overtone modes of cuts with a high coupling coefficient display a much higher Q factor than the fundamental mode. They have an "effective coupling coefficient" that can reach 15 %. We have previously shown that for conventionally designed resonators, a thermal compensation can be obtained using doubly rotated cuts (4). Other designs can be found to obtain overtone resonators with a zero F.T.C., several will be discussed in the following.

IV.1 - Doubly rotated cuts with zero T.C. for the overtone

The properties of cuts near the second locus of zero T.C. of the B mode were examined. It was noticed that quite large coupling coefficients can be obtained. But since the three thickness modes are excited in these cuts they are not very favourable for filter applications, unless a method to eliminate certain of these modes is applied. In figures 12 and 13 are represented experimental and computed thermal behaviours for several cuts belonging or near this locus.

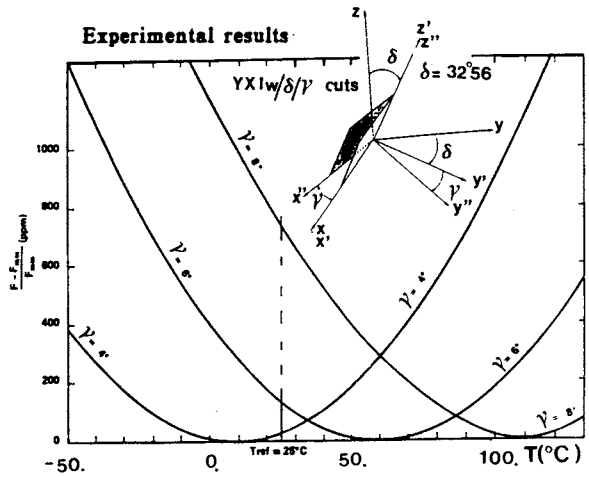


Figure 12: Thermal behaviour of doubly rotated cuts (Experimental).

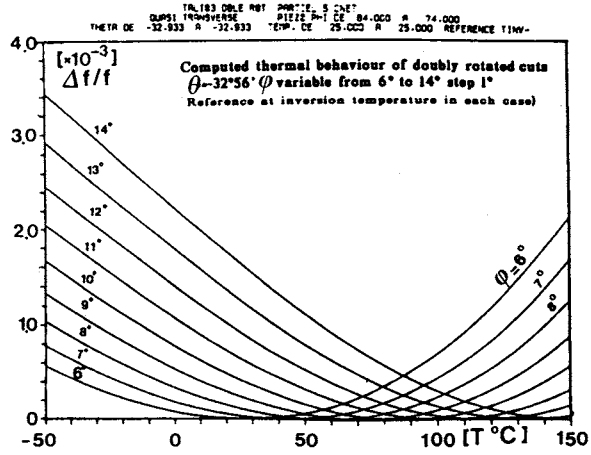


Figure 13: Thermal behaviour of doubly rotated cuts (computed).

IV.2 - Overtone resonators using a T.C. compensating layer

This type of compensation is mostly used for surface wave devices (13) for which it is difficult to find materials having simultaneously a zero T.C. and a large coupling coefficient. In the case of high frequency bulk wave resonators this method is also interesting since the thickness of the compensating layer is compatible with the existing techniques

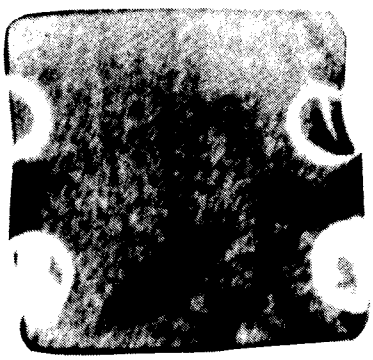


Figure 14:One resonator excited.

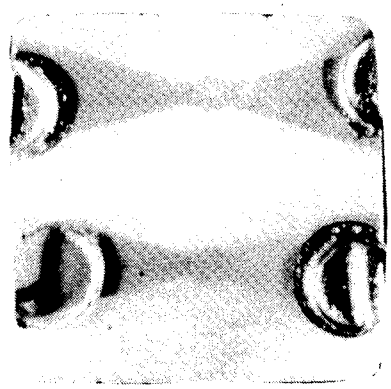


Figure 15:Integrated filter.

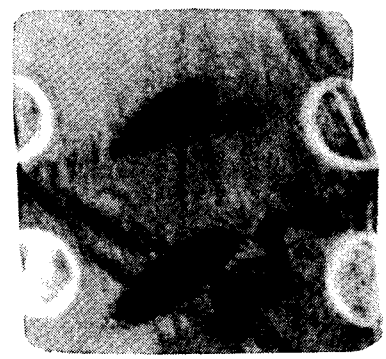


Figure 16:Filter excited.

of film deposition. The most appropriate material for a compensating layer is the amorphous silica (and also probably the amorphous aluminium phosphate). The largely positive T.C. of silica (70 to 80 . $10^{-6}°C^{-1}$ for the shear waves) is sufficient to compensate the moderately negative T.C. of the third or of the filth overtone of cuts of lithium tantalate with high coupling coefficients. Again in this case the most appropriate cuts are those around the X cut which have a first order T.C. around -40 $10^{-6}°C^{-1}$ for the third overtone.

Properties of the third overtone of the X cut.

The Q factor of the third overtone mode of X cut resonators is much higher than that of the fundamental mode (in the range 10^4 to 510^4) and the response presents much less coupling to plate modes. It was found using stroboscopic X ray topography that the lateral anisotropy of this mode is much greater than for the fundamental mode (Figure 14). The very small lateral dimensions of the electrodes required to trap no anharmonics make possible to integrate several resonators on the same plate (Figure 15). In this case to avoid any coupling betwen the resonators by the plate mode (Figure 16) it is necessary to use some form of acoustic isolation between the resonators.

An example of electrical response of resonators integrated on one plate is given in Figure 17a the response of a 2 Poles Jaumann filter using similar integrated resonators is given on Figure 17b. The 3 dB bandwith of this filter is 150kHz.

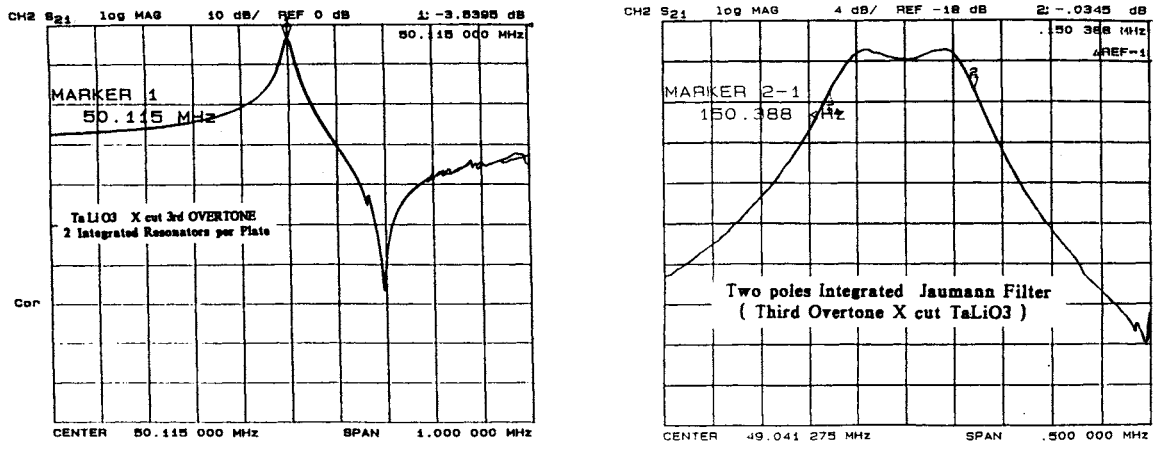


Figure 17: Response of 2 integrated resonators (a) Response of an integrated filter (b)

IV.3 - Modelisation of composite compensated resonators

As it was observed (Figure 14 and 16) that, when no coupling to plate modes are present, the vibration mode of overtone resonators is extremely confined, the analysis was conducted assuming a one dimensional propagation

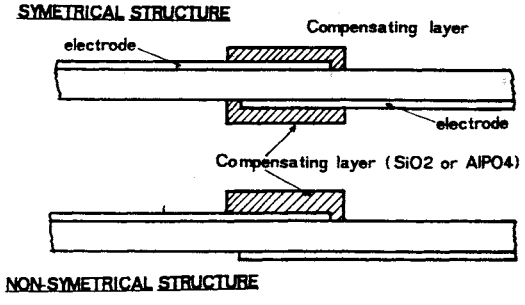


Figure 18: Compensation by a Silica layer.

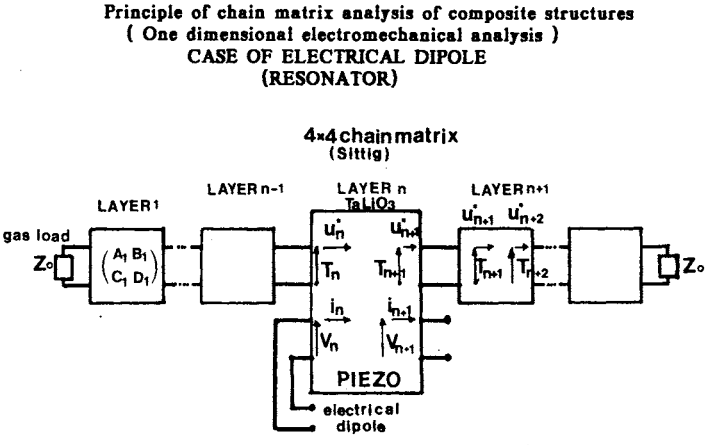


Figure 19: 1 D. Modelization of composites.

in the different layers (compensating material, electrodes, lithium tantalate) constituting the composite resonator (Figure 18). This assumption is sufficient for the purposes of finding a compensated design and also to have approximate values of the equivalent scheme.

The model can take into account all the layers constituting the resonator and the effect of gas loading (by viscosity for the shear modes) on the external surfaces. It uses exacts solutions of one dimensional piezoelectricity for the piezoelectric materials and of elasticity for the electrode layers and the compensating material. The acoustic and dielectric losses in any material are taken into account using imaginary parts for the constants (C or ε) of the material. The model uses a chain matrix formalism (Figure 19) based upon the electromechanical chain matrix (4x4) proposed by Sittig (14). The obtained result is the electrical impedance of the resonator, from the impedance computed as a function of frequency, the zero phase resonance frequencies are searched for by the program. Since the temperature coefficients of all the quantities are considered, it is possible to compute the temperature coefficients of the resonance frequencies (This is done calculating the resonance frequencies for 5 to 10 temperatures and then by extraction of

the temperature coefficients by a polynomial least square fitting). The corresponding computer program was used to analyse the most interesting structures for our application. It was found that it is possible to consider either the symmetrical or the non symmetrical structures. The non symmetrical structures have the property to have a strong resonance (if the non symmetry is sufficient) on the second overtone (and more generally on even overtones). In this case, it is appeared that the T.C. of the lithium tantalate layer is compensated by a small layer of silica. The symmetrical structures operating on the third overtone have the interest to permit the obtention of very large bandwith (large resonance-antiresonance spacing) two examples of calculated response for composite resonator are given on figures 20 and 21.

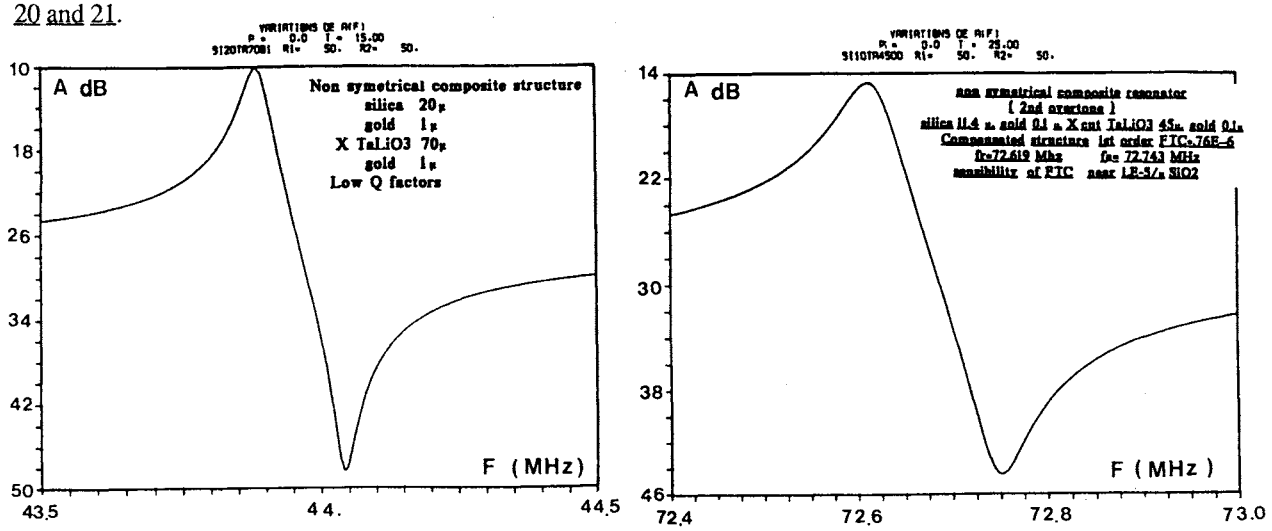


Figure 20 and 21: Examples of computed responses for composites.

IV.4 - Overtone resonators using the lateral excitation

A. Ballato (15) (16) (17) has demonstrated the interest of using the lateral excitation to obtain severall interesting properties with the thickness mode resonators (excitation of modes which are non-piezoelectric using the conventional design, selection of only one thickness mode in a cut etc...). Due to the high values of the dielectric constant and of the coupling coefficients (the lateral coupling coefficients are also very high), the use of lateral field excitation is more interesting in lithium tantalate than in any other known compensated material. The principle of lateral field excitation is recalled in figure 22. Different types of resonators were made using different design to obtain an energy trapping. Very interesting results have been obtain using plano-convex resonators. In table II are given the measured characteristics of a resonator using the B mode of the Y-32°56° cut. This cut is very close to the second locus of zero T.C. for the B mode. As can be observed in figure 23 this mode is not excited with an electric field normal to the plate.

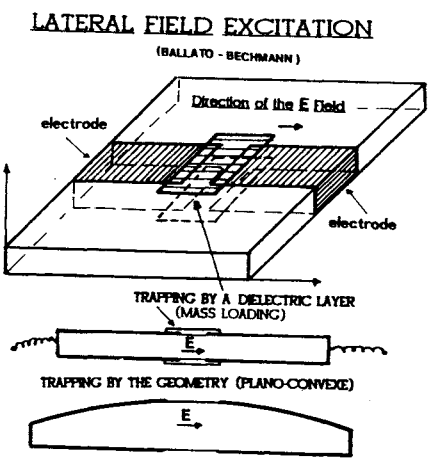


Figure 22: Lateral Field excitation.

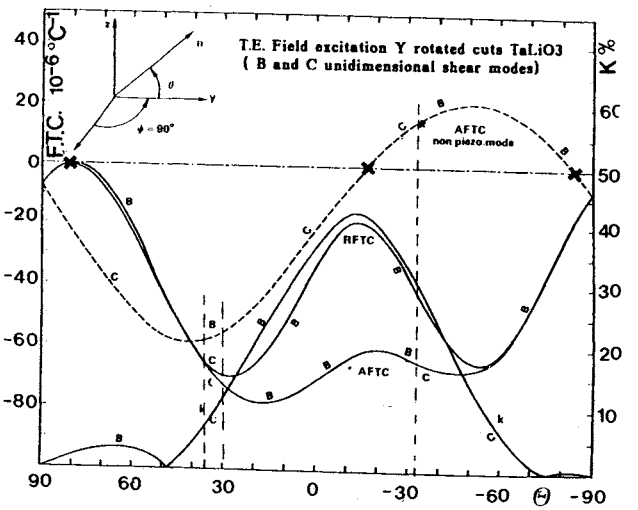


Figure 23: Properties (T.E.) of Y rotated cuts.

TABLE II

-32°56' PLANO-CONVEX RESONATOR (L.E.) 2h=1,5mm R = 150 mm

Overtone	Frequency (Hz)	Q factor	Q at 5 MHz	R (Ohm)	L (Henry)
1	1.321.625	9130	-	144.	-
3	3.614.263		(untrapped ?)	-	-
5	6.375.086	149.000	190.000	38.35	0.142
7	8.936.515	191.000	342.000	27.01	0.092
9	11.447.612	107.000	245.000	87.32	0.129
11	14.012.703	471.000	1.310.000	16.5	0.088
13	16.562.111	576.000	1.908.000	18.4	0.101
15	19.089.865	262.000	1.003.000	38.35	0.084
17	21.639.792	489.000	2.120.000	30.3	0.108
19	24.190.249	343.000	1.663.000	52.40	.118
21	26.719.321	347.000	1.856.000	73.98	0.133
23	29.266.075	457.000	2.675.000	51.70	.128
AT QUARTZ 5TH OVERTONE (T.E.)					
5	5.000.000	2.400.000	2.400.000	70	5.6

In table II we observe that overtones with very high ranks have exceptionally large Q.f products and have also very favourable characteristics for use in oscillator circuits. Their impedance level, even for the 23th overtone are much more favorable than the 5th overtone of an AT quartz plano-convexe resonator. These measurements are a confirmation of the observations made in the GHz frequency range which have indicated extremely low acoustic propagation losses for this material and an intrinsic Q.f product higher than 30 10^{12} (3).

Several response curves obtained for the overtones of one Y-32°56 resonator are shown on figure 24. It was also observed that these resonators were able to withstand extremely high excitation levels before that the electrical response shows a sligth form of non linear behaviour. All these properties make them very attractive to obtain, by direct generation in the range from a few MHz to over 150 MHz, oscillators with outstanding phase noise properties (18). The response curves for different excitation levels of the overtone of a Y-32°56' resonator are displayed in figure 25. Other preliminary experiments with L.E. plane resonators, using the z cut, have indicated the possibility to have a sufficient "effective coupling coefficient" to obtain the required bandwidth for the I.F. filter.

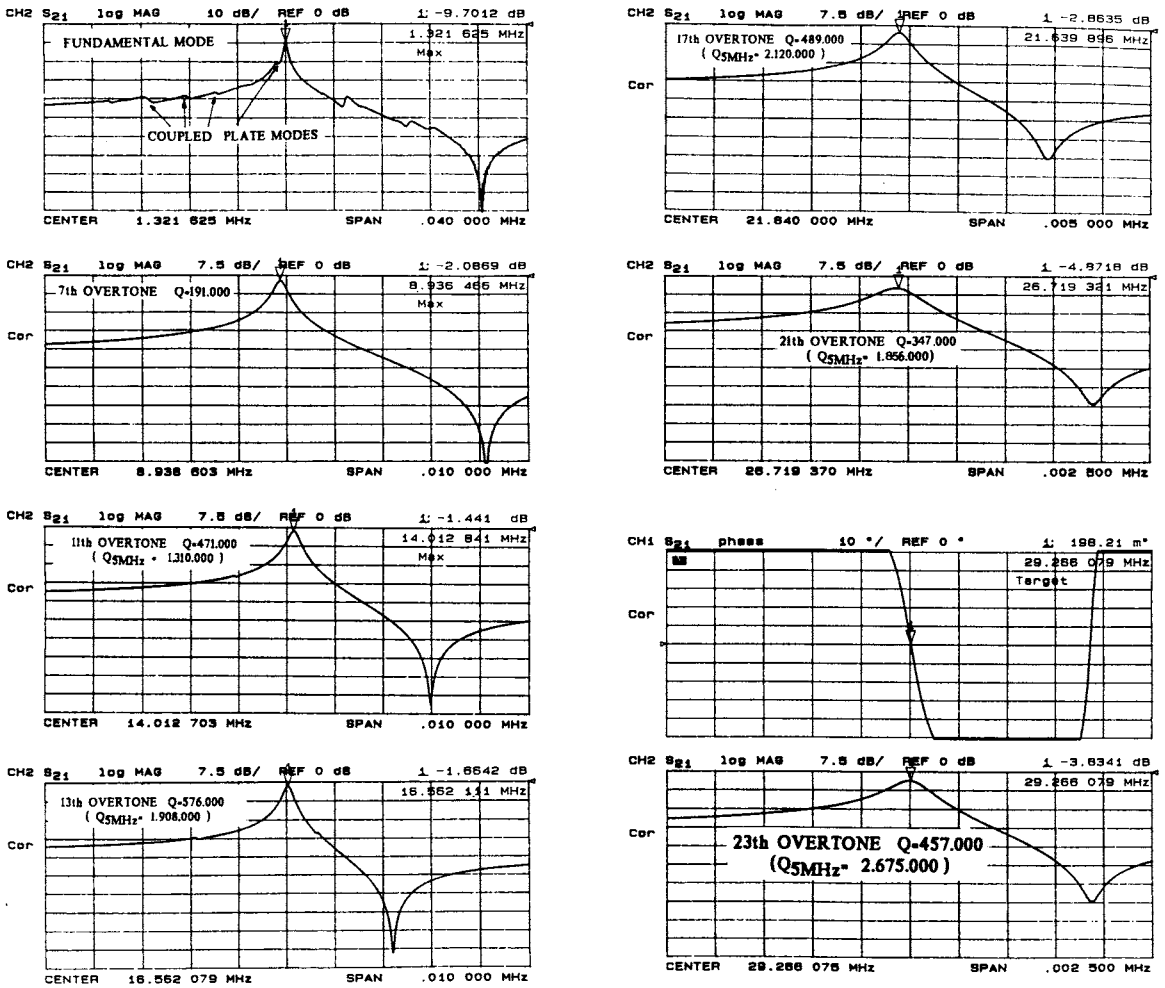


Figure 24: Experimental results for a plano-convexe resonator.

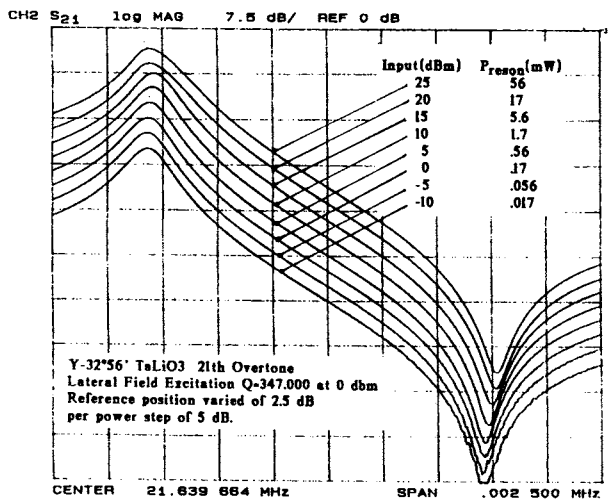


Figure 25: Influence of the excitation level.

V - CONCLUSION

Several conventional and new types of resonators using lithium tantalate where investigated in view to define the most interesting solutions, using this material, to make I.F. filters having the characteristics required by the new numerical pan european radiotelephone system. Besides several conclusions relative to this specific question, these investigations have revealed many new features of the considered devices.

Amoung them, the most important points are :

- The Q factor limitation by couplings to dissipating plate modes of the devices using a conventional design and the fundamental mode

- The possibility to obtain high frequency overtone resonators and integrated filters with a zero T.C. using composite structures.

- The major interest of lateral field excitation in devices using this material and the outstanding characteristics obtained with overtones of very high ranks.

This study has confirmed that due to its energy trapping characteristics, lithium tantalate is particularly adapted to obtain extremely miniaturized integrated devices and also that several new interesting solutions to realize the I.F. Filters for the G.S.M. radiotelephone system can be implemented using this material.

ACKNOWLEDGMENTS

The authors are indebted to A. Ballato for the revelation of the great interest of the lateral field excitation in lithium tantalate. They acknowledge the skilful assistance of J. Douis for cutting and polishing the many samples used in this study and thanks particularly M.P. Louis for the preparation of this manuscript.

REFERENCES

(1) A. Ballato Doubly rotated thickness mode plate vibrators in Physical Acoustics Vol XII Academic Press New York 1977

(2) J. Détaint, R. Lançon Proc. 30th. Annual Freq. Control Symp. 1976

(3) R.A.Moore, J.T.Haynes, B.R.McAvoy Proc. 1981 Ultrasonics Symposium p 414 (1981)

(4) J. Détaint Electron Letters vol 13 n° 1 p 20 1977

(5) J.W. Burgess M.C. Hales Proc. IEE Vol 123 n° 6 June 1976

(6) M. Onoe Proc. IEEE Vol 57 p 702 (1969)

(7) A. Zárka, B. Capelle, J. Détaint, J. Schwartzel Proc. 41st Annual Frequency Control Symp. 1987 p. 236-240

(8) B. Capelle, J. Détaint, Y. Zheng, J. Schwartzel, A. Zarka 4 th European Frequency and Time Forum (these proceedings)

(9) D.S. Stevens, H.F. Tiersten J. Acoust. Soc. Am Vol 79 n° 6 p 1811 (1986)

(10) K. Nakamura, M. Shimizu Proc. IEEE Ultrasonics Symposium p 606 (1976)

(11) J. Détaint, J. Schwartzel, C. Joly, E. Philippot Proc. 41th. Annual Frequency Control Symp. p 314 1987

(12) Y. Fujiwara and N. Wakatsuki Strip. IEEE Trans U.F.C.C vol UFCC34 n° 1 p 39-44 (1987)

(13) G. Cambon, J. Lassale, J.M. Saurel, J. Attal Proc. 1981 Ultrasonics Symposium p 364

(14) E.K. Sittig Physical Acoustics vol IX Academic Press 1972

(15) A. Ballato, J. Kosinski, M. Mizan, T. Lukaszek Proc. 43rd Annual Frequency Control Symp. p 557 (1989)

(16) E.R. Hatch, A. Ballato Lateral Field excitation of quartz plate Proc. 1983 IEEE Ultrasonics Symposium p 512 (1983)

(17) A. Ballato, E. Hatch, M. Mizan, B.H.T. Chai, R.S. Tilton, T.J. Lukaszek Proc. 38th Annual Freq. Control Symposium (1984)

(18) MM. Driscoll, C.R. Vale, R.W. Weinert Proc. 1987 IEEE Ultrasonics Symposium p 347 (1987)

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-04 Number 4

CHEMICAL ETCHING OF BERLINITE

O. CAMBON*, A. GOIFFON*, E. PHILIPPOT*, J.P. AUBRY**, J. DETAINT***

* URA D0407, U.S.T.L., Place E. Bataillon, 34095 Montpellier, France

** C.E.P.E., 44, Avenue de la Glacière , 95100 Argenteuil, France

*** C.N.E.T., PAB/BAG/MCT, 92220, Bagneux, France

I - INTRODUCTION

The problems of berlinite crystal growth and improvement of crystal quality being quite carried out (1,2) the question of the moment is the manufacturing of high frequency berlinite resonators. Indeed, previous works have shown that this material is very attractive specially in the field of high frequency resonators (1a,1d,2,3).

Mechanical lapping cannot be envisaged for manufacturing high frequency, and hence thin resonators, and ionic lapping is too much expensive. The most suitable method is therefore chemical lapping when it's possible. For this reason, a lot of works have been published on the chemical etching of quartz, (4 to 12), and in particular on AT cut, which is the most commonly used type of temperature compensated cut. No work had been undertaken on berlinite, which possesses the same temperature-compensated orientation (13). In the light of the potential interest mentioned above, we investigated chemical etching of this substance.

The work presented here was first restricted to AT cut study. Sulphuric acid was chosen as etching solution as it is the growth medium that had governed the best previous results (1c,d). In order to determine optimal conditions for chemical etching in this acid, we systematically measured the influence of temperature and acid concentration on etching rate and monitored the variations in the texture of the surface in relation to these two parameters by roughness measurements and S.E.M.

II - EXPERIMENTAL PROCEDURE

II-1 - Samples and etching system

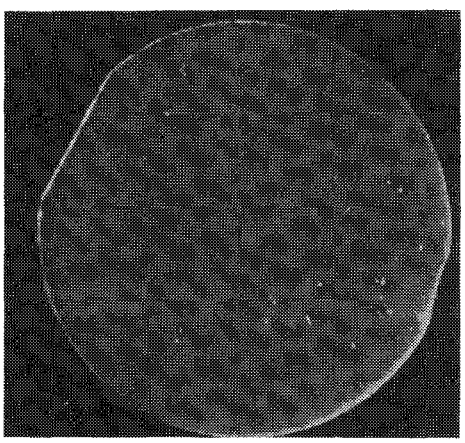


Photo 1 : AT plate of berlinite before chemical etching (5mm diameter, 150 µm thickness)

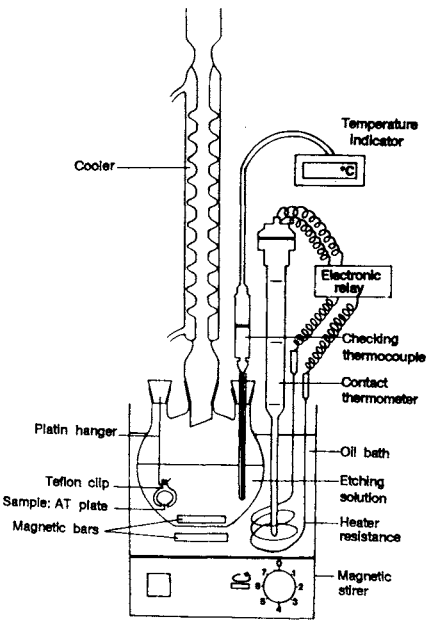


Figure 1 : System used for etching plates

Chemical etching was carried out on AT wafers cut in berlinite crystals obtained at the laboratory in sulphuric acid medium by the vertical temperature gradient method. Cutting in a specific direction, grinding and initial polishing were carried out at Thomson CEPE, Argenteuil. The samples were in the form of discs 5 mm in diameter (Photo 1) and etching is made in the system shown in figure 1.

II-2 - Checking of etching

a) Measurement of plate thickness

The decrease in thickness is followed by measuring the resonance frequency of AT plate : $F = K/e$ with $K = 1.48$ MHz.mm^{-1} (value calculated from constants of ref. 3).

b) Surface characterization

Roughness is checked using a perthometer W5B surface profilometer along Z' axis (projection of Z axis in AT plane).

SEM was made using a Cambridge S 360 microscope at low voltage (1 kV).

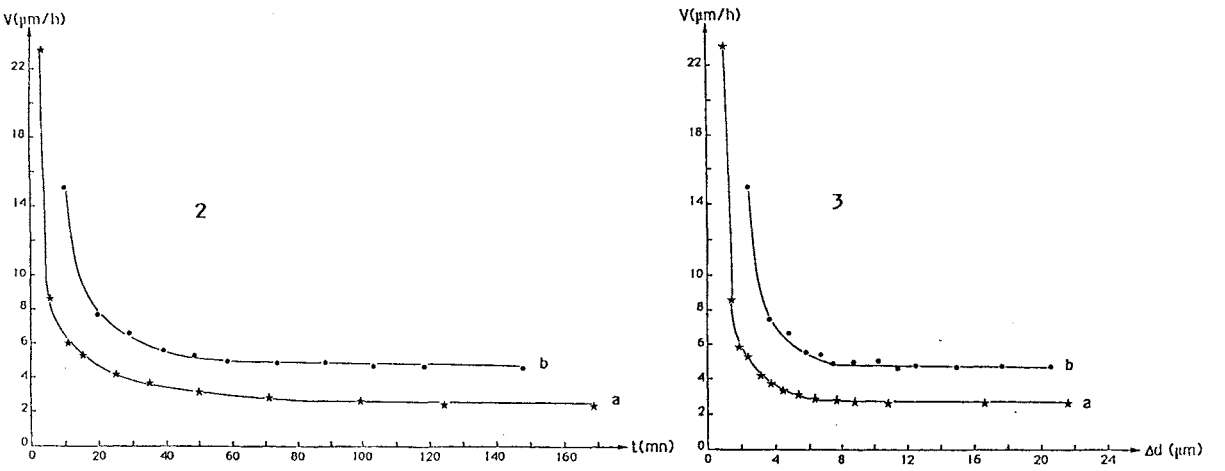
III - RESULTS AND DISCUSSION

III.1 - Kinetic studies

a) Evolution of etching rate versus time and etching depth

During this process, temperature and acid concentration were kept constant. The etching rate was defined by the ratio $\Delta d/\Delta t$ where Δd is the total thickness decrease on the two faces during the time Δt.

We present in Figures 2 and 3 the results obtained during etching at 80°C for two concentrations of sulphuric acid (14 M and 6 M). Other temperatures and acid concentrations were studied and give similar results.



Figures 2 and 3 : Etching rate (V) vs time t (figure 2), vs etching depth Δd (figure 3) of berlinite AT plates in sulphuric acid solutions at T = 80°C

a - H_2SO_4 6M b - H_2SO_4 14M

In Figure 2, the etching rates are given in terms of time. We note that the etching rate decreases quickly and becomes constant at two different times for the two acid concentrations.

The Figure 3 presents etching rates in terms of etching depth Δd. It appears that etching rate

stabilizes at a same etching depth for the two acid concentrations.

All the curves obtained can be divided in two parts :

- the first part corresponds to rapid decrease in etching rate. It can be assumed that this phase consists of etching of the surface layer disturbed by cutting and polishing,

- it can be seen in the second part that the etching rate stabilized. It might thus be the characteristic intrinsic rate of chemical etching of a given orientation of the material.

Thus, on Figure 2 the etching rate becomes constant at different times because the etching rates are different for the two acid concentrations and Figure 3 shows that the depth of the disturbed layer is approximatively 3 µm per face.

It is known that the thickness of the disturbed layer depends on the previous treatment to which the sample has been subjected (type of cutting, type and particle size of the abrasive used, polishing time etc). Studies carried out on quartz have shown that the thickness of this layer is proportional to abrasive particle size, with the proportionality factor varying from 0.2 to 0.8 d (d = diameter of the particles of abrasive) depending on the quality of quartz.

In the case of berlinite, the thickness of 3 µm that we determined also corresponds to 3 µm abrasive. As no other measurements has been available to date, we can only say that under the conditions used the thickness of the damaged layer is similar to that of the particles of abrasive.

b) Influence of temperature on etching rate

We measured the etching rate in relation to temperature for different acid concentrations.

The results shown in Figure 4 correspond to etching rates measured after removal of the disturbed surface layer 3 µm thick. These rates were thus intrinsic etching rates of the material.

Parabolic type increase in rate was observed when temperature increased whatever the acid concentration. This behaviour is similar to that observed during the etching of quartz by NH_4HF_2 and shows that the etching rate was considerably influenced by temperature and that this parameter should therefore be carefully checked during the frequency adjustment of a resonator. It can nevertheless be noted that the plot for the highest acid concentration used displays a slightly different curve, indicating less marked variation of etching rate with temperature.

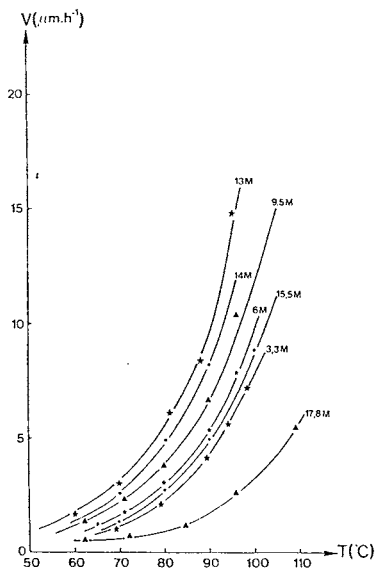


Figure 4 : Etching rates of berlinite AT plates in terms of temperature for various sulphuric acid concentrations.

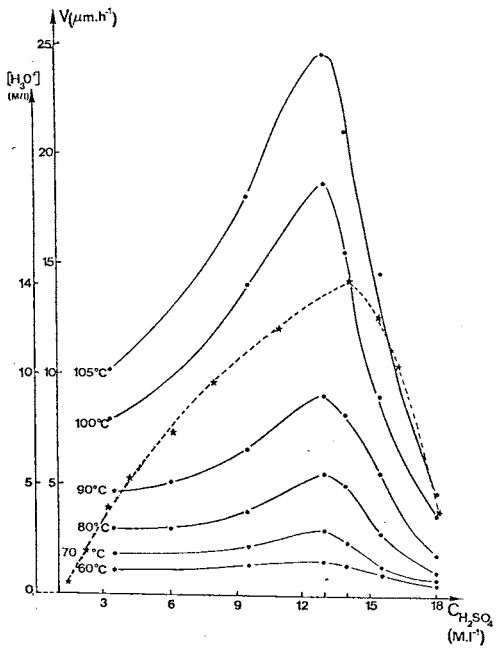


Figure 5 : Etching rates of berlinite AT plates vs sulphuric acid concentration for different temperatures (solid lines). Dotted line shows H_3O^+ concentration from YOUNG (14)

c) Influence of acid concentration

The network of curves plotted in Figure 5 was obtained when etching rates were plotted against sulphuric acid concentration at different temperatures. It was observed that when sulphuric acid concentration increases, etching rate also increases at first, reaches a maximum for a sulphuric acid

concentration of ≃ 13 M and then decreases whatever the temperature. It can be seen however that the rate variation is more marked when the temperature rises.

Although the etching mechanism of berlinite has not been yet elucidated, it can be supposed that as in the case of quartz, the breaking of M-O bonds (M = Al or P) occurs when the latter has been weakened after the formation of OH bonds caused by the presence of protons in solution. If it's the case, the rate should increase with the protons content ; this is what was observed. Indeed, YOUNG (14) measured the concentration of the different species in sulphuric acid solutions and if it is considered that $[H^+]^*$ = $[HSO_4^-]$ + 2$[SO_4^=]$, the proton concentration is maximum when molarity is around 13 M at 25°C. When the temperature increases, the ionization constant decreases in relation to HSO_4^- (15) ; however, the contribution of this ion is very small in comparison with that of the first acidity.

It can be considered as a preliminary approximation that the variation in proton concentration obeys the same variation law at higher temperature and thus accounts for the variation observed in rate. This hypothesis remains to be confirmed by the use of other strong monoacids (e.g. HCl and HNO_3) which would make possible to determine this proton content in a simple manner. We intend to do this in the near future.

d) Etching rate law

. Influence of hydronium concentration :

If the etching rate can be expressed by $v_T = k_T [H^+]^n$ relation, then $\log v_T = \log k_T + n \log [H^+]$ and the plot of $\log v_T$ vs $\log [H^+]$ must be a straight line.

In Figure 6, we have plotted the logarithm of etching rate versus the logarithm of proton concentration.

We can check that, for all temperatures, we obtain linear plots with the same slope (n) but different values of intercept with ordinate axis (k_T).

The slope, n, is the reaction order with respect to hydronium and can be computed as n = 0.5.

These first results seem to confirm that our hypothesis of etching rate in the form $v_T = k_T [H^+]^n$ is valuable with n value of 0.5.

* However proton is solvated, we write H^+ for simplification.

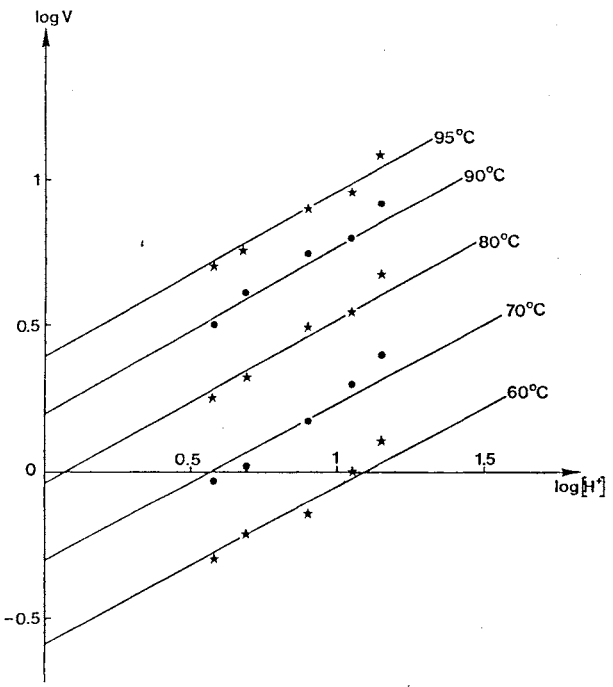


Figure 6 : Etching rate logarithm, vs logarithm of proton concentration for various temperatures

. Influence of temperature :

If we plot k_T in terms of temperature T (Figure 7), that curve shows an exponential variation and so k_T must follow an Arrhenius type evolution : $k_T = Ae^{-Ea/RT}$ where Ea is the activation energy of dissolution reaction and A the frequency factor.

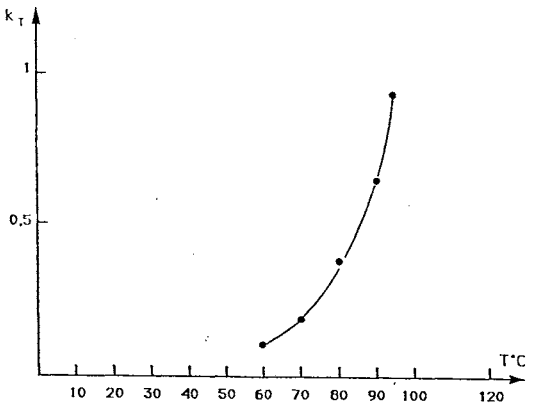


Figure 7 : Variation of k_T (etching rate constant) in terms of temperature

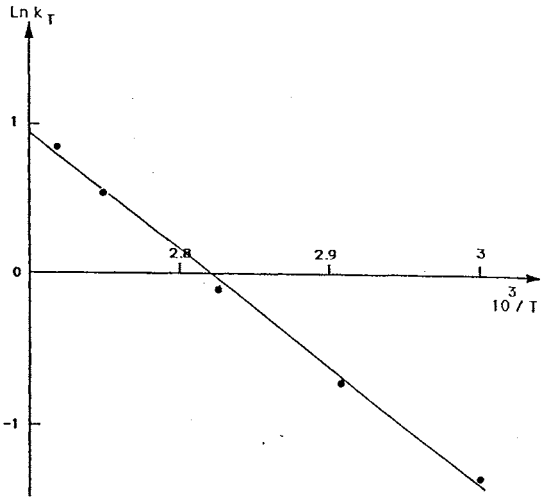


Figure 8 : Plot of Lnk_T vs $10^3/T$

If it is true, then Ln k_T = Ln A - Ea/RT. Figure 8 reports Ln k_T versus $10^3/T$. The observed linear variation is in agreement with the previous equation. From the slope of this line, the value of Ea is calculated to be 63.4 kJ.mole^{-1}.

The term A is determined by y-axis extrapolation to a value of A = $2.2.10^9$.

. Etching rate law :

Thus, we can write the etching rate expression of berlinite AT plates in sulphuric acid as :

$$V = (2.2.10^9 \, e^{-Ea/RT}) \, [H^+]^{1/2} \quad \text{with Ea} = 63.4 \text{ kJ.mole}^{-1}$$

So, our assumption seems to be checked by experimental results in sulphuric acid solutions. Nevertheless, this hypothesis needs other examinations in other acid solvents, specially with monoacids.

III.2 - Evolution of surface texture

a) Roughness measurement

In Figure 9, variation of the roughness parameter Ra is plotted in relation to the decrease in thickness for etching conditions identical to those of Figure 2 (T = 80°C, H_2SO_4 = 14 M). This evolution is representative of all the measurements carried out under other conditions of temperature and acid concentration.

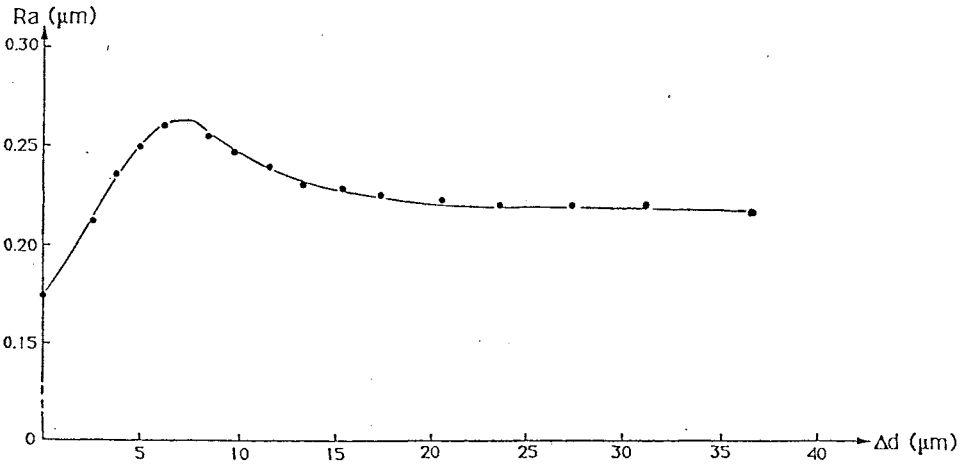


Figure 9 : Roughness parameter (Ra) vs etching depth Δd_s (one face) during etching of a berlinite AT plate.

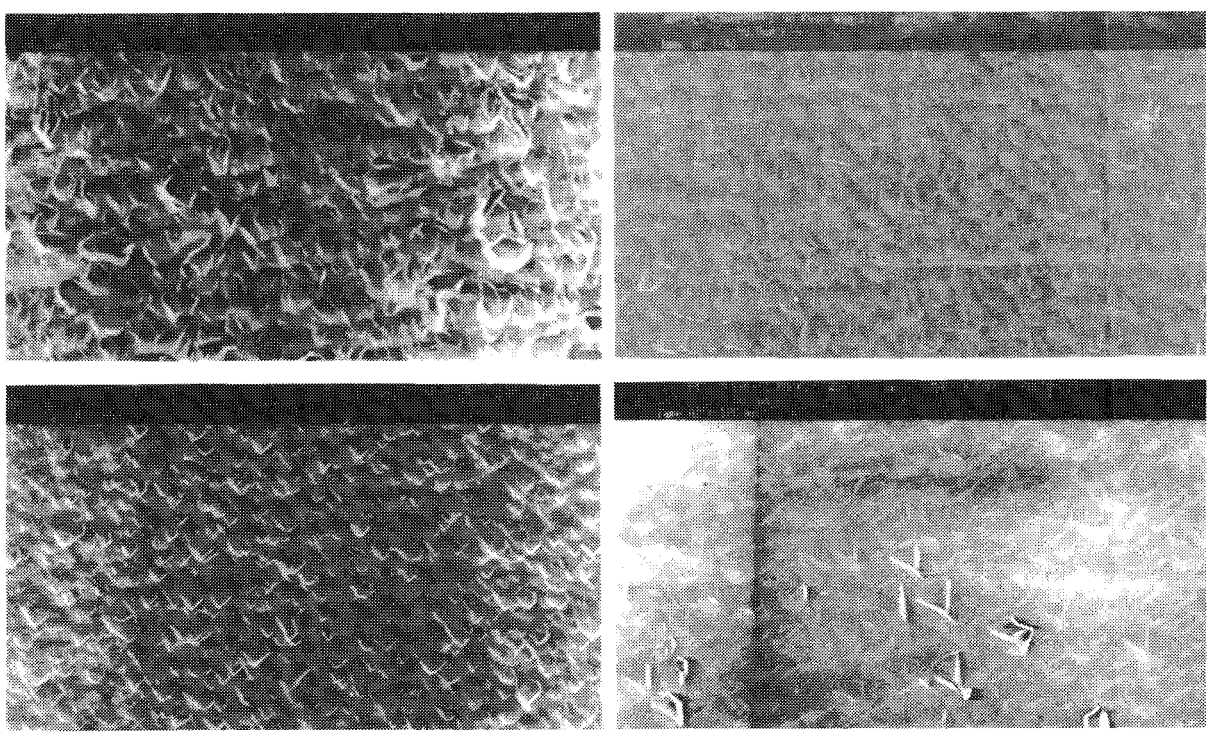
It can be seen that Ra reaches a maximum for an etching depth of about 7 µm and then falls and stabilizes at a value close to 0.22 µm when Δd is close to 16 µm. This value is superior to the initial roughness, which would seem to show that although etching using sulphuric acid is effective for setting frequency it is not necessarily optimal for fixing the surface texture.

This behaviour is similar to that observed for the same orientation in quartz (12e). It should be noted that the depth of etching in relation to maximum roughness is identical to the particular value found in the kinetic study and which can be attributed to the existence of a disturbed surface layer. This hypothesis also appears to be confirmed by observations using electron microscopy.

b) Scanning electron microscopy

During the study presented in micrographs 2, we checked the surface texture of the same AT plate etched by a 14 M sulphuric acid solution at 80°C. Observations were made every 10 minutes, but only the micrographs taken after 10, 30, 40 and 120 minutes of etching are shown here.

Before etching, the surface was formed by numerous disordered hillocks (Photo 2a). After 30 minutes the hillocks had been taken out and were oriented (Photo 2b). This behaviour increases until 40 minutes of etching (Photo 2c) when the maximum of irregularity was obtained (maximum roughness measured). Indeed, chemical etching tended to remove preferentially the weaker disturbed parts at the expense of the crystal bulk and therefore revealed all the defects due to polishing, leading to a maximum Ra value. In a second phase, we then observed the effect of polishing which removed these hillocks by wearing them away (Photo 2d). Nevertheless, it should be noted that this chemical etching in H_2SO_4 produces orientated etching figures with very marked profile, which explain that Ra remains high and even superior to its initial value.



Photos 2 : SEM micrographs during etching of a berlinite AT plate in 14 M H_2SO_4 at 80°C.
a) after 10'etching
b) after 30'etching
c) after 40' etching
d) after 120' etching.

IV - CONCLUSION

This work on chemical etching of berlinite AT plates in sulphuric acid solutions showed that, after removal of the disturbed layer due to previous cutting and polishing, etching rates increase with temperature and vary with acid concentration : they first reach a maximum for 13 M etching solution and then decrease. This behaviour have been correlated to the proton content.

Thus we have been able to determine the law of etching rate of AT plates in sulfuric acid solutions $v = (2.2.10^{9} e^{-Ea/RT}).\ [H^{+}]^{1/2}$ with Ea = 63.4 kJ. $mole^{-1}$.

Nevertheless, surface texture checked by roughness measurement and scanning electron microscopy remains a deceptive result.

This study is to be followed by the investigation of orientations effects in chemical etching in order to elucidate etching mechanism. In other way, surface texture is to be improved either by use of surfactants in sulphuric acid medium either by use of other etching solutions.

Acknowledgment :

The authors wish to acknowledge M.C. Gril (Graduate engineer at USTL Montpellier) in providing the SEM micrographs and Prof. Maurin for many valuable discussions during the course of this work.

REFERENCES

1)a) E. PHILIPPOT, A. GOIFFON, J.C. JUMAS, C. AVINENS, J. DETAINT, J. SCHWARTZEL, A. ZARKA, 42nd Ann. Freq. Control (1988) 138.
b) A. GOIFFON, C. AVINENS, E. PHILIPPOT, Rev. Chim. Miner., 24 (1987) 593.
c) E. PHILIPPOT, A. GOIFFON, C. AVINENS, X. BUISSON, R. ARNAUD, O. BIGNON, A. ZARKA, B. CAPELLE, M.T. SEBASTIAN, 3rd Eur. Freq. Time Forum, Besançon (1989).
d) E. PHILIPPOT, A. GOIFFON, M. MAURIN, J. DETAINT, J.C. SCHWARTZEL, Y. TOUDIC, B. CAPELLE, A. ZARKA, J. Cryst. Growth, (1990) in press.
2) J. HOU, B.H.T. CHAI, Ultras. Symp., (1987).
3) H. WANG, B. XU, X. LIU, J. HAN, S. SHAN, H. LI, J. Cryst. Growth, 79 (1986) 227.
4) K. BRAUER, E. MULLER, Cryst. Res. Technol., 19 (1984) K 101.
5) J. DOWSETT, D. DWYER, F. STERN, R.A. HEINECKE, A.H. TRUELOVE, Proc. IEEE Ultrasonics Symp. (1985) 301.
6) A.J. BERNOT, Ibid (1985) 271.
7) J.R. HUNT, R.C. SMYTHE, Ibid (1985) 292.
8) R.J. BRANDMAYR, J.R. VIG, 40th Ann. Freq. Control Symp. (1986) 86.
9) R.J. BRANDMAYR, J.R. VIG, 39th Ann. Freq. Control Symp. (1985) 276.
10) J.R. HUNT, 41st Ann. Freq. Control Symp. (1987) 183.
11) J.K. VONDELING, J. Mat. Sci., 18 (1983) 304.
12)a) C.R. TELLIER, F. JOUFFROY, Ibid, 18 (1983) 3621.
b) C.R. TELLIER, 38th Ann. Freq. Control Symp. (1984) 105.
c) C.R. TELLIER, 2nd Eur. Freq. Time Forum (1988) 937.
d) C.R. TELLIER, J. Mat. Sci., 17 (1982) 1348.
e) C.R. TELLIER, Surf. Technol., 21 (1984) 83.
f) C.R. TELLIER, T.G. LEBLOIS, P.C. MAITRE, J. Mat. Sci., 24 (1989) 3029.
13) J.C. JUMAS, A. GOIFFON, B. CAPELLE, A. ZARKA, J.C. DOUKHAN, J. SCHWARTZEL, J. DETAINT, E. PHILIPPOT, J. Cryst. Growth, 80 (1987) 133.
14) T.F. YOUNG, Rec. Chem. Proc. (1951) 81.
15) T.F. YOUNG, C.R. SINGLETERRY, I.M. KLOTZ, J. Phys. Chem., 82 (1978) 671.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-04 Number 5

ETCH CHANNELS IN QUARTZ MONOCRYSTALS - FACTORS OF INFLUENCE

Olivier BIGNON, Bernard BROTTIER
SICN
4 rue du Radar
BP 343
74008 ANNECY CEDEX
FRANCE

ABSTRACT

A standard method of characterizing etch channel densities in quartz monocrystals is proposed in the April 1988 issue of IEC PUBLICATION 758.

The key problem to solve, in order to characterize a batch by this method, is to perform adequately representative sampling.

In a first stage at SICN, we have attempted to characterize not a batch in particular, but our production and production methods in general. Results of measurements on samples of a number of runs of our alpha-Grade I material [alpha(3500 cm^{-1})$<$0.02 cm^{-1}] are reported and discussed, as well as a comparison with data obtained from the same material after vacuum sweeping.

In a second stage, with the aims of representatively qualify one run by controlling one sample and perhaps be able of growing material of a specified etch channel grade at will, we have begun to accumulate data on seed runs, starting from the initial natural material.

First results are presented and discussed.

I/ INTRODUCTION

The more and more frequent use of chemical machining methods on quartz needs the use of a material of which surface characteristics are known and mastered.

The most common defects are ETCH CHANNELS which appear when etching an AT cut quartz plate with ammonium bifluoride, and which are now well known and easy to identify.

Etch channels affect performances of piezoelectric devices by :
- reducing their resistance to shocks and vibrations,
- forming contamination pits, where pollutants may concentrate,
- favourising surface electrical defects under the electrodes.

It has become most important, for manufacturers of high grade quartz material, to understand the factors which affect the etch channel density in cultured quartz in order to be able of producing reliable material of specified etch channel quality.

II/ ETCH CHANNEL COUNTING METHOD

II.1. IEC 758 proposed standard

A standard method has been specified in the April 1988 edition of IEC 758 publication ; the main steps of this method are as follows :

a) sample preparation
- AT cut with θ = 35,25 $\pm$ 1°, 6.35 mm thick ($\equiv$ 0.25 inches)
- lapping with 25 micron and thus 3 micron abrasive
- removal of a total thickness of 0.35 mm of quartz.

b) sample etching
- in a saturated aqueous solution of ammonium bifluoride at 75 $\pm$ 2°C
- etching time of about 4 hours
- removal of 0.10 to 0.12 mm of material

c) etch channel counting
- use of a 0.25 x 0.25 cm grid and count in the rectangular portion of the Z zones, matching the seed's height
- use of a 30 X magnification binocular microscope
- use of a strong light source directed into the slice in its lesser X direction

II.2. Method developped at SICN

The method used at SICN is only slightly different from the one suggested by the IEC publication. The main differences are :

a) sample preparation

The AT cut sample is first lapped and then optically polished. Its finished thickness is about 5.8 mm.

b) sample etching

The etching time is limited to 2 hours, and the removed thickness is generally between 0.10 and 0.20 mm.

c) etch channel counting

The sample plate is laid on a black ground and, with the specified lighting, etch channels appear as light spots on the surface of the plate (see fig. 1 and 2).

Counting is made using a laboratory precision 0.14 x 0.14 cm grid engraved on a glass plate.

The most significant difference with the specified process is the reduced etching time.

II.3. Influence of etching time

In order to limit the time of use of the highly toxic etchant, we controlled on the same series of samples, the effect of etching time on material removal and etch channel count.

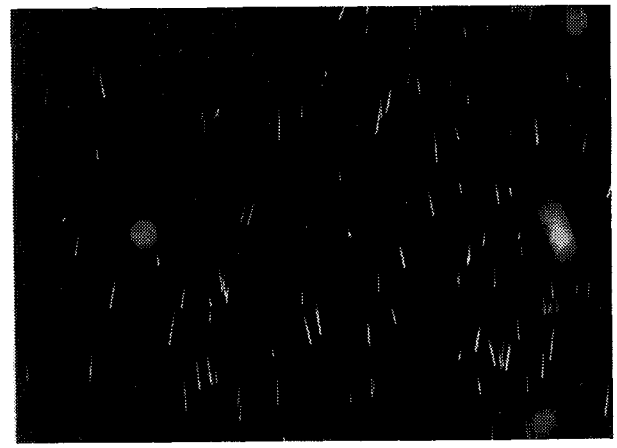
Fig. 3 shows that there is no significant difference in etch channel count when etching time exceeds one hour (and consequent material removal exceeds 60 micron).

III/ RESULTS ON STANDARD PRODUCTS

In a first stage, we have attempted to characterize not a batch in particular, but our production and production methods in general.

Samples have been taken from standard production runs of our alpha-Grade I material [alpha(3500 cm^{-1}) $\leq$ 0.02 cm^{-1}] , grown on standard seeds. Also, etch channel counts have been performed on similar material after vacuum sweeping.

Slice lighted from above



Lighting directed into the slice in its lesser X direction

fig. 1 : Etch channel view (magnification : 11,3)

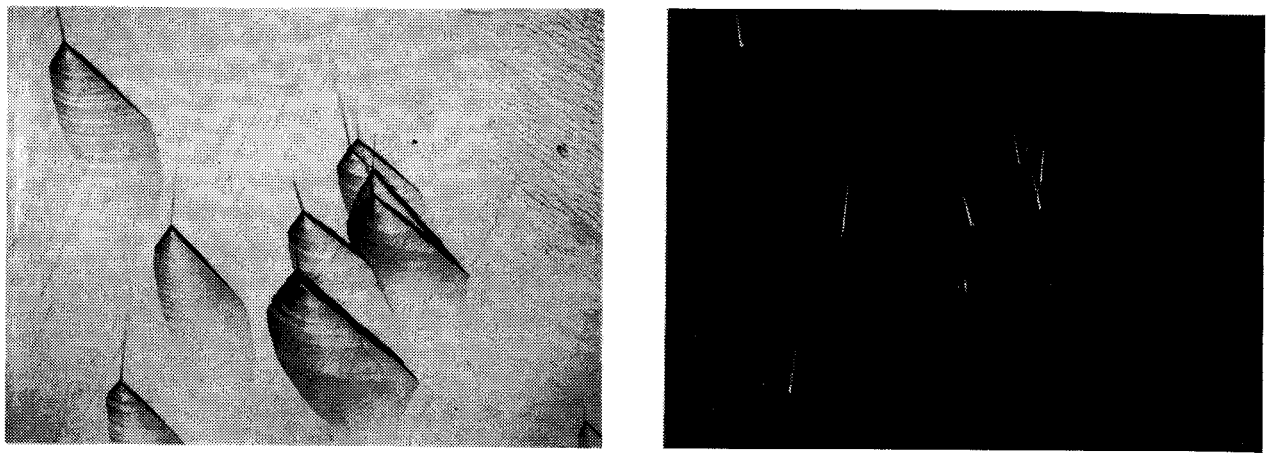


fig. 2 : Etch channel view (magnification : 50)

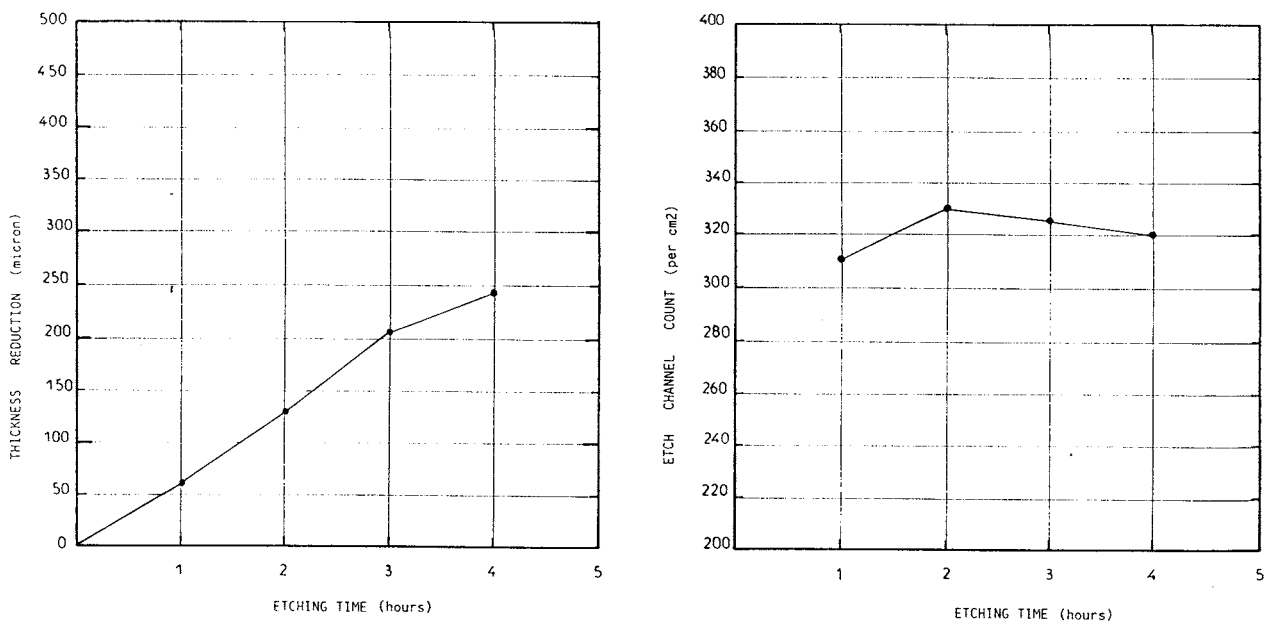


fig. 3 : Influence of etching time on
a) thickness reduction
b) etch channel count

III.1.Alpha-Grade I material

Results of counts on samples of several alpha-Grade I runs (equivalent to IEC 758 new alpha-Grade Aa) are reported on following table.

RUN	alpha (3500 cm^{-1})	Average etch channel count
563	0.018 cm^{-1}	165 cm^{-2}
922	0.019 cm^{-1}	142 cm^{-2}
718	0.016 cm^{-1}	283 cm^{-2}
268	0.019 cm^{-1}	278 cm^{-2}
263	0.020 cm^{-1}	144 cm^{-2}

Etch channel counts are between 100 and 300 cm^{-2} , which shows that standard production of high alpha-grade material classifies the products in IEC etch channel grade 3.

Dispersion, however, is quite important and seems to be dependant on two main factors.

III.1.1. Sampling problems

In a same batch, we sampled at random two blocks, out of which AT cut slices were made for etch channel counts. Results are given hereunder.

	Etch channel count		
	Average	Minimum	Maximum
Block 1	198 cm^{-2}	175 cm^{-2}	214 cm^{-2}
Block 2	304 cm^{-2}	194 cm^{-2}	436 cm^{-2}

The only differences between these two blocks, taken out of the same production batch are :

- location into the autoclave
- seed quality

Block No 2, in particular, shows a very important variation of etch channel count with Y position of the AT cut sample, which seems to be due only to inhomogeneity in seed quality.

This gives a first evidence that the desired ability to produce at will material of a given homogeneous and reproducible etch channel quality necessitates the previous control of seed blocks and seeds.

III.1.2. Distribution in X and Z' directions

Dispersion in etch channel counts also appear on a given AT cut sample, as shown on fig. 4 , which represents the typical distribution obtained on our samples. The distribution in Z' direction is rather homogeneous, but along the X axis, it clearly shows a maximum in the middle of the Z-zone (corresponding to the middle of the seed and to minima at both X ends).

Although not yet explained, this distribution is very coherent with previous observations on LANG topographies, which exhibit a strong reduction of the number of dislocations in the -X ends of the Z zone and also sometimes, in the +X end of this zone (see figure 5).

III.2. Effects of sweeping

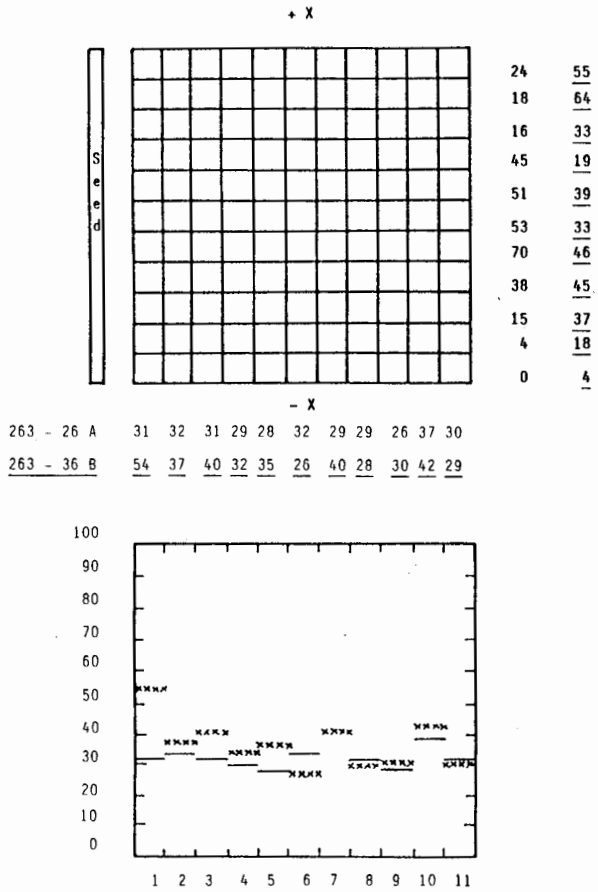
SICN's sweeping method has been developped for radiation hardening of cultured quartz, under a contract by DRET and DGRST, and with the qualification control of one of our major customers.

Alpha-Grade I material with very low inclusion count (inclusion grade 1 of IEC 758 standard) are submitted to following conditions :

- secondary vacuum
- 500°C
- 2500 V/cm Z

during a sufficient period of time.

The effect of sweeping is to move out alcaline impurities, particularly Li+ and Na+ which are initially combined with an Al+++ ion, and replace the initial Al-Li or Al-Na centers by an Al-e+ center (in case of vacuum sweeping), experimentally proved to be more stable when submitted to irradiations.



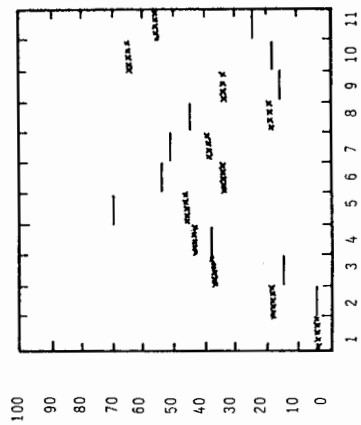


fig. 4 : Etch channel distribution in Z' and X directions

263 - 26 A ——
263 - 36 B ××××

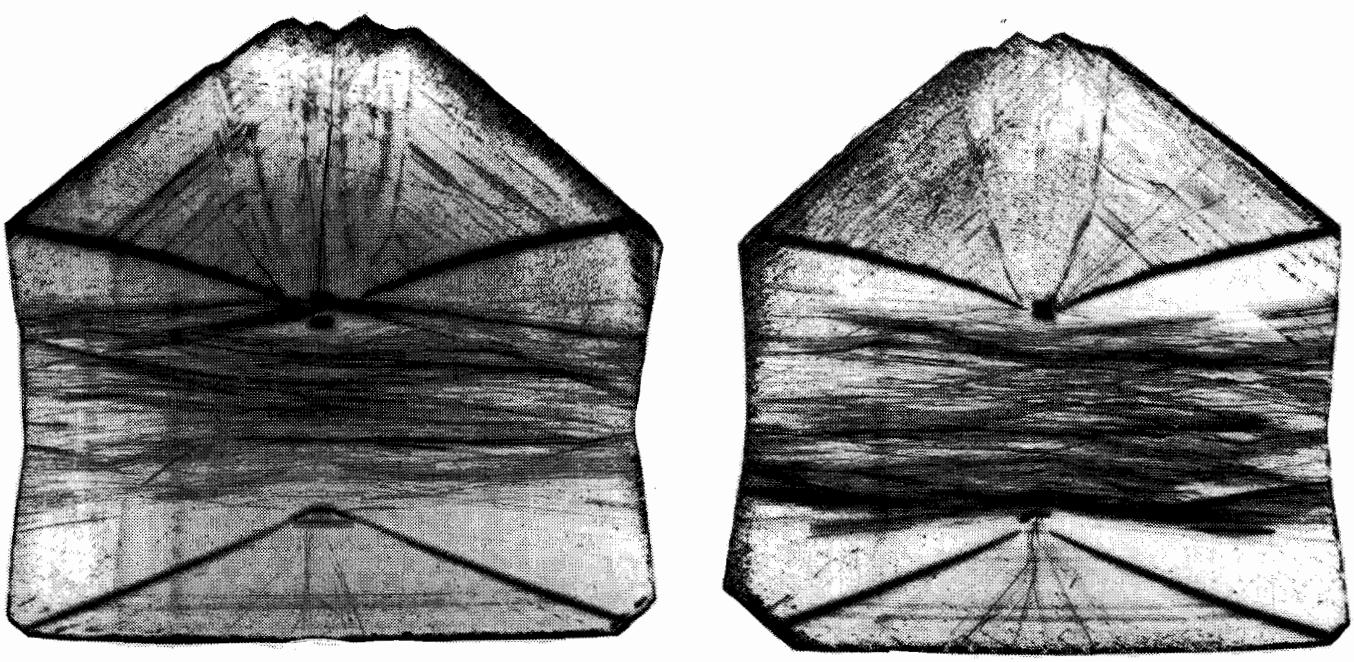


fig. 5 : Typical dislocation patterns in the pure Z zone

Sweeping, and notably vacuum sweeping, also has a very strong effect on etch channel characteristics of material, as shown by the following data.

Block	Etch channel counts		
	Before sweeping	After sweeping	Percent reduction
621 A5	/	2 cm^{-2}	/
621 B5	/	2 cm^{-2}	/
563 26A	125 cm^{-2}	9 cm^{-2}	93 %
563 36B	168 cm^{-2}	4 cm^{-2}	98 %
Average	146 cm^{-2}	4 cm^{-2}	95 %

The comparison between etch channel counts before and after sweeping is made on AT cut samples from the same block, near to one another.

All data up to now show that, starting from an IEC etch channel grade 3 material, vacuum sweeping reclassifies the product in IEC etch channel grade 1.

IV/ INFLUENCE OF SEED QUALITY

IV.1. Natural and standard seeds

Comparison of dislocation patterns and etch channel counts in cultured alpha-grade I quartz grown on natural seed on one hand, standard synthetic seed on the other hand is shown hereunder.

	natural seed	standard seed
dislocations	3 cm^{-2}	several 100 cm^{-2}
etch channels	2.8 cm^{-2}	144 cm^{-2}

Both blocks were issued from the same batch, so that the only difference, as illustrated by fig.6 and 7, was the cristalline quality of the seed.

These results confirm the observations of a number of authors, who relate the presence of etch channels to the presence of dislocations (although absence of etch channels does not mean absence of dislocations and etch channel counting cannot replace LANG topography).

They also confirm the idea that the key factor for dislocation and etch channel quality of cultured quartz is the quality of the seeds.

BARNS et al. showed that the degradation of seed quality, and consequently the increase of internal defects in cultured quartz, such as dislocations, became critical after the 6th generation.

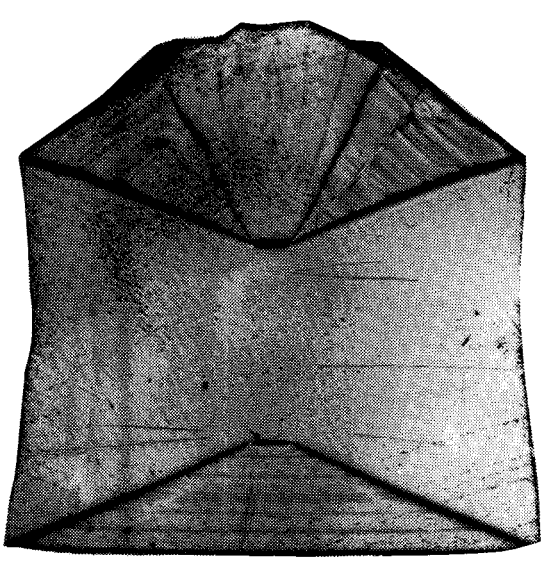
It becomes thus obvious that mastering quartz quality in term of cristalline structural homogeneity needs to begin with mastering seed quality.

IV.2. Follow-up of seed generation

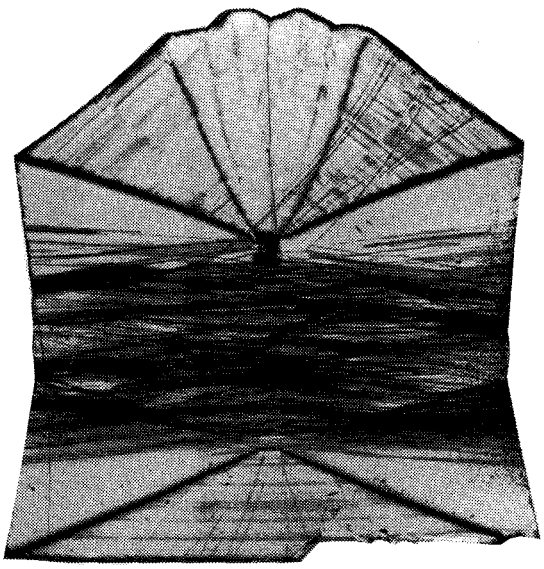
SICN recently acquired a large natural quartz block for seed generation. This block allowed to machine a number of large natural seed plates.

One of these plates has already been grown, together with a sample seed, cut out of the same block, for etch channel and X-ray control.

Only etch channel counts on a piece of natural quartz, and on one half of the sample block (the other half having been used to make sample seeds of first generation) have been achieved up to now. The values obtained are given hereunder :



Natural seed



Standard seed

fig. 6 : Natural and Standard seeds

seed run N°	alpha (3500 cm^{-1})	generation	etch channel counts
/	/	natural	0 cm^{-2}
567	0.023 cm^{-1}	first	< 5 cm^{-2}
		standard	320 cm^{-2}

These results confirm that growing quartz on etch channel free seeds give material with a very low quantity of etch channels. They also illustrate the fact, when compared with previous data, that a faster growth rate, leading to a higher alpha value, may also result in higher number of etch channel sites.

Second generation seed blocks are currently being grown on first generation seeds and data will keep being accumulated ; but already, some orientative conclusions concerning the production of low etch channel cultured quartz can be drawn.

V/ CONCLUSIONS

A very powerful way of reducing the amount of etch channels in cultured quartz is to sweep it properly. However, if we accept the idea that etch channels are related to the presence of dislocations, we have also to admit that sweeping is able to reduce the effect of the dislocations in presence of an etchant by changing the relationship factor between dislocations and etch channels, but not to actually reduce the number of dislocations.

Growing quartz on natural seeds or low generation seeds on the other hand will result in less etch channels because less cristalline defects are present. It may also be possible as first suggested by ZARKA and BUISSON and then by several other authors, that using seeds cut out of the +X zone of cultured quartz is sufficient to generate low etch channel material, because of low dislocation content.

It remains however certain that assuring homogeneity and reproducibility of the fabrication of low etch channel quartz products at reasonable cost will need fine understanding of the growing process and the way a dislocation is generated during the autoclave run.

ACKNOWLEDGEMENT

The authors wish to thank the MRT (Ministère de la Recherche et de la Technologie), the DRET (Direction des Recherches et Etudes Techniques) for their financial backing, Messrs. ZARKA and CAPELLE for having made X-ray topographies and commented on them and the personnel of the laboratory of SICN at Annecy.

REFERENCES

. BARNS, KOLB, LAUDISE, SIMPSON, KROUPA, J. Crystal Growth, 34, (1976)
. ZARKA, LIU LIN, BUISSON, J. Crystal Growth, 54, (1981), 394
. BALASCIO, ARMINGTON, 8th Quartz Devices Conference and Exhibition (1986)
. MARTIN, 41st Annual Frequency Control Symposium, (1987), 167
. HUNT, 41st Annual Frequency Control Symposium, (1987), 183
. STERN, DOWSETT, CARTER, WILLIAMSON, 43rd Annual Frequency Control Symposium, (1989), 634
. JOHNSON, IRVINE, FOISE, 3rd European Time and Frequency Forum, (1989), 337

Applications

APP-04 Time Transfer: Future Prospects

Chairman: R. Sydnor

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-04 Number 1

REMOTE TIME AND FREQUENCY COMPARISONS NOW AND IN THE FUTURE

David W. Allan
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado 80303

ABSTRACT

Time metrology has moved from milliseconds to picoseconds in the last four decades, and frequency metrology has moved from nine significant digits to sixteen. The ability to synchronize remote clocks has improved dramatically as well. With the implementation of GPS (Global Positioning System), the full long-term frequency stability, as well as the frequency accuracy of the best atomic clocks, can now be transferred to remote sites. In the future, GPS's selective availability, an intentional degradation of system performance, will adversely affect the usefulness of GPS time and frequency transfer for the average civilian user.

This paper discusses various alternatives for clock synchronization and syntonization, and makes some comparisons between various techniques used in synchronizing and syntonizing clocks. In the process, it reviews concepts of time stability and accuracy, and frequency stability and accuracy. The future of comparison systems is considered. An Appendix of definitions is provided to support the concepts developed.

Contributions of the U.S. Government; not subject to copyright.

INTRODUCTION

The synchronization of clocks is a subject which has been widely treated throughout the years. With the development of very accurate means for satellite time transfer, the subject has substantially gained in importance.

Time transfer systems (or clock synchronization systems) are often given a single numeral designation, characterizing their precision, or accuracy, in microseconds or nanoseconds. This characterization is often ambiguous or inadequate, and thus it is now important to clarify the factors involved in clock synchronization or comparison systems.

BACKGROUND

Characterizing the measurement system is essential if a remote (slave) clock is intended to be optimally synchronized or syntonized to a master clock. In this situation, optimum design of the servo which locks the slave to the master clock requires a characterization of all of the contributing elements. We are not here generally concerned with measurement noise (divider or counter-noise), although such noise can be problematic in some instances.

A free-running clock can almost always be better characterized than one whose output is servo controlled to another clock. Hence, it is better to have the servo control be a computed output or an external micro-phase stepper in order to provide a synchronized or syntonized output which does not perturb the free-running clock. [1] A local set of clocks can be better characterized if there are at least three of them of about the same quality. [2] Algorithms can be employed to intelligently combine the readings of a set of clocks so that the algorithm-computed time and/or frequency can be more stable than that of the best clock in the set. In addition, algorithms can be designed to test for abnormal clock behavior and for desensitizing the computed time to any abnormal behavior or failures. [3]

If the clocks, as well as the comparison system, are well characterized, then an ensemble of clocks can be constructed from a set of remotely located clocks. With full characterization of all components, the system of clocks and its associated comparison can be optimized for overall performance. While often applied to local ensembles, this concept has apparently not yet been applied to an ensemble whose member clocks are in different locations. There are some long-term plans to do this using GPS. Potentially significant gains are available in the proper application of this concept to the generation of TAI.

Figure 1 illustrates a straightforward comparison system or time-and-frequency-dissemination system which measures the time-and-frequency differences between Clock 1 and Clock 2. The word comparison will be used for simplicity to represent any generic dissemination system. Of concern is the characterization of the full noise in the comparison, including measurement noise, clock noise, and noise introduced in the comparison path and system. In figure 2, an additional concern arises in designing a servo loop to slave a

remote clock to a master clock. The data from the comparison may not be available immediately; hence in the feedback loop the measurement noise, path deviations, and delay in acquiring the comparison data will affect the servo design very fundamentally. Practical delays in acquiring comparison data range from milliseconds to times longer than a month. For example, the delay time (data acquisition time) for servo controlling Coordinated Universal Time at NIST (UTC(NIST)) to the international UTC scale is more than 1 month. Although this paper does not address servo design theory, it is important to stress that the measurement noise and path noise characteristics, and the delay in acquiring comparison data, play very important roles in servo design.

Appendix A provides a few relevant definitions (for example, "precision," "accuracy," "stability"). In characterizing systems for comparing clocks which are located remotely from each other, it is important to consider concepts such as time accuracy, time stability, time prediction error, frequency accuracy, and frequency stability. Each has a unique interpretation.

Conceptually, time accuracy is the time difference between the readings of two clocks at some time in a given reference frame. One of the clocks is often defined as perfect, so that the assessment is of the accuracy of a clock relative to that "ideal" clock. We can imagine the transport of a perfect, portable clock to accomplish this time-difference measurement. Time accuracy is often limited by systematic errors in the comparison system, such as uncertainties in cable delays and propagation-path-length uncertainties. In addition, systematic differences between clocks will contribute to time inaccuracy. Time accuracy can never be better than time stability and is often much worse.

One of the best ways to observe time stability is to plot the time residuals, often denoted x(t), that exist between two clocks after the systematic errors have been subtracted. In addition to the usual kinds of random variations which affect clock and comparison system performance, time stability is often affected by environmental variations. People commonly measure time stability as the rms deviation of the time residuals from a linear regression to the time deviations. This practice, which can be very misleading, is discussed in some detail in the body of this paper. If there are periodic terms affecting a time comparison system, then the spectral density of the time or the phase fluctuations may be a very good measure. We may also measure the effect of these periodic terms using $\sigma_y(\tau)$. [2] For time stability there is often a τ (averaging-time) dependence. Also, $\tau \mathrm{mod}\sigma_y(\tau)$ is a useful measure of the time stability of a comparison system.

The quantity $K\tau\sigma_y(\tau)$ is a useful measure for estimating the time prediction error of a clock. Often a particular power-law, spectral-density model is dominant for the signal variations from the clocks and/or the comparison system. Under the assumption of optimum prediction over the data spacing interval, τ_0, the value of K is $1/\sqrt{3}$ for white-noise PM, 1 for white-noise FM and for random-walk FM, and 1.2 for flicker-noise FM. In the case of white-noise phase modulation, the quantity $\tau \mathrm{mod}.\sigma_y(\tau)/\sqrt{3}$ is the optimum rms time prediction error for an average of n x(t) measurements ($\tau = n\tau_0$).

Frequency accuracy for a given primary standard is not a function of integration time and is properly stated as a single number. But the ability of a comparison system to determine absolute frequency difference between two standards is often a function of the sampling, or integration time τ. The frequency accuracy of a comparison system is also a function of the data processing method. This leads to the idea that there is an optimum method for estimating the absolute frequency difference between two remote clocks or for controlling the frequency of a remote clock.

Frequency stability, like time stability, is observed by looking at a plot of the fractional frequency offset, y(t), where $y(t) = (\nu(t)-\nu_o)/\nu_o$; $\nu(t)$ is the time varying frequency output of a clock and ν_o is the clock's nominal frequency. In practice, measured values of y(t) are observed over some averaging time, τ. It is often very useful to observe a y(t) plot at different averaging times. The frequency stability of a comparison system can be quantified in the same way clocks are characterized, using a $\sigma_y(\tau)$ or $\mathrm{mod}\sigma_y(\tau)$ plot. It is sometimes useful to measure the spectral density of the frequency fluctuations to supplement the time-domain methods, in order to ascertain the presence of different kinds of noise. The kind of noise observed in comparisons between two clocks, and the noise which may be added by the comparison system, will determine how to optimize estimates of characterization parameters (both systematic errors and noise) for the clocks and the comparison system. One important example of a characterization parameter is the relative frequency drift between two clocks. See reference [2] for examples.

CHARACTERIZATION OF COMPARISON SYSTEMS

Figure 3 shows the improvement in the U.S. primary frequency standard since the development of cesium beam technology. The trend line shows an improvement of about a factor of 10 every seven years. Further improvement is expected--although not certain. There are now good indications that standards based on trapped and cooled ions will yield dramatic improvements. The estimated ultimate potential for these devices is an accuracy of about one part in 10^{18}, but practical considerations will make this limit difficult to achieve.

In the past, the accuracy of operational comparisons between primary standards lagged behind the accuracy of the standards. Further improvements in primary standards were thus of limited use. However, during the last decade, the development and application of GPS time transfer and two-way satellite time transfer dramatically changed the picture. With the excellent comparison accuracy available in the GPS common-view technique or in the two-way technique, comparison accuracy at integration times of a few days and longer is now ahead of clock accuracy. This was a major breakthrough for international time and frequency comparisons, and the GPS common-view technique has become the defacto international standard for comparisons. [5] A decision by GPS system operators to intentionally degrade performance as observed by civilian users (called selective availability) raises questions which are important in time transfer applications.

Time transfer using the two-way satellite technique now appears very attractive alternative to primary timing centers. More information is needed on the accuracy and long-term stability of this comparison

technique since primary investigations have not focused on these. [6], [7] Most of the published results concern short-term time stability.

Important factors for all of these comparison systems include cost, simplicity of use, and means for correctly assessing comparison accuracy. The ideal comparison system is one which provides the time difference, the frequency difference, and the relative time and frequency stability of the clocks and which makes clear the uncertainties associated with the comparison system. If the comparison system is to be widely used, the cost should be low. Of course, there is no single system which now meets this ideal. Figure 4 illustrates the performance of some of the common comparison techniques now being used. Both $\sigma_y(\tau)$ and $\mathrm{mod}\sigma_y(\tau)$ are used to characterize the frequency stability of these comparison systems, because, in some cases, white-noise phase modulation (PM) is the limiting random process and therefore characterization using $\sigma_y(\tau)$ is ambiguous.

When white-noise phase modulation is the predominant noise in a comparison system, some important equations for optimal estimation of time and frequency differences between the clocks are

$$\hat{x}(i) = a_0 + a_1 \cdot i \qquad \text{and} \tag{1}$$

$$s_x = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (x(i) - \hat{x}(i))^2} \quad . \tag{2}$$

Here $\hat{x}(i)$ is the optimal estimate of the time difference between the clocks at the measurement point i. The coefficients a_0 and the a_1 are determined by minimizing the variance around the linear regression line, so the meaning of optimum is for a minimum variance. The x(i)'s are the measured time differences over the N measurements. The confidence of the estimate of the intercept a_0 is

$$s_0 = 2s_x/\sqrt{N}, \tag{3}$$

and the confidence of the estimate of the mean value $x = \frac{1}{N} \sum_{i=1}^{N} x(i)$ is $s_{mean} = s_x/\sqrt{N}$.

The confidence of the estimate of the slope (a_1, the frequency difference) is

$$s_1 = \frac{\sqrt{12}}{\tau_0 N^{3/2}} s_x = 2\ \mathrm{mod}\sigma_y(\tau). \tag{4}$$

Equation 1 is the classical equation for a linear regression, which is often computed as a fit to the time residuals. The application of this equation is optimal only for white noise processes. It is assumed that there are N values each separated by τ_0. In this case, the standard deviation (given by equation 2) is a measure of the time stability at the data sampling rate -- sometimes called the precision of the time difference measurements. The expression N-2 in the denominator shows that two degrees of freedom have been removed with the estimation of a_0 and a_1 in equation 1. The mean-value confidence interval in Equation 4 is half that of the intercept and is the optimum estimate of the time difference between the clocks at the midpoint time., Clearly, it is more efficient to use the mean rather than the intercept to cite the time difference in a comparison. The solution to equation 1 at the midpoint is equal to the mean value. Equation 5 shows the value of using $\mathrm{mod}\sigma_y(\tau)$ to determine the confidence of the estimate of the frequency difference, a_1. If the residuals are not white, then the τ dependence will not be $\tau^{-3/2}$, and the linear regression will not give the optimum estimate of the time and frequency difference of the clocks. If the residuals are white, the value of $\mathrm{mod}\sigma_y(\tau)$ gives the proper value of the confidence for any averaging time, τ. The rapid improvement ($\tau^{-3/2}$) gained in estimating the absolute frequency difference by increasing the averaging time is clearly illustrated by the use of $\mathrm{mod}\sigma_y(\tau)$.

Linear regression analysis is often used to model processes whose residuals do not have a white spectrum. In this case, the linear regression coefficients and their confidences can often be very misleading. A $\mathrm{mod}\sigma_y(\tau)$ diagram will indicate whether or not using linear regression analysis is legitimate; if not, then it gives a measure of the effects of the degradation (caused by the actual random processes) on the estimate of the frequency difference between the two remote clocks.

Figure 5 is a plot of the rms time prediction error seen in currently available clocks and oscillators. The data has been used in an optimum fashion to predict the future over an interval, τ_p. The rms time deviation can be defined in many ways. This is one useful approach. The next four Figures, 6, 7, 8, and 9, are plotted with exactly the same ordinate and abscissa as Figure 5. They can be overlaid to see the various systematic effects, either in the clocks or in the comparison system. Figure 6 shows that the ordinate is labeled with both the white PM level (usually arising from the comparison system) and the time inaccuracy. The time accuracy number provides a hard limit for comparing the time difference between two clocks. In contrast, the white PM level is a function of integration time, and, if other processes are not limiting, knowledge of the time difference improves as the square root of the number of measurements averaged -- consistent with equation 4. If the residuals are white PM, we may also write (from the concept of time averaging of measurements) the following equation

$$s_{rms}(\tau_0) = \frac{\tau^{3/2}}{\sqrt{3\ \tau_0}}\ \mathrm{mod}\sigma_y(\tau), \tag{5}$$

where s denotes the classical standard deviation of x(i) taken τ_0 apart ($\tau = n\tau_0$) as in Equation (2). Since the numerator in Eqüation (5) is constant for white PM, the improvement in $s_{rms}(\tau_0)$ is proportional to $\tau_0^{-\frac{1}{2}}$. This is not surprising since τ_0 is the period over which the phase (or the time) has been averaged. If τ_0 becomes the full data length, then, as expected, equation 5 is the standard deviation of the mean. Here

again, a mod$\sigma_y(\tau)$ diagram provides a good visualization of the estimate of the time difference uncertainty and of the time stability (as limited by the clocks and/or the comparison system).

Figures 5, 7, 8, and 9 are included for the reader's convenience. Figure 7 shows the time difference as a function of time for two clocks whose frequencies differ by various fixed amounts. In this case the abscissa could also be the prediction interval. Figure 8 shows the rms time deviation as a function of the prediction interval as caused by flicker noise frequency modulation (FM) (a common noise in clocks). Notice that the slope is the same as for frequency offset. The factor 1.2 is the K factor for flicker noise where optimum prediction has been assumed. Figure 9 shows the large time deviation that results from frequency drift. The labels for the different lines are fractional frequency drifts per day expressed as powers of 10. The quadratic nature of the time deviation resulting from frequency drift often causes this kind of error to be the predominant long-term systematic error.

Figure 10 is a plot of τmod$\sigma_y(\tau)$ as a function of τ. With $\tau = n\tau_0$, this shows whether or not we benefit from averaging n values of the x(i) time difference measurements. This new approach illustrates the benefit of averaging the time difference measurements, whether the instabilities are in the comparison system or in the clocks. If the measurement noise residuals represent a white PM process, then the time stability will improve as the square root of τ. If it is a flicker PM process, there will be no improvement with averaging. If the plot degrades with increasing τ (slope greater than 0), then there are probably nonstationary processes perturbing the comparison system. In the case of Loran-C, we see a double hump at one half day and at one half year caused by diurnal variations and annual variations. There are longer-term, common-view GPS data than are plotted showing that the time stability does not continue to improve as the square root of τ. In this latter case, the nonstationary processes are probably related to ionospheric modeling errors and errors in the Kalman estimates of the satellites' ephemerides. Multipath distortion at the antenna can sometimes cause several nanoseconds of bias in the time accuracy, but does not change the slope or the level in a τmod$\sigma_y(\tau)$ plot (that is, the bias is constant).

For two-way satellite time transfer, the noise limit does not continue decreasing as indicated by the short-term results in Figure 10. Daily deviations of the order of a few nanoseconds have been observed, but these will likely be reduced as the systems are improved and better characterized. This characterization of the two-way satellite time transfer technique will be very important for the future -- especially for averaging beyond one day. A determination of the time transfer accuracy of this technique will be very important as well. Theoretically, for both the time stability and the time accuracy, two-way satellite time transfer should provide the best operational means for comparing widely separated clocks. The primary drawback to this technique is the need for broadcasting from each station, a requirement which adds cost and involves licensing with government agencies.

THE FUTURE OF COMPARISON SYSTEMS

The best means for comparing widely separated clocks involves satellite techniques. For clocks in proximity (that is, within a modest number of kilometers) perhaps optical fibers will provide the best comparisons. [8] As higher accuracy and more stable clocks are developed, it will be necessary to use higher frequencies to achieve better phase resolution in the comparisons.

It appears that the GPS system could be pushed to a time accuracy approaching a few nanoseconds. For short-baseline comparisons, studies suggest that we might achieve errors as low as 0.1 ns. [9] Time stabilities for GPS common-view comparisons yield τmod$\sigma_y(\tau)$ of about 1 nanosecond times $\tau^{-\frac{1}{2}}$, where τ is in days. At τ = 1 day, this product actually ranges from 0.8 to 8 ns for the many international time stability measurements which use the GPS common-view method. With ionospheric measurement receivers, and more exact post-ephemeris data for the satellites, the GPS common-view technique could yield a comparison limit for frequency accuracy approaching 10^{-17}. This would require about three months of integration under the assumption of ideal white-noise phase modulation. Codeless-ionospheric-measurement receivers, which measure the real-time, path-dependent ionospheric delay, are now becoming available for GPS. There is also the promise that precise post-measurement ephemerides will be made available to the civilian sector. With these advances the GPS common-view, time-and-frequency transfer could be even better than it is today. But the price for this will be additional processing and significant delay in access to data needed to calculate all errors. Table I summarizes the errors in GPS common-view transfers in the face of selective availability, both with and without additional compensation for errors.

How well the systematic errors in two-way satellite time transfer can be understood is yet to be determined. In theory, this technique should be better in both time stability and time accuracy than the GPS common-view technique. The method could provide an order of magnitude of improvement in performance.

An older, often overlooked experiment which has significant bearing on time transfer improvement is the Scout Rocket Experiment which involved flight of a hydrogen maser. [10], [11] This experiment used a microwave Doppler cancellation method and an ionospheric calibration system. From the published data it is estimated that time stability, τmod $\sigma_y(\tau)$ over several hours was about 10 ps. With the stability available from a satellite-borne hydrogen maser, cycle ambiguity of the clock's microwave signal could be resolved from pass to pass or from day to day. Such performance in an operational satellite could yield frequency comparisons over 24 hours of 1 part in 10^{-16}. If the residuals for the comparison were white PM from day to day, it would take only a few weeks to measure frequency difference of a few parts in 10^{-18}. At this level, relativity considerations become very important, and they will be very difficult to calculate.

CONCLUSION

In order to synchronize (or syntonize) a system of clocks optimally, it is necessary to characterize both the stability of the clocks and those of the comparison system. The characterization of the random variations in clocks is pretty well understood, but that of comparison systems is not. Often the standard deviation of the time residuals for both clocks and comparison systems is not convergent, in which case this is not a useful measure. This paper has presented some effective ways to describe and to characterize comparison systems. These allow us to better specify time and frequency comparisons. This issue is becoming more important as system synchronization and syntonization requirements become more stringent.

We have described time accuracy, time stability, time predictability, frequency accuracy, and frequency stability as separate and distinct concepts, with important relationships between these concepts being presented. These ideas have implications for accurate time comparisons. For example, knowing the kinds of random instabilities in clocks and in the comparison system or dissemination system allows us to optimally estimate absolute time and frequency differences between widely separated clocks. With the anticipation of more accurate frequency standards, very careful design as well as characterization of comparison and dissemination systems will be required to take advantage of the improved standards. Even with the current accuracy of time comparisons, there is a need for better specification of the performance of comparison and dissemination systems. We have presented one approach here with the hope that discussion will be stimulated leading to the adoption of a standard method for characterizing the accuracy and stability of the comparison or dissemination process.

ACKNOWLEDGEMENTS:

The author is indebted to Wilfred K. Klemperer, Trudi K. Peppler, Donald B. Sullivan and Matt Young for contributions and suggestions. Marc A. Weiss and Dick D. Davis helped with the data acquisition and analysis for which I am grateful. Thomas Weissert was extremely helpful with the figures.

APPENDIX:

DEFINITIONS

ACCURACY

The degree of conformity of a measured or calculated value to its definition (see Uncertainty).

AGING

The systematic change in frequency with time caused by internal changes in the oscillator.

DRIFT

The systematic change in frequency of an oscillator with time.

ERROR

The difference of a value from its assumed correct value.

FREQUENCY INSTABILITY

The spontaneous and/or environmentally caused frequency change within a given time interval.

PRECISION

The degree of mutual agreement among a series of individual measurements; often, but not necessarily, expressed by the standard deviation.

REPRODUCIBILITY

(A) With respect to a set of independent devices of the same design, the ability of these devices to produce the same value.
(B) With respect to a single device, put into operation repeatedly without adjustments, the ability to produce the same value.

UNCERTAINTY

The limits of the confidence interval of a measured or calculated quantity.

REFERENCES:

[1] D.W. Allan, M.A. Weiss, and T.K. Peppler, "In Search of the Best Clock," IEEE Transactions on Instrumentation and Measurement," IM-38, 624-630 (1989).

[2] D.W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction, of Precision Clocks and Oscillators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 647-654 (1987).

[3] M.A. Weiss and T. Weissert, "A New Time Scale; AT1 Plus Covariance," submitted to Proc. of 20th Annual Precise Time and Time Interval Planning Meeting (PTTI), Redondo Beach, CA, Nov. 30 - Dec. 1, 1989.

[4] D.J. Wineland, J. Bergquist, J. Bollinger, W. M. Itano, D. Heinzen, S. Gilbert, C. Manney, and C. Weimer, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions," Proc. of 43rd Annual Symposium on Frequency Control, Denver, CO, May 31-June 2, 1989, pp. 143-150.

[5] D. W. Allan, et. al., "Accuracy of International Time and Frequency Comparisons via Global Positioning System Satellites in Common-View," IEEE Transactions on Instrumentation and Measurement, IM-34, 118-125 (1985).

[6] D. Hanson, "Fundamentals of Two-Way Time Transfers by Satellite," Proc. of 43rd Annual Symposium on Frequency Control, Denver, CO, May 31-June 2, 1989, pp. 174-178; D.A. Howe, et al., "NIST-USNO Time Comparisons Using Two-Way Satellite Time Transfers," 43rd Annual Symposium on Frequency Control, Denver, CO, May 31 - June 2, 1989, pp. 193-198.

[7] M. Imae, H. Okazawa, T. Sato, M. Urazuka, K. Yoshimura, and Y. Yasuda, "Time Comparison Experiments with Small K-Band Antennas and SSRA Equipments via a Domestic Geostationary Satellite," IEEE Transactions on Instrumentation and Measurement, IM-32, 199-203 (1983).

[8] G. Lutes and M. Calhoun, "Simultaneous Transmission of a Frequency Reference and a Time Code Over a Single Optical Fiber," submitted to Proc. of 20th Annual Precise Time and Time Interval Planning Meeting (PTTI), Redondo Beach, CA, Nov. 30-Dec. 1, 1989.

[9] P.F. MacDoran, D.J. Spitzmesser, L.A. Buennagel, "SERIES: Satellite Emission Range Inferred Earth Surveying," Proc.of the Third International Geodetic Symposium on Satellite Positioning," Las Cruces, N. Mexico, February 1982.

[10] R.F.C. Vessot and M.W. Levine, Gen. Relativ. Gravit. 10 (1979) 181;
R.F.C. Vessot, et al., Phys. Rev. Lett. 45 (1980) 2081.

[11] D.W. Allan, C.O. Alley, N. Ashby, R. Decher, R.F.C. Vessot and G.M.R. Winkler, Ultra-Accurate International Time and Frequency Comparison Via an Orbiting Hydrogen-Maser Clock, Journal De Physique, Colloque C8, 395-413, 1981.

TABLE I. GPS COMMON-VIEW TIME-TRANSFER ERROR SOURCES

(WITH SELECTIVE AVAILABILITY ON)

SOURCE	COMMENTS	RMS TIME ACCURACY (ns)
CLOCK DITHER	CANCELS IN C-V MODE	--
EPSILON	DEPENDS ON THE BASE-LINE	30 to 50
IONOSPHERE (BDCST)	DEPENDS ON TOD AND COORD.	5 to 40
TROPOSPHERE	DEPENDS ON ELV. AND WEATHER	2 to 5
MULTIPATH	DEPENDS ON GROUND PLANE AND REFL.	4 to 8
RECEIVER	DEPENDS ON THE MAKE AND MODEL	1 to 100
C-V TIME TRANSFER ERRORS (NO COMPENSATION)		31 to 120

(WITH SELECTIVE AVAILABILITY ON AND WITH COMPENSATION)

SOURCE	COMMENTS	RMS TIME ACCURACY (ns)
CLOCK DITHER	CANCELS IN C-V MODE	--
EPSILON	COMPUTED EPHEMERIS (Some Days After)	3 to 5
IONOSPHERE	WITH IONOSPHERIC CALIBRATOR	2 to 3
TROPOSPHERE	DEPENDS ON ELV. AND WEATHER	2 to 5
MULTIPATH	WITH CHOKE-RING ANTENNA GND. PLANE	2 to 4
RECEIVER	DEPENDS ON THE MAKE AND MODEL	1 to 100
C-V TIME TRANSFER ERRORS (WITH COMPENSATION)		5 to 100

The right column lists rms estimates for each of the time accuracy error elements with the sum at the end of each column being the square root of the sum of the squares. EPSILON is the intentional insertion of errors in the broadcast ephemeris. The meanings of other terms in the table are:

C-V - GPS common-view mode
TOD - Time of Day
GND - Ground
Elv. - Elevation
Refl. - Reflections
BDCST - As Broadcast

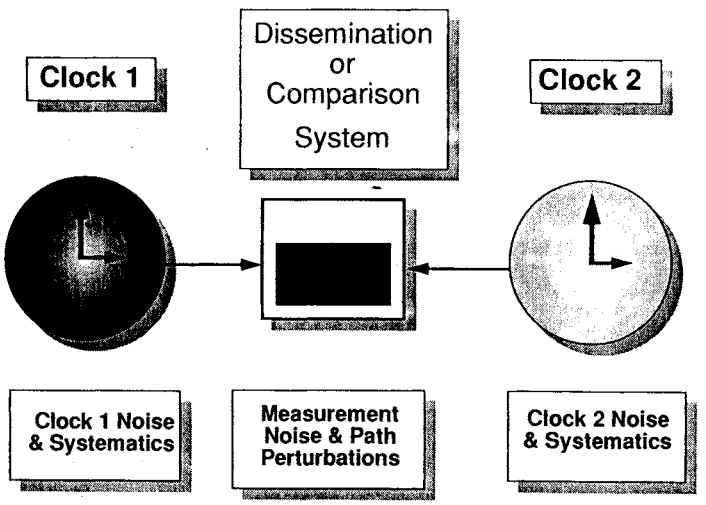


Figure 1. This figure shows two clocks, some arbitrary distance apart, being compared by some generic comparison system. In principle, the comparison system can be co-located with either or both of the clocks or be apart from both. In general, the measured values coming from the comparison system will have variability due to clock noise, delay variations in the connecting links, and variations in the comparison system itself. Characterizing the performance of the links and the comparison system is important. Otherwise, the determination of what variations come from the clocks and what come from the comparison system and the links would be impossible.

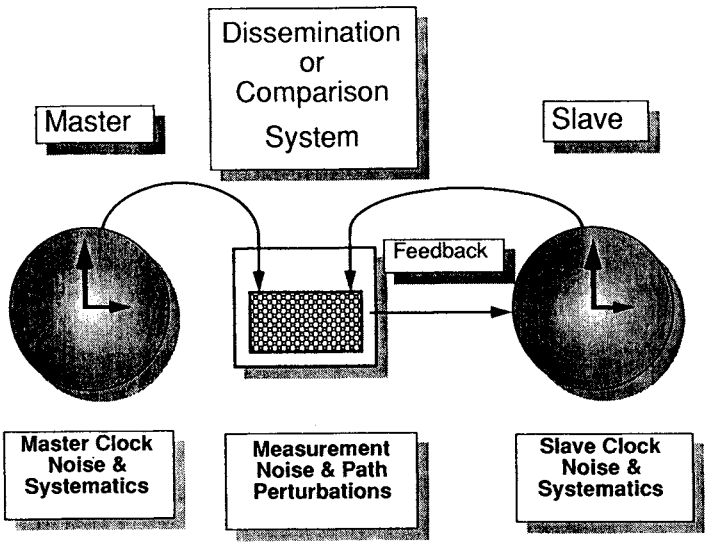


Figure 2. This figure is similar to Figure 1. Again, we are measuring the time and frequency difference between two clocks located some distance apart. In this case we wish to servo control the time and/or frequency of the slave to the master. A proper characterization of the links between the clocks in combination with the comparison system is essential for the proper design of a feedback system to control the slave clock. Another important parameter for the feedback design is the delay associated with acquisition of data by the comparison system.

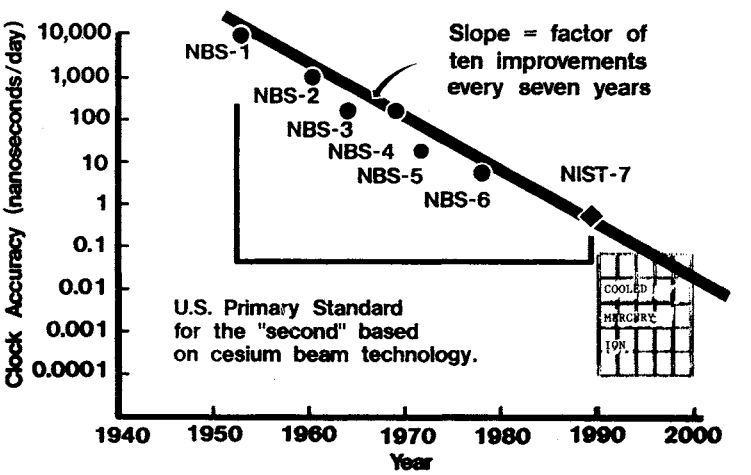


Figure 3. Improvement in atomic frequency standards of the U.S. The overall trend is a factor of 10 improvement every seven years. If this trend line continues, and there is good indication that it may, then more careful attention is needed both in the design as well as in the proper characterization of systems used to compare these with other national standards. Note: one nanosecond per day corresponds to a fractional frequency change of about a part in 10^{14}.

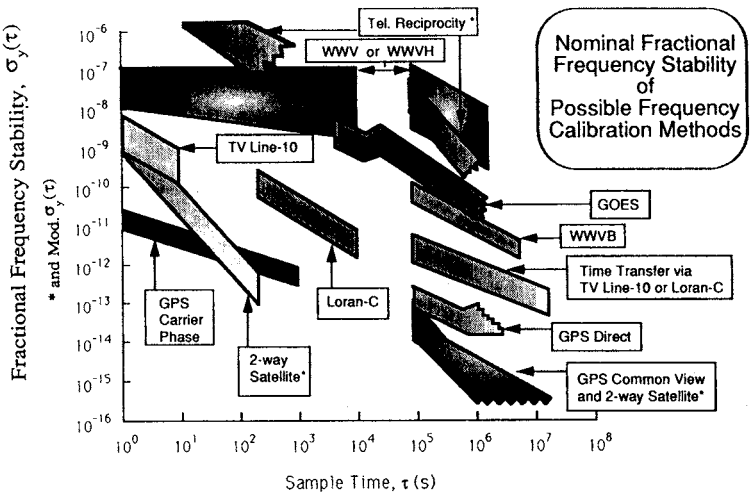


Figure 4 (Caption to follow)

Figure 4. Nominal frequency stability for several important comparison systems. The stabilities are characterized using $\sigma_y(\tau)$ except where indicated by an "*" in which case Mod$\sigma_y(\tau)$ is used. The latter case Mod $\sigma_y(\tau)$ was used in those involves situations where white noise PM is predominant for some range of sample times τ. The "Tel. Reciprocity" data were analyzed under the assumption of reciprocity of the path (measure the round trip time and divide by two to calibrate the path delay). The short-term data were measured locally and the long-term data were measured between Colorado and Hawaii via communication satellite. We often found that telephone modems contributed more noise than the path. What is plotted is the composite. The WWV and WWVH time-and-frequency transmissions at 2.5, 5, 10 and 15 MHz (WWV also broadcasts at 20 MHz) are limited in their stabilities by sky-wave-path variations. GOES East and GOES West are NOAA weather satellites broadcasting UTC(NIST) on two slightly different frequencies near 468 MHz. Here, the stability is limited by the knowledge of the satellites' ephemerides. WWVB is NIST's 60 kHz time-and-frequency broadcast service; in this case the propagation path stability is limited by the fluctuations in the earth-ionosphere waveguide. The TV Line-10 method involves line-of-sight transmissions in the TV band. It can operate with an atomic clock at the transmitter or with two clock sites receiving the TV Line-10 arrival times concurrently and subtracting one set of numbers from the other. Stability limitations here are often caused by the receiving equipment. Loran-C is a ground-wave navigation signal (at 100 kHz) operated by the U.S. Coast Guard. The time is monitored and controlled with respect to UTC(USNO). The stability is limited by propagation path variations. Two-way satellite time transfer uses spread-spectrum modems operating with different up-link and down-link carrier frequencies in one of several different bands (C, Ku, and K). The short-term stability for two-way satellite time transfer is basically limited by signal-to-noise and bandwidth considerations. Currently, the long-term performance seems to be limited by equipment instabilities. One can only extract frequency information from the "GPS Carrier Phase" measurements, and the stability seems to be limited by the GPS on-board clocks. Time and frequency stability of directly received GPS signals is limited mainly by variations in the GPS Kalman-state estimates for the system. If one is using an L1 GPS timing receiver only, then the ionospheric modeling errors can contribute additional instabilities. In some cases, signal multipath errors and/or receiver instabilities can also contribute significant instabilities. Using GPS in the common-view mode cancels out the GPS clock instabilities and cancels some of the broadcast satellite-ephemeris instabilities. The stability limits for the common-view mode arise from the same mechanisms as for GPS direct measurements except that some errors are reduced by common-mode cancelation.

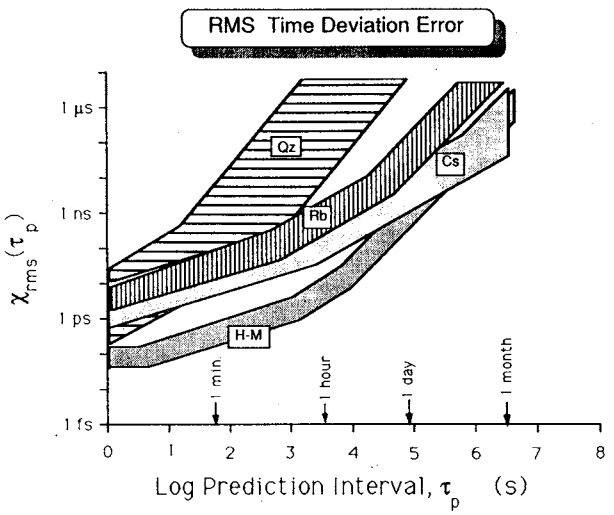
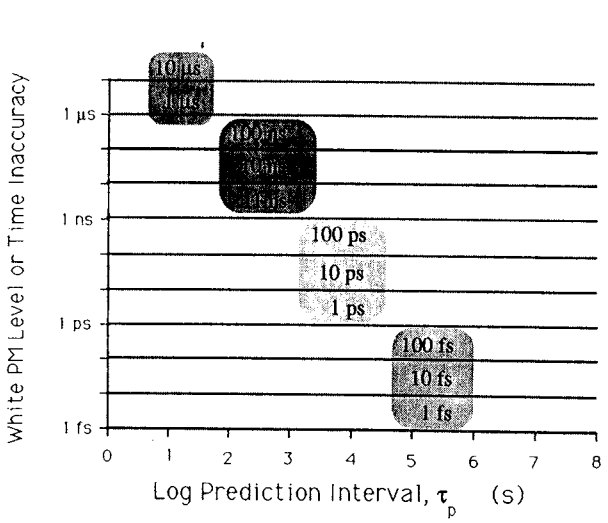
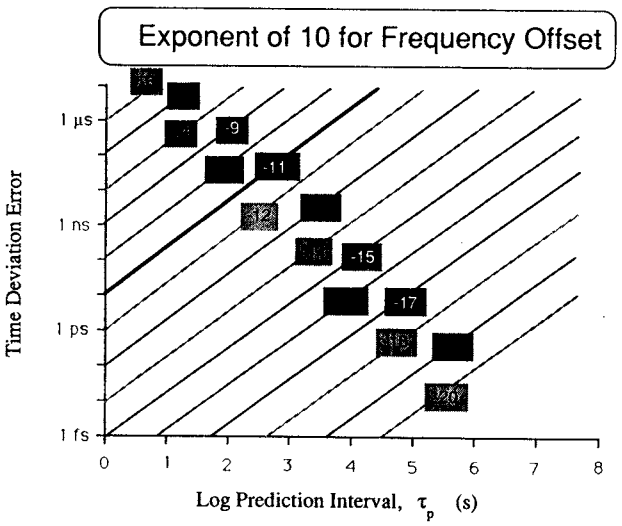
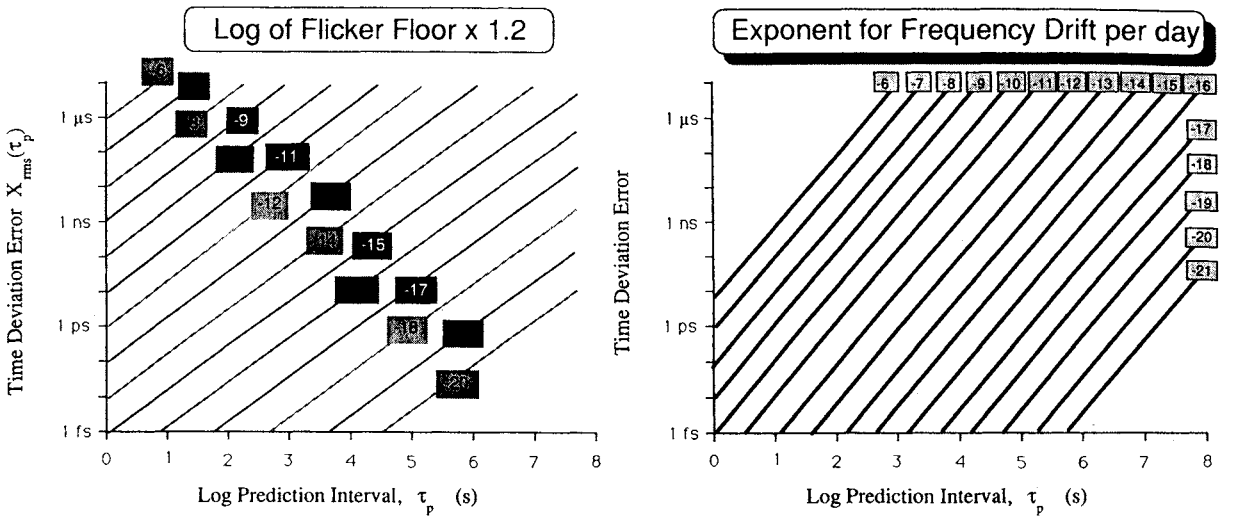


Figure 5. Time prediction error, $x_{rms}(\tau_p)$, as a function of the prediction interval for commercially available precision clocks. Qz denotes quartz-crystal-oscillator clock; Rb denotes rubidium gas-cell clock; Cs denotes cesium-beam clock; and H-M denotes active hydrogen maser clock. The prediction error is calculated from $K\tau\sigma_y(\tau)$ with K chosen for an optimum prediction estimate. The value of K depends on the noise type.





Figures 6 and 7 (caption to follow)



Figures 6, 7, 8 and 9. The ordinates and abscissas of these four plots are the same as those for Figure 5. Figure 6 can represent either the time accuracy or the white noise PM level. The time accuracy is often limited by systematic effects and averaging does not improve it. The white noise PM is well represented by the standard deviation of the measurements, and, if this is the limiting noise, then averaging will improve the knowledge of the time as the square root of the number of values averaged. Figure 7 is the time accumulation over some interval, τ_p, due to a systematic frequency difference (or offset) between two clocks being compared. Figure 8 is the rms time deviation resulting from a random flicker FM process -- often observed in long-term clock comparisons. The 1.2 (1/√ln2) factor is the K factor for flicker noise FM. "Flicker Floor" means the value of $\sigma_y(\tau)$ where there is a τ^0 dependance, that is, where there is no further improvement in stability with increasing τ. The curves in Figures 7 and 8 have the same slope (+1) even though they arise from different mechanisms. Figure 9 demonstrates the long-term significance of time deviation errors resulting from a linear frequency drift in a clock. The plus-two (+2) slope corresponds to the quadratic departure of the time of a drifting clock. If frequency drift exists in a clock, this error along with environmental perturbations is often the main cause of long-term time error.

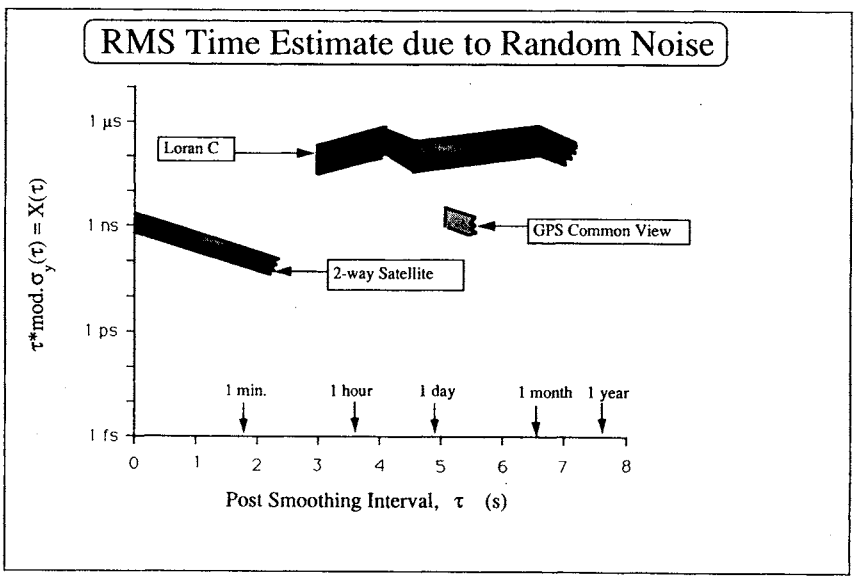


Figure 10. This type of plot can be used to determine whether or not smoothing or averaging the data is beneficial. We have here defined the time stability as the product τ mod $\sigma_y(\tau)$. For flicker noise PM, white noise FM, flicker noise FM and random-walk noise FM the standard deviation of the time residuals grow without bound as the data length increases. Hence, the standard deviation is not a good measure. The above product is a good measure since it is convergent and is data-length independent. This measure can also show the effects of systematic errors and of environmental perturbations as well as the different kinds of noise processes that may be driving the instabilities in the comparison system and/or in the clocks. The different comparison methods are explained in the caption of Figure 4.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-04 Number 2

DESIGN CONSIDERATIONS FOR KU-BAND TWO-WAY SATELLITE TIME TRANSFER TERMINALS

D. Kirchner

Department of Communications and Wave Propagation
Technical University Graz (TUG)
Inffeldgasse 12, A-8010 Graz, Austria

ABSTRACT:

Two-way satellite time transfer is a key technique for precise and accurate time comparisons over long distances. The use of pn-coded signals as generated by the MITREX-modem allows the use of small, low-power terminals. Design criteria for such terminals and the possible use of VSAT's will be discussed. Furthermore some cost estimates are given.

1. INTRODUCTION

Two-way techniques are used for high accuracy point-to-point time transfer. Both laboratories require transmit- and receive stations and exchange timing signals via the transponder of a geostationary satellite. The main advantage of this technique is that - if the signals from both stations are transmitted simultaneously - one can neglect the motion of the satellite and hence station-to-satellite delays cancel out. The accuracy of the time transfer mainly depends on the knowledge of the signal delays in the transmitting and receiving equipment of the ground stations /1/.

In past years several experiments using this technique have been carried out. The highest precisions could recently be achieved using pn (pseudo noise)-coded signals as generated by the MITREX-modem and transmitted via transponders of communication satellites /2/. The importance of this method is stressed by recommendations of the Comité Consultatif pour la Definition de la Seconde (CCDS) and a resolution of the Conférence Générale des Poids et Mesures (CGPM). Nevertheless at present this method is not in wide use. The most important reasons for this are the fact that prospective users are not familiar with the necessary administrative procedures required to obtain access to a space segment and that local telecommunication authorities are sometimes reluctant to issue such permits. Also for many users it is not clear enough which satellite terminals meet the requirements of the MITREX-modem and the various space segments and how much they will have to spend for this equipment. This point shall be discussed in more detail.

2. SPACE SEGMENT

Ku-band transponders (receive frequency range: 10.95 GHz - 12.75 GHz, transmit frequency range: 14.00 GHz - 14.50 GHz) are available worldwide on Intelsat (International Satellite Organisation) satellites and on Eutelsat (European Satellite Organisation) satellites in Europe and on various domestic and experimental satellites making use of certain frequency bands dependent on the satellite system and the service region.

The frequency bands used in Europe are the following:

Receive frequency bands (down-link):	Transmit frequency bands (up-link):
10.95 GHz - 11.20 GHz	14.00 GHz - 14.50 GHz
11.45 GHz - 11.70 GHz	
12.50 GHz - 12.75 GHz	

For different frequency bands and satellite systems the receive frequencies are related to the transmit frequencies in different ways. The difference between transmit- and receive frequencies is called transponder translation or conversion frequency.

For the Eutelsat satellites presently in use (ECS-1, ECS-2, ECS-4, ECS-5) the nominal conversion frequencies are 3.30 GHz, 2.55 GHz and 1.50 GHz as can be seen from the frequency plan given in Fig. 1. According to the stability of the on-board oscillator this frequencies may change by some kilohertz per month and by about ±50 kHz over the lifetime of the satellite /3/. This behaviour of the conversion frequency also applies to other satellite systems.

The resolution of the allocated frequencies depends on the service used and is for example for the Intelsat Business Service (IBS) and the Eutelsat Satellite Multiservice System (SMS) for "occasional use" assigned on a 22.5 kHz grid (see Fig. 2) requiring such a frequency agility in the user equipment. For other services for example a step size of 1 MHz is sufficient.

All these and other relevant informations are available from the operational service section of the specific satellite system the prospective user is interested in.

3. MITREX-MODEM

The MITREX (Microwave Time and Ranging Experiment)-modem has been developed by Prof. Ph. Hartl, Germany, for point-to-point time transfer of highest precision and accuracy using geostationary satellites and in order to perform high accuracy ranging to these satellites /2/. It makes use of the spread-spectrum technique to transform the one pulse per second (1PPS) of a clock into a binary sequence with a chip-rate of 2.5 MHz and provides eight different pn-codes. The modulation scheme used is BPSK (binary phase shift keying) and according to the chip-rate of 2.5 MHz to transmit the signal a bandwidth of 5 MHz is sufficient, but may be further reduced by appropriate filtering.

The modem uses a fixed intermediate frequency (IF) of 70 MHz whereby the center frequency of the received signal is only allowed to deviate by some kilohertz from the above mentioned 70 MHz in order to keep delay variations below 0.5 ns /4/. A new microprocessor controlled version of the modem allows for larger frequency deviations and has an electronic counter with high resolution installed /5/. Also this modem, however, provides no real frequency agility.

The signal delay of the modem is practically constant for a wide range of the operational parameters. For a carrier-to-noise power density ratio (C/N_0) above 55 dBHz the typical measurement uncertainty is below 1 ns with a limit of about 0.1 ns.

4. SATELLITE TERMINAL

Using satellites operating at Ku-band frequencies the MITREX-modem can be used with earth stations employing antennas of about 1.8 m diameter, state of the art low-noise pre-amplifiers (LNA's) and high-power amplifiers (HPA's) of a few Watt or even less and providing an IF of 70 MHz. For a sample link budget see /6/.

Because of the fixed transmit frequency of 70 MHz and the narrow receive frequency range of the MITREX-modem the satellite terminal has to provide the necessary frequency agility in order to match the assigned frequencies and to compensate for the varying conversion frequencies of the satellites. The step size for the adjustment of the uplink frequency is determined by the satellite service used and for the downlink the step size has to be in the kilohertz range (1 kHz for the original version) to meet the requirements of the MITREX-modem.

These conditions are easily fulfilled by earth stations of "conventional type" but at fairly high costs (see Table 1). Another and substantial cheaper solution makes use of "Very Small Aperture Terminals" (VSAT's) which are now produced in large quantities for a wide range of communications applications /7/.

Fig. 3 shows the functional block diagram of the transceiver of a VSAT revealing typical properties of these systems such as limited frequency range, fixed relationship between uplink- and downlink frequency, 1 MHz step size and integrated design for use as an out-door unit (ODU) attached to the antenna structure. However, employing two independent synthesizers in the up- and down converter and exchangeable low-noise converters (LNC's) in order to cover all frequency bands of interest one can use such a terminal with different satellite systems, assuming the modems provide the necessary additional frequency agility. Fig. 4 gives the frequency plan for a modified version of the transceiver shown in Fig. 3 which allows for the independent use of any transmit and receive frequency within the reqired transmit and receive bands (in Fig. 4 for example 14.00 GHz - 14.50 GHz and 10.95 GHz - 11.20 GHz). This modification and exchangeable LNC's for all commonly used frequency bands are available from the manufacturer.

The necessary frequency agility can be achieved by additional up- and down conversion stages between the MITREX-modem and the transceiver employing synthesizers with the required resolution to adjust any assigned frequency and to meet the requirements of the MITREX-modem (see Fig. 5).

5. DELAY CALIBRATION

In order to measure the difference of the delays in the transmit- and receive chains of the stations the method proposed in /8/ can be used (see Fig. 6.). The necessary devices needed in front of the antenna and the cables should be permanently installed.

In order to achieve the highest accuracy the delay differences between the stations involved should be measured from time to time by a portable station /7/.

6. CONCLUSION

For two-way satellite time transfer VSAT's are a useful and cheap alternative to earth stations of the conventional type. The shortcomings of standard VSAT's such as commonly used for communications can easily be eliminated by the use of independent synthesizers in the up- and downlink and by obtaining the necessary frequency agility - neither provided by the VSAT nor by the MITREX-modem - by means of a frequency conversion stage between the MITREX-modem and the VSAT-transceiver employing synthesizers with the required resolution. The costs for such a terminal are only fractions of the costs one has to spend for a station of conventional design (see Table 1).

7. ACKNOWLEDGEMENTS

The helpful discussions with many colleagues and the loan of a MITREX-modem from the U.S. Naval Observatory are deeply appreciated.

The work was made possible by various grants of the Austrian Council for Scientific Research, the Austrian Academy of Sciences and the Jubilee Fund of the Austrian National Bank.

8. REFERENCES

/1/ Kirchner, D. and W. Riedler:
Timing by Satellite: Methods, Recent Developments and Future Experiments, ESA SP-245, pp.175-186, 1985

/2/ Hartl, Ph., N. Gieschen, K.M. Müssener, W. Schäffer and C.M. Wende:
High Accuracy Global Time Transfer via Geosynchronous Telecommunication Satellites with MITREX, J. Flight Science and Space Res., Vol. 7, No. 5, pp.335-342, 1983

/3/ ECS Data Book, ESA, 1983

/4/ Kirchner, D., H. Ressler and A. Söring:
Messungen mit dem MITREX-Modem, Actes du Congrès de Chronométrie, CP 6, pp.71-75, 1986

/5/ Thiel, K.H.:
Personal Communication, March, 1990

/6/ Veenstra, L.B and D.W. Hanson:
Two-Way Satellite Time Transfers between and within North America and Europe, Proc. 43th Annual Symp. on Frequency Control, 1989

/7/ Howe, D.A.:
High Accuracy Time Transfer via Geostationary Satellites: Preliminary Results, IEEE Trans. on Instr. and Meas., IM-37, pp.418-423, 1988

/8/ De Jong, G.:
Accurate Delay Calibration of Satellite Ground Stations for Two-Way Time Transfer, Proc. 3rd European Frequency and Time Forum, pp.198-203, 1989

	"Conventional type" (individual design)		"VSAT type" (large quantities)	
Antenna plus feed and pedestal		3,000		
Upconverter	1 kHz steps	35,000		
HPA		7,000		
Downconverter	1 kHz steps	35,000		
LNA		5,000		
Filters, waveguids, etc.		5,000		
Assembling		10,000	1 MHz steps	27,000
Adaptations for 1 kHz steps		---		7,000
		100,000		34,000
		=======		======
Calibration			6,000.00	
Steerable antenna (azimuth, elevation, polarization)		35,000		

Table 1. Cost estimates in US $ for Ku-band stations of conventional and VSAT type. Both stations cover the same frequency bands and have 1 kHz step size, however with the VSAT one has to exchange the LNC's to change the frequency bands and it provides no remote frequency control. Therefore a steerable antenna is only resonable for the station of conventional type.

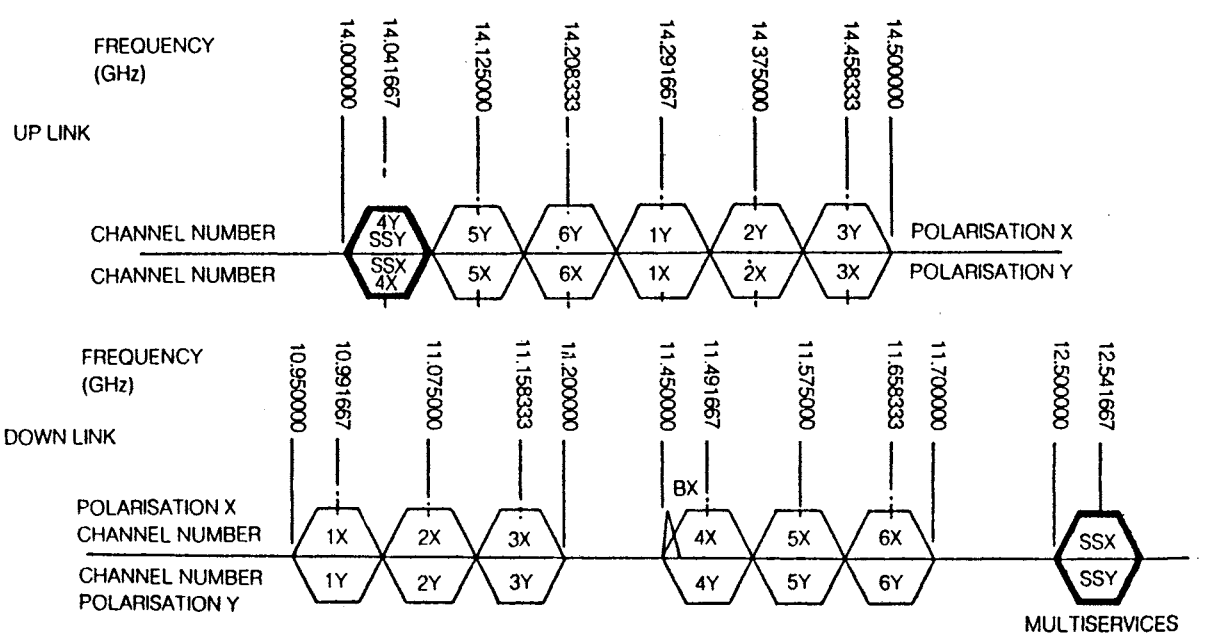


Fig. 1. EUTELSAT (present EUTELSAT satellites ECS-1, ECS-2, ECS-4, ECS-5) frequency and polarization plan (from /3/).

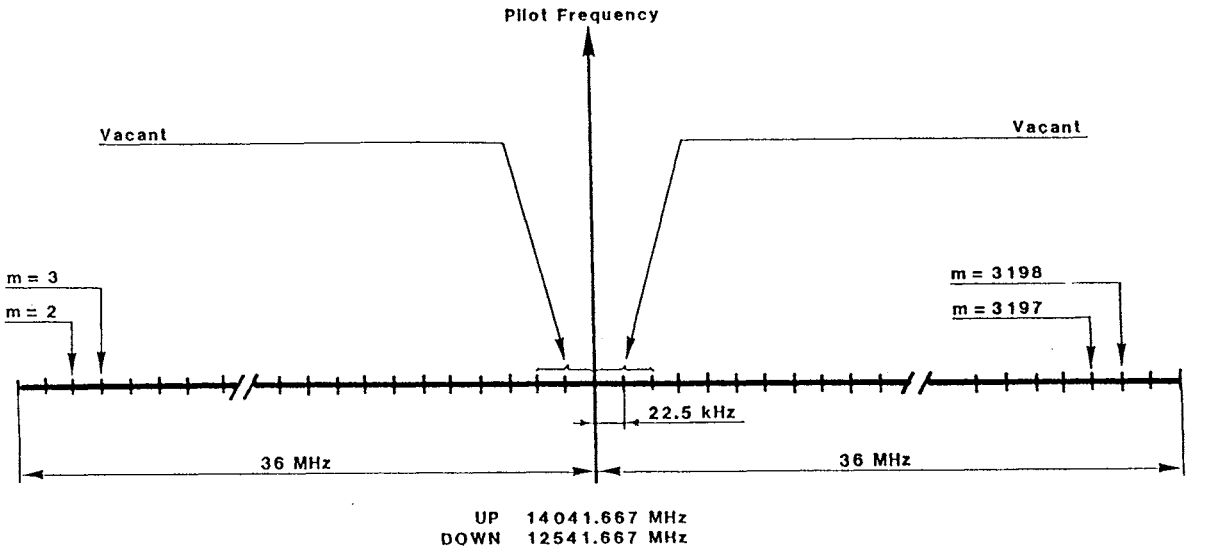


Fig. 2. EUTELSAT transponder frequency plan (from /3/).

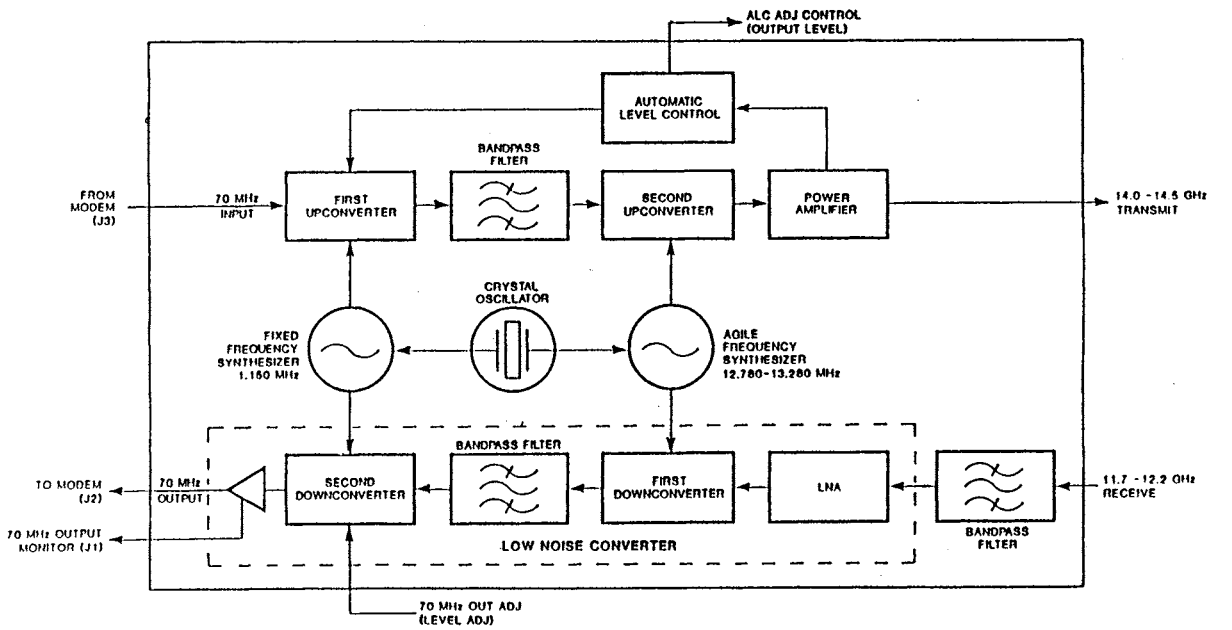


Fig. 3. ASAT 1214 Functional block diagram (from ASAT 1214 Operation and Maintenance Manual, SEE Technologies).

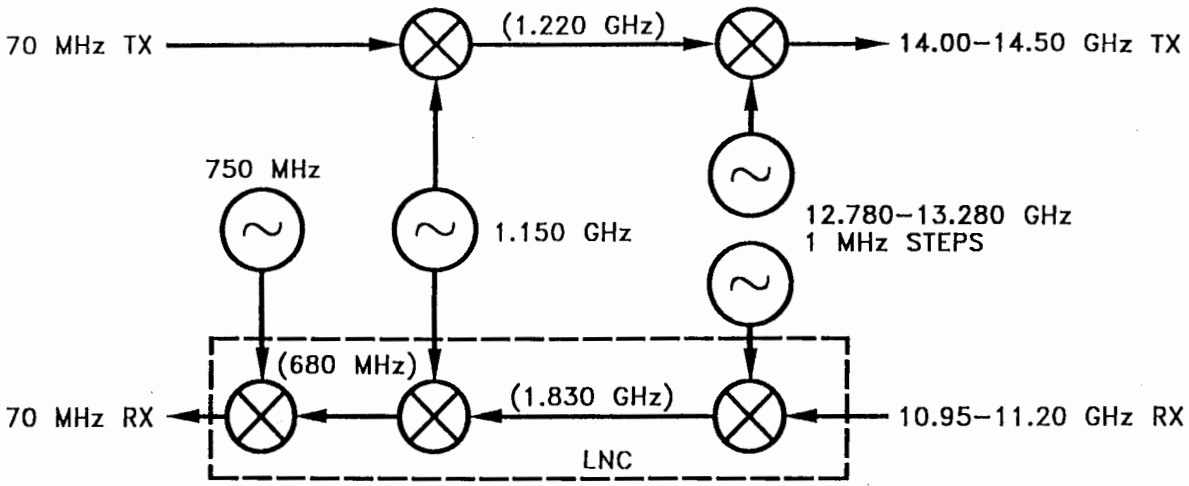


Fig. 4. Frequency plan of ASAT 1214 using independent synthesizers in the up- and down converter and adapted for the use of another receive frequency band (10.95 GHz - 11.20 GHz instead of 11.70 GHz - 12.20 GHz as given in Fig. 3.).

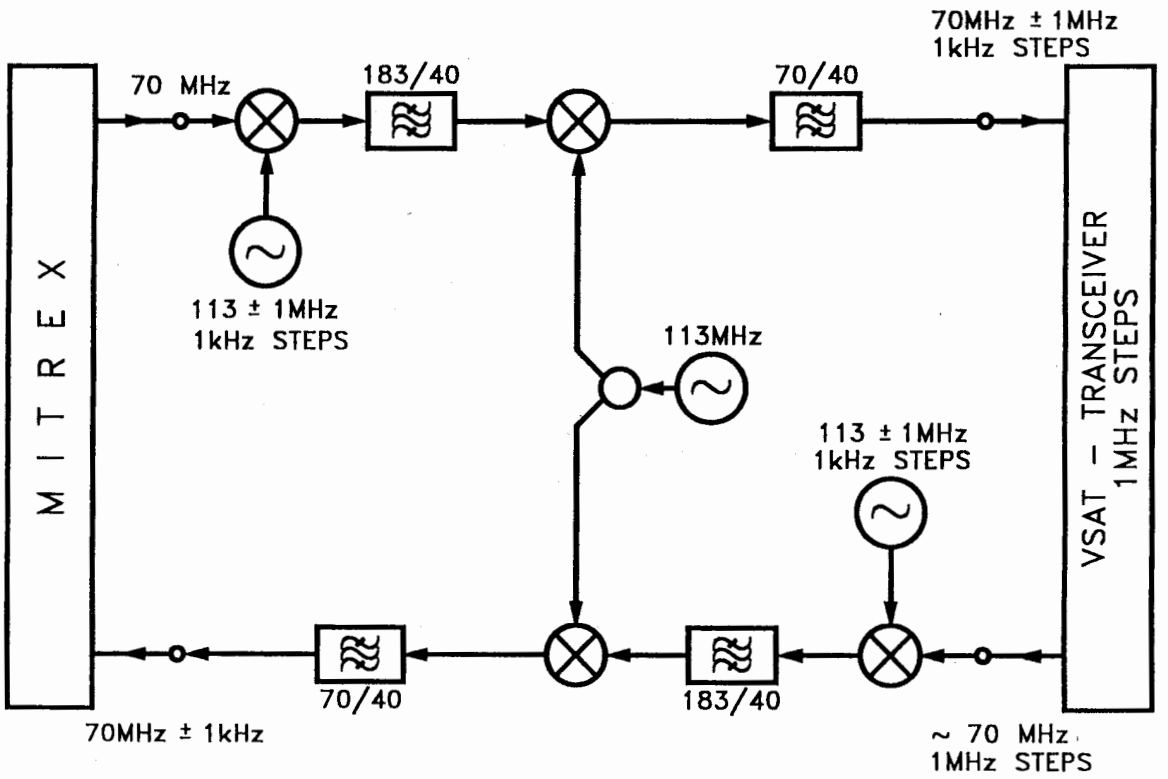


Fig. 5. Frequency conversion scheme between MITREX-modem and VSAT in order to obtain the necessary smaller frequency steps in the up- and downlink.

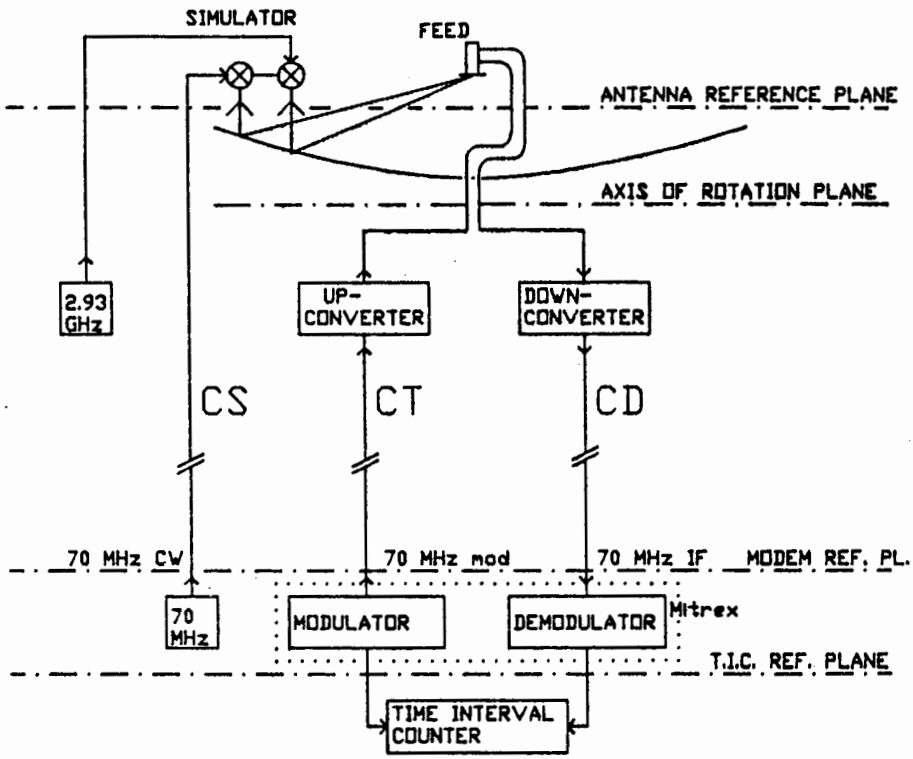


Fig. 6. Set-up for station delay calibration (from /8/).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP 04 Number 3

FURTHER DEVELOPMENT OF THE ECS TIME SYNCHRONIZATION EXPERIMENT

O. Buzek, J. Cermak (URE CSAV, Prague)
J. Vondrak (AsU CSAV, Prague)
V. Pettiti, F. Cordara (CNR IEN, Torino)
A. Cenci (Telespazio, Roma)

Further results of the undergoing time synchronization experiment between the time scales of some European laboratories based on the passive television method using the ECS-5 geostationary satellite are presented. A practical attempt has been made to reduce the measurement uncertainty due to the satellite motion using methods based on the satellite ranging from a ground station and on eliminating the effect of the periodical motion by averaging two values measured at half sidereal-day intervals.

Session APP-04 Number 4

PRAREE- The Extended Version of the Precise Range and Range Rate Equipment

Philipp Hartl
University Stuttgart
Institute of Navigation
D-7000 Stuttgart
F.R. Germany

Abstract

The Precise Range and Range Rate Equipment PRARE is an experiment of the Federal Republic of Germany, funded by the Minister of Science and Technology (BMFT) for it's first use during the ESA Earth Observation Mission ERS-1. It is a spaceborne, dual frequency, two- way microwave ranging system for satellite orbit and geodetic position determination with dm- and cm-precision. The follow-on system for use on Polar Platforms is called PRAREE. It aims at cm- and mm- accuracy to meet the needs of the next decades. In order to achieve this accuracy the system improves the ranging precision by applying PRN- ranging modulation with a clockrate of 100 MHz. For correction of signal propagation effects it uses parallel to the microwave links a one-way optical link, which operates with ranging PRN modulation signals fully coherent fully coherent to the one of the microwave link. This allows by comparison of its difference in the signal delay to determine the influence of the wet component of the troposphere. Also the dry component of the troposphre can be measured; for this purpose a two- colour laser link is foreseen. These laser links apply semiconductor lasers of moderate power.

PRAREE system goals

The Polar Platforms will give opportunities to perform new experiments with and developments of new instruments and instrument systems. A new quality of performance can be expected to become reality by making use of the advantages of the powerful platforms. There will be less constraints on the instruments with respect to mass, volume, electric power, data rate and data storage capacity. Synergetic effects can be expected from the fact that many other sensors will be aboard, which will deliver complementary information. Also the infrastructure will be become much better and more efficient. The missions will allow to analyse the system performances and to introduce improvements in an evolutionary process by replacement of existing instrument modules by new ones.

PRAREE follows this idea: In this spirit a proposal was submitted to ESA in response to the Announcement of Opportunity (AO) of ESA /1/ , the objective of which is to improve existing ranging techniques and to develop new methods, which allow to meet the increasing accuracy demands of geoscience.

It is true that the present PRARE system offers already very high accuracy in the order of a few cm and some fractions of mm/sec in velocity. But for the next two decades at least an order of magnitude better values are requested. It is expected for position determination on a global basis and for satellite orbit determination.

Many improvements are needed to achieve this goal. The uncertainties of the gravity field must be excluded and the signal propagation effects need to be modelled or measured much better than today. Also the precision of the ranging instrument itself must be improved by new developments.

From the presently running feasibility study one can expect the following performance data

- Microwave ranging error of raw data < 2 cm for 1sec integration time
- Microwave plus optical ranging: 0.5 to 1 cm for 1sec integration time
- Micorwave ranging rate resolution <0.04 mm/s for 15 sec integration intervals

The system will be capable to operate with up to 10 ground stations simultaneously. The importance of this performance can be seen from table 1 which identifies scientific goals in comparison on possible performance qualities of various tracking concepts.

Table 1: Scientific goals in comparison with the present performance of various tracking concepts

SCIENTIFIC TOPIC	REQUIRED ACCURACY	LASER	(DIFF.) GPS	MW DOPPLER	PRARE	PRAREE
Orbit Determination						
Real-Time On-Board	10 m on-board	o	****	o	o	****
Near Real Time Oper. Center	1m/6h Op.Center	o	***	*	**	****
Off-Line Oper. Center	2-10cm/7d	***	*	*	**	****
Solid Earth						
Intraplate Tectonics	<1mm - 5mm/yr	*	*	o	**	****
Interplate Tectonics	<1mm - 5mm/yr	*	*	o	**	****
Post Seismic Strains	< 1 mm/d	o	*	o	**	****
Earth Tides	< 5 mm	*	*	o	**	****
Tidal Loading Effects	< 5 mm	*	*	o	**	****
Reference to Geocenter	< 1 cm	*	o	o	**	****
Ocean Physics						
Sea-Surface Topography (+RA)	< 1cm	*	*	o	**	****
High Precision Geoid (+RA)	< 1cm	*	*	o	**	****
Circulation and Eddies (+RA)	< 1cm	*	*	o	**	****
Submarine Topography	< 5cm	o	**	o	**	****
Sea Floor Positioning	< 5cm	o	**	o	**	****
Coupling of Ocean and Atmosphere	< 1cm	*	*	o	**	****
Local Ocean Tides (+RA)	< 1cm	*	*	o	**	****
Atmosphere						
Atmospheric Propagation	< 1cm	***	*	o	**	****
Ionosphere (TEC)	< 5 mm	o	**	o	****	****
Troposphere (Wet Term)	< 5 mm	o	o	o	o	****
Atmospheric Loading	< 1cm	*	*	o	**	****
Angular Momentum Budget	< 0.1 msec	*	*	o	**	****
Support of RA for Atmospheric Correction	< 5 mm	o	*	o	**	****
Climatology						
Eustatic Sea-Level Changes	< 1 cm/yr	*	*	o	**	****
Ocean Circulation and Green House Effects	< 1 cm/yr	*	*	o	**	****
Geodesy						
Global Point Positioning	< 10 cm/0.5 yr	***	*	*	****	****
National Control Networks	1 cm/d/500 km	*	***	o	**	****
Engineering Surveying	1 cm/5 km	o	***	o	**	****

Valuation Criteria:

****	Full capability including required accuracy and near real-time operational requirement
***	Similar to (****) provided that additional data transmission network is present
**	Capability with modest accuracy and near real-time operational requirement
*	Similar to (**) provided that additional data transmission network is present
o	No capability with respect to accuracy and real-time operational requirement

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-04 Number 5

TIME TRANSFER IN THE UNITED KINGDOM VIA A DIRECT BROADCASTING SATELLITE

P R Pearce, J A Davis, D R Allen, B R Swabey, R Jefferies,
D J E Knight and P Menary*

National Physical Laboratory, Teddington, Middlesex, TW11 0LW, UK.
* British Telecom International, Rugby Radio Station, Hillmorton, Rugby, CV23 0AS, UK.

ABSTRACT

The introduction to the United Kingdom of satellite TV via direct broadcasting satellites offers a simple technique of time transfer at the microsecond level. A time transfer by common view of the Astra satellite requires only inexpensive domestic receiving equipment, including a small 65cm diameter dish. Time comparisons are possible, both across the United Kingdom and with other European countries.

We report time comparisons between NPL Teddington and Rugby Radio Station using caesium atomic clocks as references at both locations. Time transfer in common view mode via the Astra TV satellite and via a local BBC TV station are compared. Uncertainties in the time transfer method via the Astra satellite are evaluated and the sources of error limiting the accuracy are discussed.

INTRODUCTION

The use of TV satellite signals for time comparisons has been investigated over three decades [1], [2], including work at NPL [3], proposals for the use of high powered satellites [4], and a collaborative common-view experiment currently in progress [5]. The present investigation of the use of signals from a direct-broadcasting satellite was stimulated by the low cost, compactness and portability of the receiving equipment, and by the desire to investigate an alternative method to the MSF 60kHz broadcast for frequency calibrations of caesium standards at a distance from NPL. There is also a growing requirement for low cost epoch synchronization of stations belonging to distributed communication networks such as the Pan-European cellular telephone system. These require an order of magnitude greater accuracy than can readily be obtained from the modulation pattern of the MSF 60kHz signal.

One attribute of the present TV signals from geostationary satellites is that the carrier frequency (11GHz) is high enough almost to neglect the effects of ionospheric fluctuations. In addition, signals for direct broadcasts are strong (the Astra satellite has a 50dBW EIRP), and continuously available, allowing continuous monitoring with automatic equipment. Our program was directed to finding an effective and simple method of applying the TV signal to caesium clock calibrations, if possible without detailed correction for the position of the satellite. This paper reports measurements made between 8th November 1989 and 22nd February 1990.

METHOD

Experiments were set up to assess time transfer via the Astra satellite using three low-cost domestic receivers. Initial experiments were made with two receivers side by side at NPL with a common clock, to explore delays and noise in the receiving, pulse extracting, and timing equipment. The third receiver system was installed at Rugby Radio Station, which provided a 120km separation and a local caesium clock which was already subject to daily comparison with NPL via the ground-based TV transmitter at Sandy Heath. Later experiments at NPL included the use of separate caesium clocks on side by side receivers to establish a baseline for comparisons of caesium clocks at a distance. The caesium clocks were Hewlett Packard model 5061A, one of which was an option 004 high stability version.

A schematic diagram of the apparatus is shown in Figure 1. It is based on the standard domestic Astra receiving system, comprising 65cm dish, 11GHz downconverter and TV set. A pulse extractor, based on a design developed at the Van Swinden Laboratories, extracts the frame synchronisation pulse from the video output signal of the TV set. A 1ns resolution counter timer is referenced to a 5MHz caesium-derived standard frequency. This is used to time the difference between a pulse originating from a caesium clock or other clock under examination, and the frame synchronisation pulse extracted by the pulse extractor. The caesium clock pulse occurs every second, whereas the frame synchronisation pulse occurs every 20ms. There is a drift between these two chains of pulses due to the reference oscillator used to generate the 20ms frame synchronisation. System time is obtained from a radio-controlled clock linked to the UK MSF broadcasts. The signal strength entering the satellite receiver is also monitored. The results obtained from the counter timer are then processed by computer.

DATA LOGGING AND ANALYSIS

A 1-s pulse from the local caesium clock starts the counter timer. The time interval to the next frame syncronization pulse, which should not exceed 20ms is measured. A delay of about 0.5s is allowed before reading the counter timer buffer. Thus one data reading is obtained at each station every second. In order to avoid noise in the transfer originating from the TV frame synchronization pulse control oscillator, accurate system clocks are used to observe the TV signal at the same second.

Errors can occur when the 20ms pulse occurs close to the time when the 1s clock pulse arrives. Because of the differences in delays to different observing ground stations and the differences in the clocks being compared, the frame synchronization pulse may occur before the 1s pulse at the first station and after it at the second, the result being that different frame synchronisation pulses are being read and the frame-synchronisation reference oscillator is being used to count time between these pulses. This results in an increase of an order of magnitude of the noise level observed in the readings. For short-distance time transfers within the UK this effect happens for only a short time (2-3 hours every 3-4 days); whereas for longer time transfers across most of the Astra footprint, up to 5% of the observing time could be hampered by this effect. Referencing the pulses to MSF with a suitable offset so that all ground stations observe the same TV synchronization pulse appears to be a solution. In our measurements data with a 20ms slip was normally rejected, leading to the results at line(g) rather than line(h) of Table 1 below.

Another source of error is from occasional failure of the pulse extractor boxes to trigger when the satellite receiver signal is weak so that the interval recorded exceeds 20ms, perhaps by several frame intervals. Such points were detected and deleted at each ground station before returning the data to NPL.

Data was recorded every second for two minutes each hour to give 120 readings, and in periods of 14 minutes, four times an hour, the mean of each period being based on 840 readings. The latter data was split into 60s sections, and the mean of each section was recorded. The differences between these section means for the two separate channels were used in comparisons. If any trigger failings occurred then the whole of the minute section was rejected. The mean of 14 of these section differences was used to produce one data point. The 15th minute was used for computation, data checks, and monitoring the signal strength.

Allan variances were calculated both from the differences of single 1-s data comparisons between two channels and using the means of the differences of the 14 blocks of 60s (ie 840s) data as values. The sqrt(Allan variance) entries in Table 1 have a 1s averaging time, using single data points, or 840s averaging time using blocks of data.

COMPARISONS AT NPL

To investigate noise and variable delays in the satellite TV receiving systems, two similar systems were set up side by side at NPL as shown in Figure 1. The start pulse for both counter timers was provided from a common caesium 1s pulse signal. In order to identify contributions from the different elements, tests were made on composite systems, using a single satellite receiver, at left, and progressively working up from the lowest transverse dotted line in Figure 1.

(i) The output from a single receiving system was split between two separate counter timers. The mean delay difference between the two counter timers was approximately 0.7ns. This was within their quoted accuracy, and negligible for these experiments.

(ii) The video output signal from one TV set was split between two pulse extractor boxes and counter timers. The values of σ_y, Δt, and total uncertainty for comparisons between the two counter timer readings of the composite system are given in row (a) of Table 1. The value of the delay difference between the two counter timers showed steps of about 5ns. These we attribute to the pulse extractor boxes.

(iii) The output from a single satellite dish and downconverter was split between separate TV sets, pulse extractors and counter timers, row (b) of Table 1. The TV sets contributed additional noise, and a differential delay of approximately 5ns. The delay steps we attribute to the pulse extractor boxes were still present but no recognisably different delay steps were observed.

(iv) The timings from two completely independent systems were compared, (Table 1(c)). A five day span of data is shown in Figure 2. Besides the 5ns delay steps attributed to the pulse extractor boxes there were occasional larger steps of about 20 ns. These, unlike the 5ns steps, seemed to correlate with high signal strength at the satellite receiver downconverter, and appear to originate in this part of the system. The total uncertainty for the timing difference between the separate receiver systems, including systematic delays observed over five days, becomes ±15ns(Table 1c). This provides a benchmark for comparing the noise of spatially separated receivers.

For the data obtained above, plots of $\log_{10}(\sigma_y)$ against $\log_{10}(\tau)$ for averaging times of 1-10,000s give straight lines with gradient -1. This suggests that all the short-term noise originating in the satellite receiving apparatus is random in nature. The estimated contribution of the random noise of each device to the total uncertainties is ±13ns, ±9ns, and ±20ns for the pulse extractor box, TV set and satellite receiver respectively, considering the Allan variance with 1s averaging time. See the 1s Δt(ns) column of Table 1. $\text{Log}_{10}(\sigma_y)$ against $\log_{10}(\tau)$ plots for longer averaging (1000s - 100,000s) and using the mean values of 14 minutes of comparisons gave approximately straight lines with smaller negative gradients (0.7-0.9). This suggested that the noise was not completely random but the effects of the 'delay steps' were being observed. Further operational tests were made.

(v) One of the NPL systems was degraded by tilting the receiving dish away from the direction of the Astra satellite until the power collected by the dish was reduced to half its original value. Results are given in row (d) of Table 1. The Sqrt(Allan variance) for short averaging times was similar to the value before degrading, suggesting that the random noise remained constant. However the number of 'bad' data points increased, suggesting that the pulse extractor box was not triggering so effectively.

There was a correlation between low signal strength in the satellite receivers and rainy or overcast weather conditions, due to attenuation of the 11GHz satellite downlink signal by water in the atmosphere. The variation in the power level in normal use was no more than 10% of its mean value.

(vi) In order to examine the effects of the atmosphere over a relatively short distance, a common clock experiment was performed at the NPL site, with one of the receiving systems moved approximately 0.5km from the other. The Sqrt(Allan variance) (Table 1(e)) was similar to that obtained with the receivers side by side, ie, no additional noise or variable delays were observed.

(vii) A further experiment was carried out with separate caesium clocks on side by side receivers at NPL, Table 1(f), to provide a baseline for comparisons of separate clocks at remote locations.

COMPARISONS BETWEEN NPL AND RUGBY RADIO STATION

A typical time sequence of comparisons made between the caesium clocks at NPL and Rugby Radio Station via common view of the Astra satellite is shown in Figure 3. A periodic 24hr variation in the delay is seen, this being due to the imperfect geostationary orbit of the satellite about the earth. The magnitude of this variation slowly changes, and a typical maximum amplitude of 200ns (peak to peak) was observed. Approximately every 15 days the satellite position is corrected through its station keeping, and on occasions the periodic motion was reversed. The satellite then drifted through a near-perfect geostationary orbit before becoming again misplaced.

The difference in delay times between signals reaching the NPL site and Rugby Radio Station was approximately 450 microseconds with approximately 100 microseconds being due to the difference between the local clocks. The value of the delay calculated from the known positions of the two ground stations and the Astra satellite differed from this by 700ns, with a maximum uncertainty of the order of 500ns. In order to obtain better absolute time transfer, the ground station would need to be calibrated, possibly with a portable caesium clock time transfer, or the positions of both the satellite and the ground stations would need to be calculated accurately.

Plots of $\log_{10}(\sigma_y)$ against $\log_{10}(\tau)$ for the Astra satellite comparisons are given in Figure 4. The upper curves A and B were derived from sets of data read at 1s intervals for about 2 days, allowing plots for τ = 1s to τ = 30,000s. The lower curves, C and D, were computed from means of the blocks of 840 1s data points lasting for about 10 days, allowing (overlapping) plots for τ = 900s to about 3 days. Plots A and C refer to separate clocks at NPL and at Rugby, plots B and D refer to separate clocks on side by side satellite receivers at NPL. For averaging times of 1s to 5,000s the plots A and B are linear with gradient -1, as for comparisons between the two NPL channels above, suggesting that almost all the short-term noise in the time-transfer system is due to random noise in the satellite receivers and timing equipment. For τ longer than 10,000s, the plots A and C show oscillatory behaviour, with minima (plot C) occurring at averaging times which are whole numbers of sidereal days. This is due to averaging the effect of the 24 hour periodic motion of the satellite. Other sources of noise include the caesium clocks, the atmosphere and ionosphere. Plot D is for side by side receivers at NPL with separate local clocks, showing the best that might be expected from any time transfer. The effects at the 11GHz carrier frequency of the atmosphere and ionosphere should be limited to a few ns [6], which may be neglected in our comparisons.

The most accurate frequency comparison of caesium clocks at NPL and Rugby will be obtained from averaging times of 1 day or multiples of 1 day, from plot C. However, the value of the Sqrt(Allan variance) for clocks side by side, plot D, for averaging time 84000s (1 day) is 2×10^{-13}, somewhat lower than the value of 8×10^{-13} for the comparisons between NPL and Rugby Radio Station. This suggests that averaging over 24 hours does not completely eliminate the effects of satellite motion. Non-periodic drifts in the satellite's longitudinal position will not have been corrected.

A comparison is made daily at 1100h between the caesium clocks at NPL and Rugby Radio station via common view of the BBC Sandy Heath transmitter (Figure 5). To compare directly with only 1100h comparisons via the Astra satellite would make the readings highly susceptible to corrections of the satellite orbit, and so the mean of 24hr of Astra readings was taken (Figure 6). The data is improved because the instability in the clocks (variation from uniform drift rate) is less than the errors induced by satellite motion. The reception of the Sandy Heath link at NPL is poor and highly weather-dependent due to the low-lying position of NPL, and there is a limit of resolution of 100ns in the readings obtained. A best fit was obtained for the variation of the delays observed from the two clock-comparison methods. The maximum discrepancy obtained was 150ns and the standard deviation of the differences in readings was 65ns. Most of this variation can be accounted for by the noise on the Sandy Heath link.

We conclude from the Sandy Heath comparison that the stability in the Astra link for daily comparisons is better than ±65ns and probably of the order of ±50ns. The largest systematic error causing variations of the delay is from the satellite motion. Any north-south or altitude variation will produce a slowly-changing 24hr periodic motion. This variation may be averaged out or corrected to reduce its value well below the typical maximum of 200ns observed. The longitudinal variation of the satellite will show no periodic effects but will introduce a systematic error. The NPL-Rugby link being almost north-south will be fairly insensitive to these variations. However a 0.05 degree error, the maximum allowed by the satellite station keeping would produce an error of 70ns. An error of 0.03 degrees in a north south direction would be required to produce a 100ns delay variation (half the 200ns peak to peak

delay variation); such an error in the satellite longitude may be more typical, and would result in a delay of 40ns. Taking this as an estimated systematic uncertainty and combining with the ±15ns total uncertainty between the two receiver systems, Table 1(c), leads to an estimated total uncertainty (1-σ) for the NPL-Rugby link of approximately 50ns, Table1(g).

CONCLUSIONS AND FUTURE WORK

As is stands, the uncertainty in the Astra link for daily comparisons is estimated to be of the order of ±50ns. In order to improve on this, and for more accurate transfer of time epoch, and for transfers over longer distances, the position of the satellite needs to be determined more accurately. There are two possible methods.

The ranging information obtained for station keeping might be used to fix the satellite position, if made available. For Astra satellite ranging the uncertainties in the position of the satellite are of the order of 50m and would result in errors in time transfer of the order of 1ns for long-path time transfers within the UK. This is far less than the errors observed in the satellite receivers and counting equipment.

The second method involves placing four or more widely-spaced ground stations within the footprint of the satellite. Assuming the maximum error in measuring the delay to each of the satellites is ±15ns, the maximum uncertainty introduced by the satellite receiving equipment, and that the time scales at each location are known to ±10ns then the errors in the position of the satellite will be of the order of 1000m. This would result in an error of 20ns for long-distance time transfers within the UK.

Improvements may possibly be made in the design of the pulse-extractor boxes and in the choice of the domestic TV sets used, to minimise the noise and differential delay steps observed. In the near future, the British Satellite Broadcasting system begins broadcasting, so that direct comparisons between two or more independent satellite-based common-view systems becomes a possibility.

In conclusion, the use of domestic receiving equipment for common view observation of a direct broadcasting satellite provided a time transfer link between NPL and Rugby Radio Station stable to ±50ns for daily comparisons. This technique might form the basis of a low cost time-transfer service within the UK, useful for the frequency calibration of caesium clocks at remote locations. A timing resolution of sub-microsecond accuracy is made available by building on the millisecond resolution of the MSF 60 kHz signal. A further use might be for synchronising stations of digital communication networks.

REFERENCES

[1] Tolman J., Ptacek V., Sancek A., Stecher R., : Microsecond clock comparison by means of TV synchronized pulses, IEEE Trans. Instr. and Meas. vol. IM - 16, pp 247-254, Sept 1967.

[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 October 1985.

[4] Takahashi F, : A proposal of accurate time dissemination using TV signals from a broadcasting satellite, Journal of the Radio Research Laboratory, vol. 23, No 112, Nov 1976, pp 267-283.

[5] 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, Third EFTF, Besancon, 1989, pp 204-214.

[6] Gray D A : Time transit variations in line of sight Tropospheric propagation paths, Bell system technical journal, July-August 1970, pp 1059-1068.

TABLE 1

EXPERIMENTAL ARRANGEMENT	σ_y 1s $x10^{-8}$	840s $x10^{-12}$	Δt(ns) 1s	840s	Total uncertainty* ns
COMPOSITE SYSTEM (a) Separate pulse extractor box	1.3	2.0	13	1.7	±5
(b) Separate pulse extractor box plus TV	1.6	4.0	15	3.3	±6
SEPARATE SYSTEMS AT NPL (c) Normal systems	2.5	3.2	25	2.7	±15
(d) One signal degraded	2.5	6.3	25	5.3	±30
(e) Common clock separate locations	2.5	4.0	25	3.3	---
(f) Separate clocks common locations	3.2	6.3	32	5.3	---
SEPARATE SYSTEMS AT RUGBY AND NPL (g) Normal systems	2.5	4.0	25	3.3	±50
(h) 20ms slip	5	-----	50	---	---

* Root sum of squares of random and estimated systematic components.

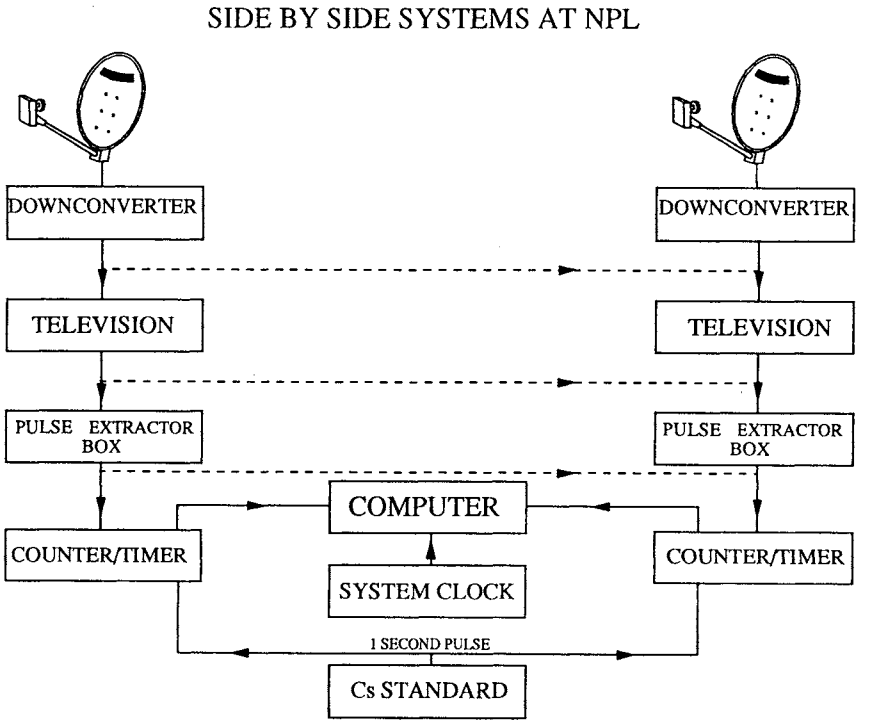


FIGURE 1

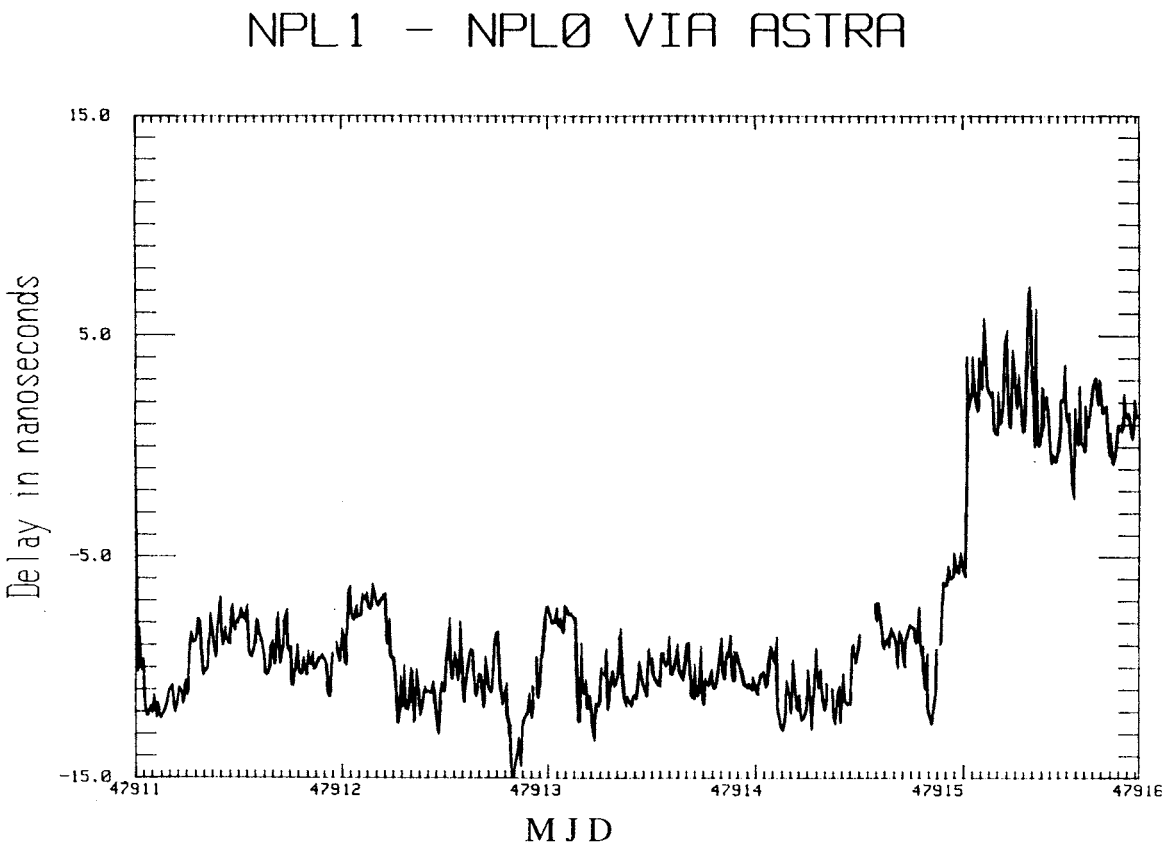


FIGURE 2

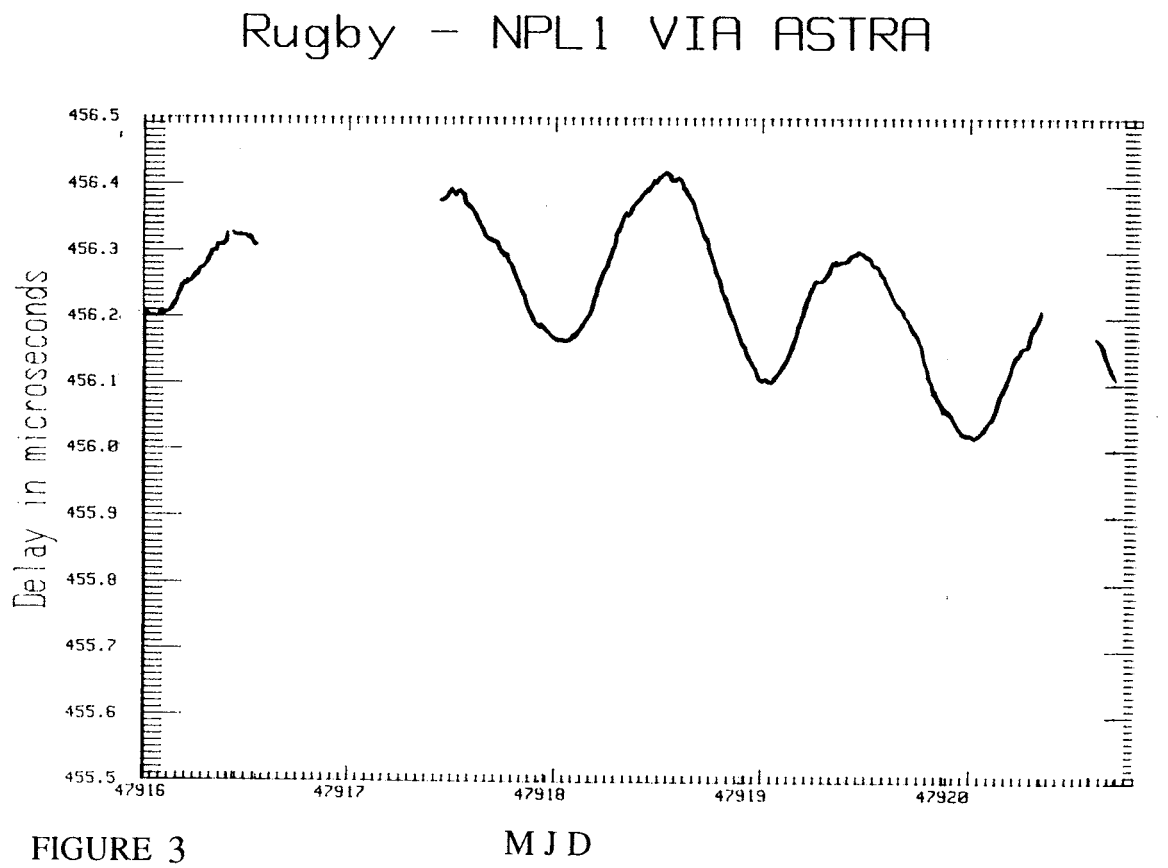


FIGURE 3

PLOTS OF LOG10(σ_y) AGAINST T

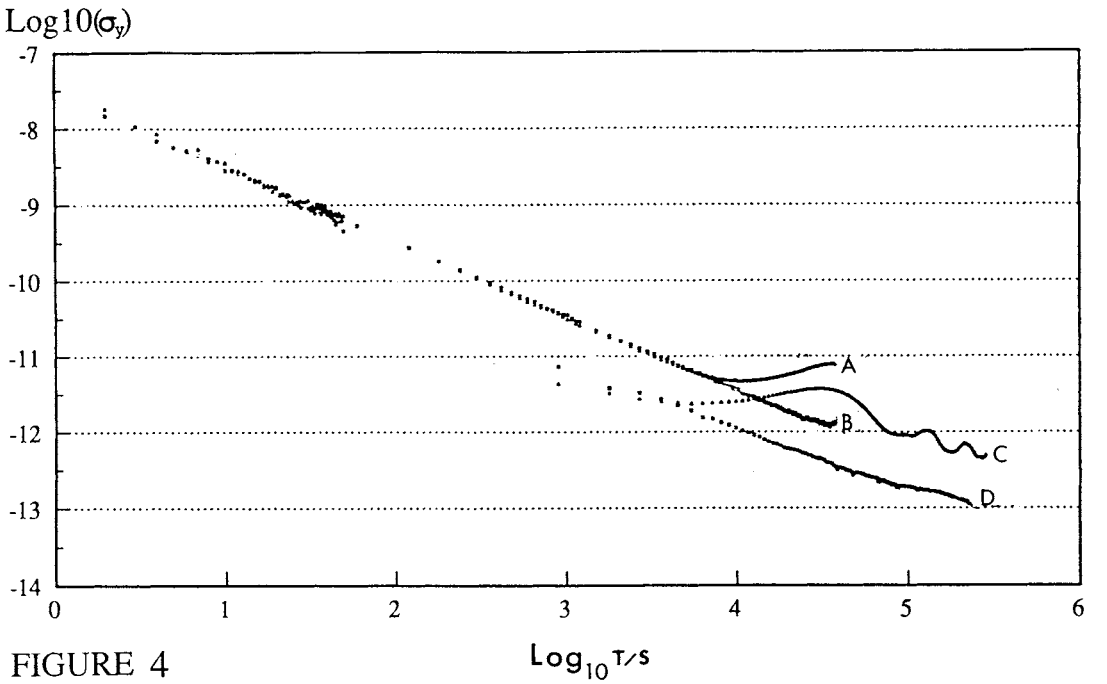


FIGURE 4

NPL - RUGBY VIA SANDY HEATH

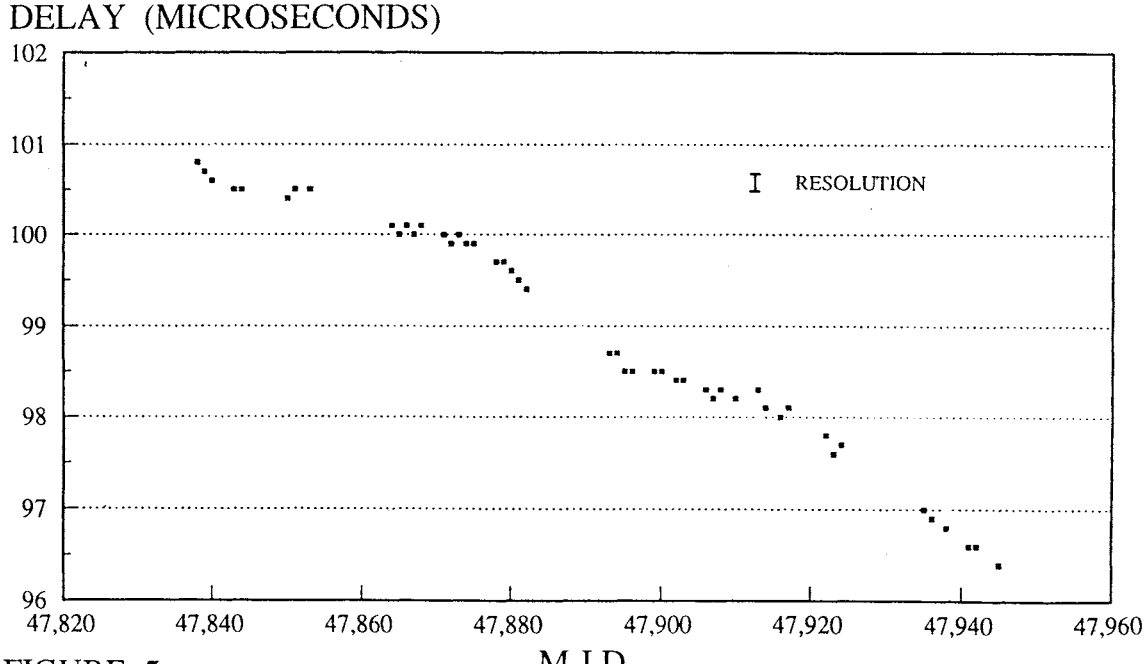


FIGURE 5

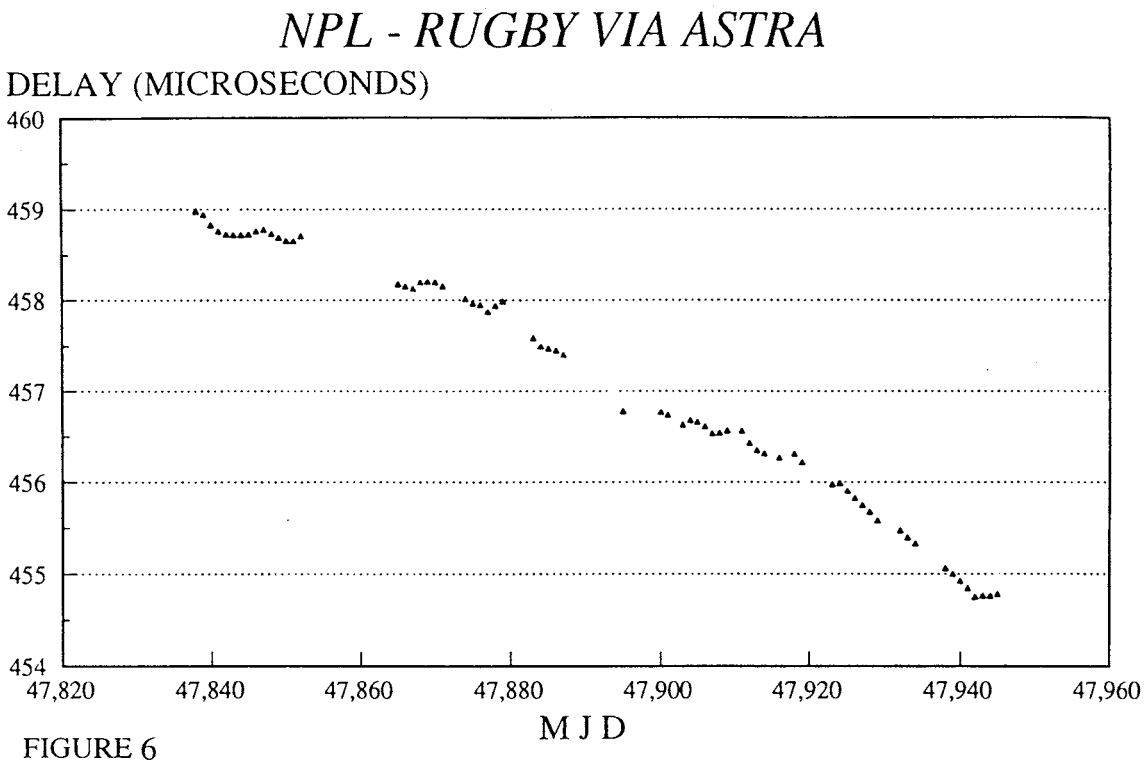

NPL - RUGBY VIA ASTRA
DELAY (MICROSECONDS)
460
459
458
457
456
455
454
47,820
47,840
47,860
47,880
47,900
47,920
47,940
47,960
M J D

FIGURE 6

Resonators, Oscillators and Materials

ROM-05 Material and Processing

Chairman: W. Zingg

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-05 Number 1

Technology and characteristics of chemically milled miniature quartz crystals

B.Studer, W.Zingg

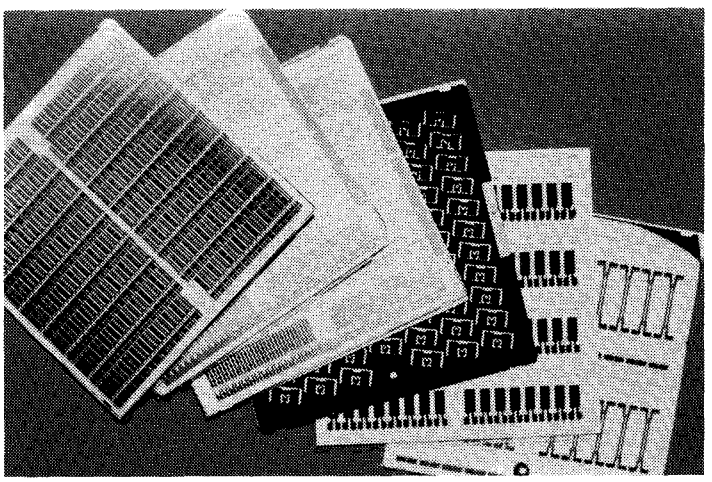
Micro Crystal, Div. of ETA, CH-2540 Grenchen, Switzerland

Abstract

Almost 20 years ago, semiconductor process techniques have been applied to the fabrication of quartz tuning fork resonators. The objective of the batch processing was to manufacture large quantities of cheap miniature and subminiature crystals, for main use in wrist watches.
Although the photolithographic technology was initially developed for watch crystals, its advantages for general purpose quartz crystals have since been recognized and several types of resonators industrialized. The benefits of this technique are lower costs, better dimension control and the possibility to build complex shapes. In this paper, the basic steps of a tuning fork resonator processing are described. Characteristics of typical miniature crystals covering a frequency range from 10 kHz up to 30 MHz, such as tuning forks, length extension mode, ZT-cut and strip AT-cut crystals are discussed.
A further group of photolithographically processed resonators are sensor crystals such as double ended tuning forks, used as force, pressure or acceleration transducers and the temperature sensor tuning fork.

1. Introduction

The fast developing tendency to miniaturization and low cost mass production components in electronics had its effect also on the developments of quartz crystal resonators. The most promising approach for the achievement of these objectives is the application of the manufacturing techniques developed by the semiconductor industries during the last thirty years. With photolithography and chemical etching operations, up to 400 resonators can be produced on a single wafer. Initially, this technology had been introduced for the production of tuning fork crystals for wrist watches [1], the so called "3rd generation " watch crystals. In the meantime, the photolithography has become the standard manufacturing technology for such crystals. The annual world production of these resonators is estimated to approximately 1 billion pieces. The swiss watch industry has adopted this technology 12 years ago; nowadays, the swiss production of these resonators amounts to 40 million pieces a year.After the first experiences with the batch processing of watch crystals, also other resonators have been miniaturized and industrialized, especially in view of the evolution of the surface mounting technology (SMT) in various fields of electronics. The first step was an extension and standardization of the frequencies of tuning fork resonators. Today, fundamental mode tuning forks cover a frequency range from 10 kHz to 200 kHz, whereas overtone tuning forks are used for the 200 kHz to 520 kHz range.



For frequencies between 0.5 MHz and 2 MHz, a length extensional resonator has been developed. Up to 1.5 MHz, a fundamental mode is used, for higher frequencies, the resonator oscillates at an overtone mode [2, 3].
The process steps for the ZT-cut crystal [4, 5, 16, 17] developed for the 2 MHz to 4 MHz range, are currently industrialized.
For higher frequencies, a chemically milled strip type AT crystal is produced; at this time, the range between 8 and 30 MHz is covered.

Besides the above mentioned resonator types, crystals with complex shapes or electrode geometries can be manufactured with the photolithographic process. The advantage of high precision tolerances is especially used for the production of DETF (Double Ended Tuning Fork) [9, 10, 11, 12] force transducers and TTF (torsional tuning fork) temperature sensors [6, 7, 8].

2. Manufacturing Process

The typical manufacturing process for chemically etched quartz resonators consists of three major steps, shown in Fig. 1:

2.1 Wafer cutting

Although this process step is most similar to conventional crystal manufacturing, a few differences need to be considered. In order to be able to cut 1.5" wafers, large raw quartz bars are needed. Material, which develops channels or pits during etching, will result in significantly lower yields. The quartz quality ("Q") as measured by infrared absorption does not correlate with material related yield losses. Etch defect density has to be measured on the surface of polished and etched wafers. Therefore, trial runs through the actual process are the best method for the definition or selection of the raw material. At the presence of microcracks in the material or chipped edges, the large, thin wafers used for the technology break easily. In order to avoid these defects, lumbered bars need to be processed very carefully; all edges and corners of the lumbered bars should be rounded. Sawing and lapping operations are similar to the processing of conventional blank manufacturing, although the size of the wafers (1.5" x 1.5") requires more precision and caution. Double side lapping is inconvenient for final wafer-thicknesses below 100 µm, due to the mechanical forces applied to the wafer edges. After mechanical lapping and polishing, the wafers are usually chemically polished, in order to eliminate any microcracks or surface defects.

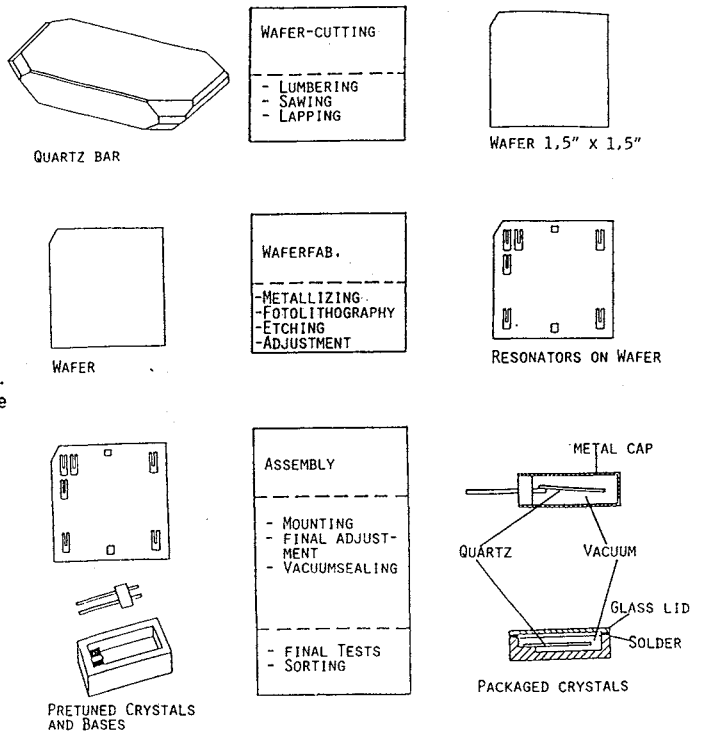


Fig. 1
Manufacturing Process (schematic)

2.2. Wafer Fab

The central part of a chemically milled resonator production line is the so called "wafer-fab". Approximately 30 to 50 individual process steps are required to produce functional resonators. This operation includes all metallization, etching and photolithographical processes. The most important steps are shown in Fig.2:
After cleaning, the quartz wafers are metallized (Cr and Au)in a high vacuum evaporation system. A uniform layer of photoresist is deposited on both wafer surfaces by a spinning process. Then, the shape of the resonator is engraved photolithographically into the metal layer. This process involves UV-exposure on a double-sided mask aligner, developing and metal etching. All these are well-known operations in the integrated circuit manufacturing. However, two basic differences need to be underlined: in quartz manufacturing, not one, but both substrate surfaces need to be treated simultaneously, and the quartz wafers are rectangular and very thin, compared to the round IC-wafers.
In a second turn, the electrodes' geometry is defined in another photoresist layer, but not yet etched into the metal plating. Subsequently, the crucial operation is the chemical milling of the quartz substrate in a bath of ammonium fluoride and hydrofluoric acid at 85°C. The shape of the crystal resonator can be controlled within tight tolerances, the etching of single crystal quartz material being very anisotropic (Fig.3) [13, 14]. The critical dimensions, e.g. the tuning fork tine width, can be kept within a tolerance of +/- 1 µm.

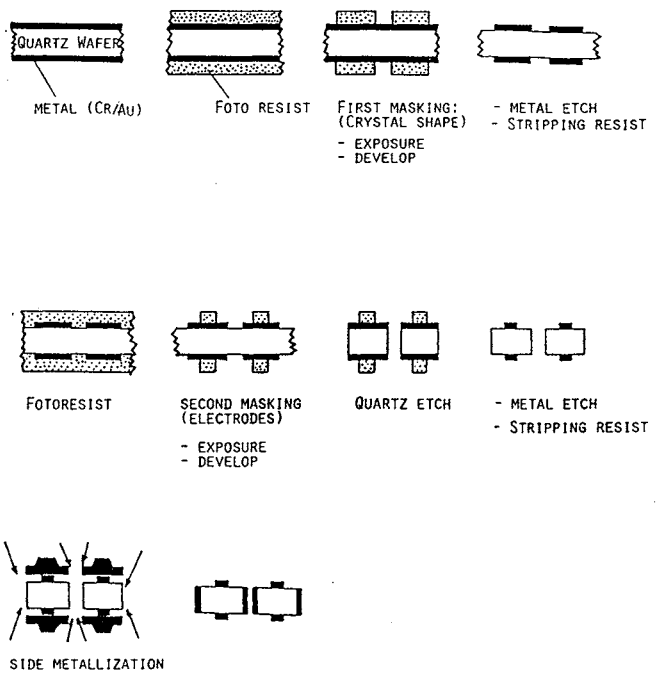


Fig. 2
Photolithographical Process
(cross-section: Tuning fork tines)

This tolerance value corresponds, for a typical tine width of 230 µm, to a frequency distribution of +/- 4,000 ppm in one production lot. After the chemical milling of the resonator shape, side electrodes and electrical connections between major faces are deposited by high vacuum evaporation through mechanical quartz masks with appropriate apertures.

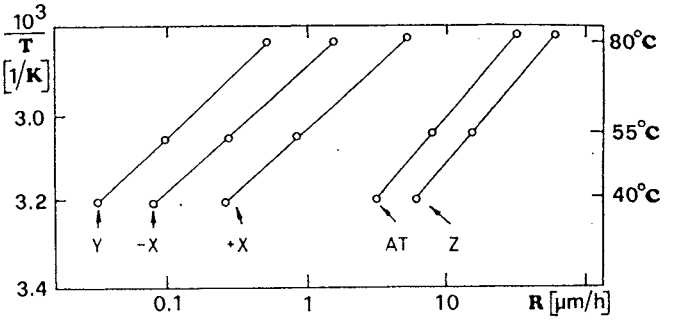


Fig.3
Etching rates in Quartz for different crystallographic directions.

2.3. Tuning

Still on wafer, a first frequency adjustment of the individual crystals takes place. The adjustment is made with a YAG-laser, evaporating small amounts of metal from the tuning areas, located at the end of the vibrating tines. Laser trimming is actually used only for low frequency crystals like tuning forks, length extension mode resonators and ZT-cut crystals. Thickness shear AT-cut crystals are adjusted by additional plating through mechanical masks. This first frequency adjustment operation on the wafer results in a frequency tolerance of +/- 100 ppm. Tighter tolerances require a final tuning after the mounting of the crystal onto the electrical feed-throughs of the package. For all the described photolithographic processes, a control of particle contamination is high priority. Therefore, this production line is located in clean rooms. The typical requirements for the atmosphere around a working table are:

- particle concentration: class 100 or better
- temperature: 25°C +/-1°C
- humidity: 40% +/- 5%
- illumination: yellow (no UV)

2.4 Packaging

All chemically milled crystals are mounted in specially developed miniature packages. Typical packages are multilayer ceramic chip carriers with glass lids, and small tubular metal cans.

The assembly operations include the following process steps:

- Mounting of the resonator into the base with conductive epoxy
- Final frequency adjustment by laser trimming or additional plating
- degasing in high vacuum
- sealing under high vacuum with tin or tin-gold solder

3. Characteristics of chemically milled quartz resonators

3.1 Fundamental tuning forks

Tuning fork crystals vibrating in a fundamental flexural mode in the xy-plane can be made for frequencies between 10 and 200 kHz. The frequency of a clamped-free quartz beam is given by (I), where $s'_{22} = 1.27\ 10^{-11}\ m^2\ N^{-1}$ is the elastic compliance of the quartz along the beam or tuning fork axis, ϱ is the density of Quartz and α_n is the modal constant of the n-th flexural mode. ($\alpha_0 = 1.875$ fundamental mode, $\alpha_1 = 4.694$ 1st overtone mode). Typically achievable quartz parameters for the fundamental tuning fork type are given in Fig. 4: The inversion temperature T_0 depends on cutting angle and can be varied between -10 and +40°C.

$$(I)\quad F_n = \frac{\alpha_n^2}{2\cdot\pi}\cdot\frac{1}{\sqrt{12\cdot\varrho\cdot s'_{22}}}\cdot\frac{w}{l^2}$$

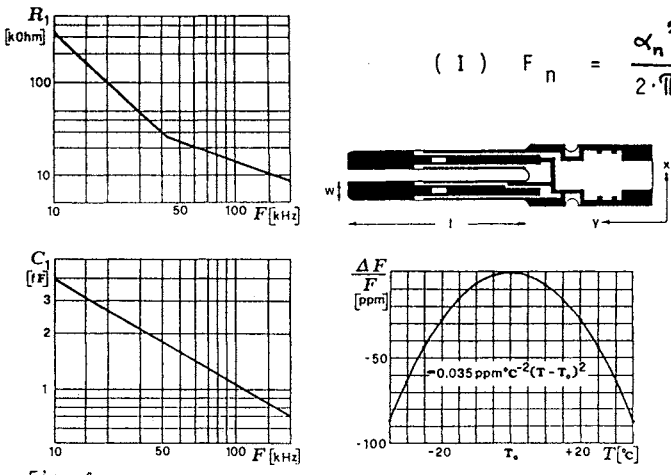
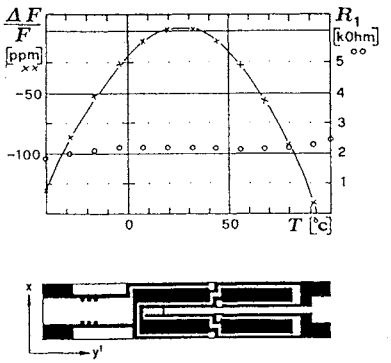


Fig. 4
Parameters for fundamental tuning fork

3.2 Overtone Tuning Forks

In a frequency range of 200 kHz up to 520 kHz, the first overtone flexural mode tuning fork is used. For the same tine geometry, the first overtone frequency is approximately 6.27 times higher than the fundamental mode. Typical parameters of an overtone tuning fork are given in Fig. 5. The overtone tuning fork crystal has the same temperature coefficient as the fundamental tuning fork. The two notches at the extremity of the tines are optimized for minimal energy loss in the mounting structure.

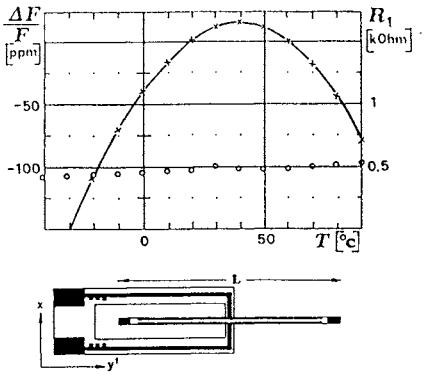


F_S = 240.2 kHz
R_1 = 2.1 kOhm
C_1 = 2.3 fF
C_o = 1.6 pF
Q = 130 000
P = 0.1 uW

Fig. 5
Typical Parameters of a 240.2 kHz overtone tuning fork

3.3 Length Extensional Resonators

Using the same quartz cut (Z-cut) as for tuning forks, higher frequencies can be obtained with length extensional resonators [2, 3]. With either fundamental or overtone vibrations, the frequency range between 500 kHz and 2 MHz can be covered. The monolithic structure consisting of the resonator itself and the integrated mounting structure, is a typical example for photolithographically processed quartz crystals. No other technology has the advantage to reproduce as easily such complex structures. The frequency-temperature behaviour is the same as for tuning forks. Typical parameters for a 1 MHz crystal are given in Fig.6.



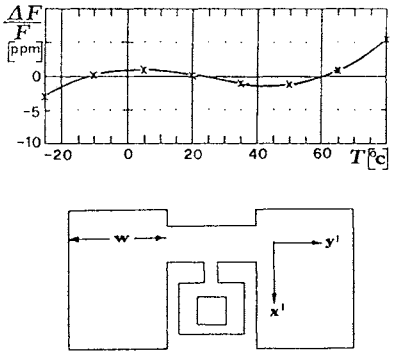
F = 2.75 MHz mm $\frac{1}{L}$

F_S = 1.0 MHz
R_1 = 500 Ohm
C_1 = 1.4 fF
C_o = 1.1 pF
Q = 200 000
P = 0.1 uW

Fig. 6
Typical Parameters of a 1 MHz length extensional crystal

3.4 ZT-Resonators

The ZT-cut resonator [5], Fig. 7, vibrates in a width-extensional mode, whose frequency is given by (II). ZT-cut resonators can cover a frequency range from 2.1 MHz to 4 MHz. The main advantages of this crystal are its excellent thermal behaviour, low g sensitivity and low aging. Since no side electrodes are needed, the photolithographic operations are simplified as compared to tuning forks. Frequency adjustment can be carried out with tight tolerances by laser tuning on wafer [15]. Since each frequency requires a thorough individual design with entirely specific geometry (i.e. a whole set of masks), the development of any frequency in the mentioned range is rather expensive.



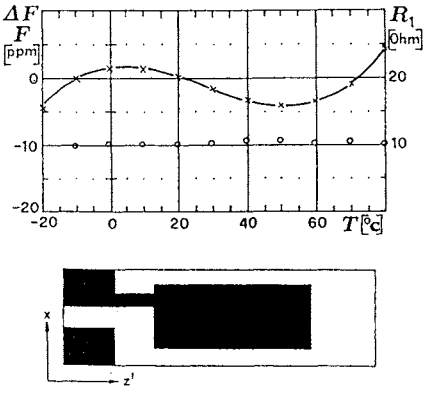
(II) $F = \frac{2820 \text{ kHz mm}}{w}$

F_S = 2.56 MHz
R_1 = 250 Ohm
C_1 = 0.82 fF
C_o = 2.3 pF
Q = 300 000
P = 5 uW

Fig. 7
Typical Parameters of a 2.56 MHz ZT crystal

3.5 AT-strip resonator

Several types of miniaturized AT-strip resonators have been developed and produced by conventional means. Most of these have convex surfaces, bevelled edges or tilted faces. The main surfaces of a chemically milled AT-crystal are usually flat. For each individual frequency, a specific design occurs. It is necessary to take into account the coupling of thickness-shear and XY-flexural modes. In the last four years, chemically milled fundamental AT-crystals with the longitudinal axis parallel to the Z'-direction have been developed. These crystals are now batch-processed on 1.5" sq wafers and are available in the 8 MHz to 30 MHz range at Micro Crystal. Typically achievable parameters are shown in Fig. 8.



F_S = 16.0 MHz
R_1 = 10 Ohm
C_1 = 8.5 fF
C_o = 3 pF
Q = 120 000
P = 1 mW

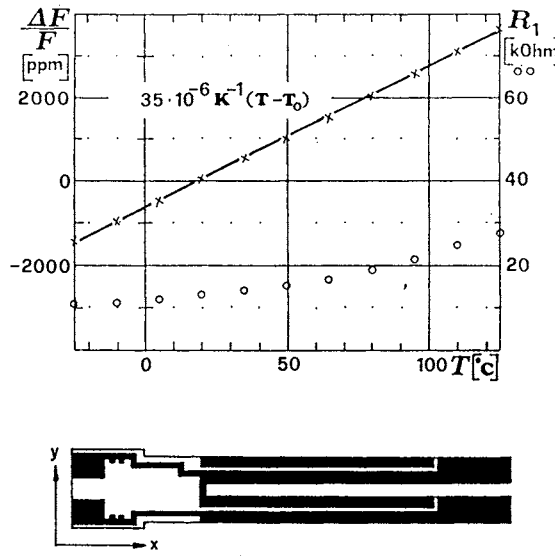
Fig. 8
Typical Parameters of a 16 MHz Strip AT crystal

4. Resonating Quartz Transducers

Today, an increased demand for miniaturized low cost electronic sensors, transducers and actuators has developed. Capabilities of quartz transducers, whose frequencies depend on physical parameters like temperature, force or acceleration have widely been demonstrated [6, 7, 8, 9, 10, 11].Basically, quartz is a material with remarkable mechanical (elastic) and thermal properties. Furthermore, the quasi-digital signal of a vibrating sensor is of inherent interest in a number of applications. Photolithographic technology and its capability to generate quartz resonators with almost any imaginable geometry, offers the possibility to produce low cost resonating quartz sensors.
In this paper, only two representative examples are described:

4.1 Temperature Sensors

A torsional tuning fork quartz resonator used as a temperature sensor has been available on the market for several years. The crystal has a nominal frequency of 262.144 kHz and its frequency varies almost linearily over more than 100 degrees (Fig. 9).
The small size allows it to be encapsulated either in a metal can with a 1.5mm diameter or in a 7 x 2.7 x 1.6mm ceramic carrier.



F_s = 262.144 kHz
R_1 = 15 kOhm
C_1 = 0.4 fF
C_o = 1.5 pF
Q = 100 000
P = 1 uW

Fig. 9
Typical Parameters of a temperature sensor tuning fork

4.2. Force Sensors (DETF)

Double ended tuning fork (DETF) crystals change their frequency, when a force is applied along the axis of the beams. The relative frequency change produced by a force K can be calculated with a modified Raleigh method [10], taking into account the flexural potential energy, the axial force potential energy and the kinetic energy. An approximation to the first order yields in (III)

$$\frac{\Delta F}{F} = 0.148 \frac{1}{E}\left(\frac{l}{w}\right)^2 \frac{K}{w \cdot t} \qquad (III)$$

where E [N/m2] is the longitudinal elastic modus of the beam
l [m] is the length of the beam
w [m] is the width in the plane of flexural vibration
t [m] is the thickness
K [N] is the introduced force.
(without shunting arms)

The change in frequency may, however, be quite large (up to 10%).

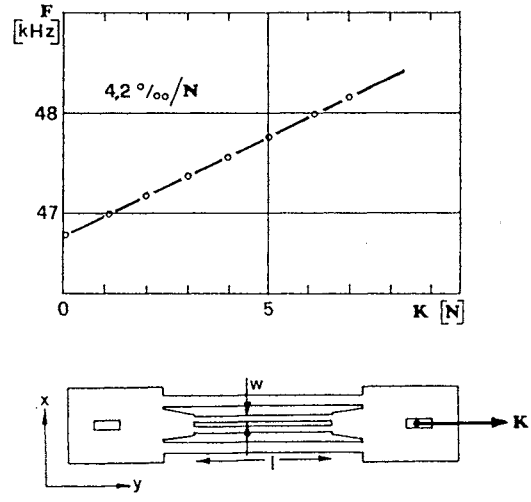


Fig.10
Force sensitivity of a double ended tuning fork crystal

Conclusions

The above described products represent only a choice of a variety of photolithographically produced quartz crystals. Other types are already available at other manufacturers or under development.
The limits of this technology have not yet been reached.In the future, and especially for the production of miniature crystals, the photolithographic process will certainly be adopted more and more. This will lead to an extension of the frequency range and to new applications, especially in the sensor field. A further field of application of the photolitho-etching technology of quartz material is the production of micromechanical components such as mirror galvanometers, etc., which have not yet been introduced.

References

[1] J.H.Staudte, Proceed. 27th ASFC (1973), 50-54, "Subminiature tuning fork resonator"
[2] R.J.Dinger, Proceed. 35th ASFC (1981), 144-148, "A miniature quartz resonator vibrating at 1 MHz"
[3] R.J.Dinger and J.G.Michel, US patent 4 608 510, "Piezoelectric micro-resonator"
[4] J.Hermann, C.Bougeois, Proceed. 33rd ASFC (1979), 255-262, "A new quartz crystal cut for contour mode resonators"
[5] J.Hermann, Procced. 39th ASFC (1985), 375-380, "A novel miniature ZT-cut resonator"
[6] R.J.Dinger, Proceed. 36th ASFC (1982), 265-269,"The torsional tuning fork as a temperture sensor"
[7] R.J.Dinger, J.G.Michel and C.E.Leuenberger, US patent 4 437 773, " Quartz Thermometer" .
[8] R.J.Dinger, C.E.Leuenberger and J.F.Chapuis, (Sensor 1983, Basel), "A cost-effective quartz temperature sensor and its application to temperature measurement devices"
[9] E.P.EerNisse, US patent 4 215 570, "Miniature quartz resonator force transducer"
[10] W.Albert, Proceed. 28th Int'l Instrum.Symp.,Las Vegas, (1982), 33-44, "Vibrating quartz crystal beam accelerometer"
[11] S.S.Chuang, Proceed. 37th ASFC (1983), 248-254, "Force sensor using DETF quartz crystals"
[12] H.F.Erdley, US patent 3 238 789, "Vibrating bar transducer"
[13] P.Suda, A.E.Zumsteg and W.Zingg, Proceed. 33rd ASFC (1979), 359-363, "Anisotrophy of etching rate for quartz in ammonium bifluoride"
[14] J.R.Vig, Proceed. 4th ASFC (1982), 47, "Etching quartz crystals"
[15] W.Zingg, B.Studer, J.Hermann, 1st European Time and, Frequency Forum (1987), 31-35, "Fabrication de résonateurs ZT"
[16] US patent 4 313 071
[17] US patent 4 450 378

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-05 Number 2

X-RAY DIFFRACTION STUDY OF QUARTZ CRYSTALS WITH INSITU ELECTRIC FIELD

M. T. Sebastian and H. Klapper
Institut für Kristallographie, RWTH, 5100 Aachen,
Federal Republic of Germany.

Abstract

X-ray diffraction study of alpha quartz crystals with insitu electric field along different crystallographic directions showed a strong enhancement in the diffracted intensity. The effect is prominant in natural crystals when the electric field is along the c-axis. This is evidenced by the X-ray rocking curves recorded from several crystals. The X-ray topographs recorded with insitu electric field showed extinction contrast consisting of striations parallel to the c-axis. This contrast disappears during irradiation in the absence of electric field. The increase in the diffracted intensity and the contrast on the topographs are explained by a model on negative space charge collection.

Introduction

Among the oxides, quartz is a unique material. It is extremely stable, and found in nature as large crystals or grown in the laboratory with even higher quality. Its most prominant property is piezoelectricity in combination with excellent mechanical behaviour. This property allows quartz to be used in a variety of electronic devices, including crystal oscillators and filters for precision frequency control, sensors of temperature, pressure and mass, discriminitors and filters in instrumentation and telecommunication. It is essential that the frequency be accurately controlled over the entire radio frquency spectrum to avoid interference between transmitters operating in the same geographical area or on adjacent channels. The continuing demand for space in the available radio-frequency spectrum has resulted in closer spacing of the assigned frequencies with accompanying requirements for better frequency control. Maximum utilisation of the the limited frequency spectrum requires the highest possible frquency stability of both transmitters and receivers. In most of these applications imperfections play an important role, and quartz crystals of very high quality are often required. For example, the quality factor Q of a quartz resonator directly depends on the point defects present in the crystal. X-ray diffraction Lang topography is a useful non-destructive technique for investigating imperfections. Several studies on quartz have been reported (see for example ref. /1/).

Recently several authors /2-8/ observed an anomalous increase in the diffracted intensity and contrast on the topographs recorded from alpha quartz under the influence of electrostatic fields of about 10 to 300 KV/cm. In the present paper we report a detailed X-ray topographic investigation of the electric field related intensity increase and X-ray topographic contrast in α-quartz to study the influence of:

(a) the effect of electric-field direction
(b) the effect on the different hkil topographs

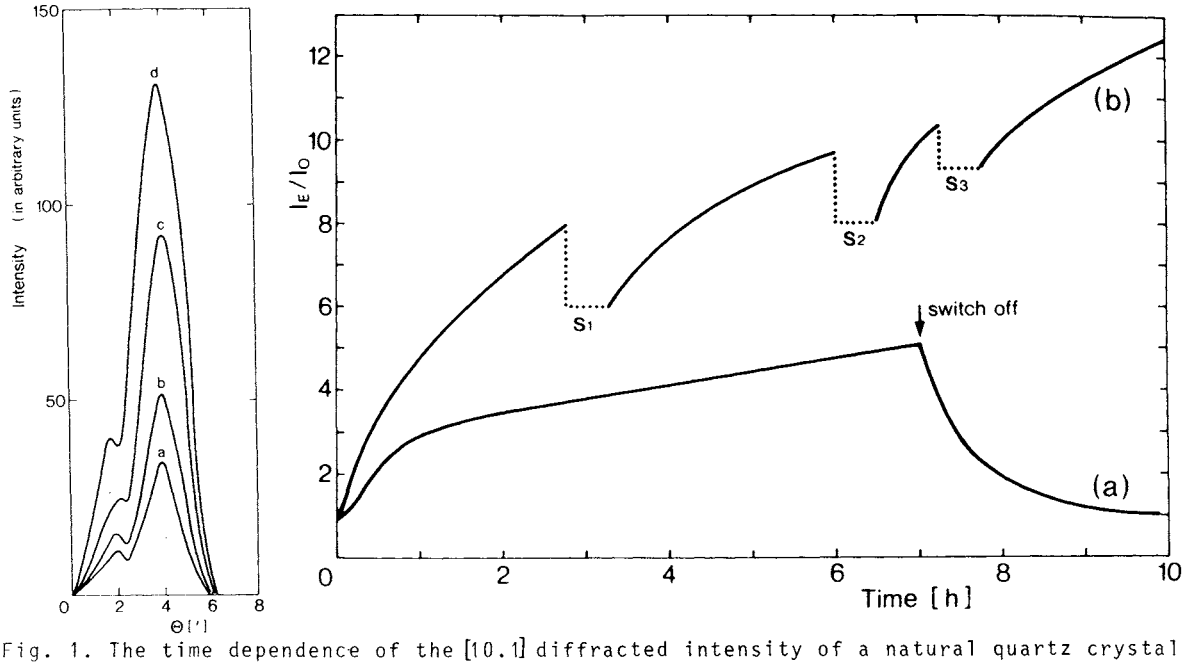


Fig. 1. The time dependence of the [10.1] diffracted intensity of a natural quartz crystal with 4 KV/cm along the 10.1 direction. The markings S_1,S_2 and S_3 represent the recording of three section topographs. The crystal was continously irradiated with X-rays to record curve (a) and was irradiated only during measurement for curve (b).

Fig. 2. 10.1 X-ray rocking curves of a Y-cut natural quartz crystal recorded with 4 KV/cm along the c-axis using MoK_κ radiation. (a) No E; (b) 2 KV/cm; (c) 4 KV/cm; (d) 6 KV/cm.

(c) the effect of irradiating the crystals

EXPERIMENTAL

In the present investigation quartz crystals, both synthetic and natural, of about 1mm thickness were used. A static electric field of 1 to 6 KV/cm was applied along different crystallographic directions using soft graphite electrodes(cut from a graphite felt obtained from Sigri GmbH, Meitingen, FRG). The use of soft graphite electrodes make very good electrical contact also with curved crystal edges. X-ray topographs (Lang technique) and rocking curves (obtained by the step wise rotation of the crystal on the Lang camera) were recorded using MoK_α radiation. The X-ray generator was operated at 40 KV and 20 MA. The exposure time for recording the topogrphs was about 2 hours.

RESULTS

X-ray rocking curves recorded from all crystals showed a strong enhancement in the diffracted intensity. The effect is more pronounced when the electric field is along the c-axis. On applying a dc field the diffracted intensity increases rapidly in the beginning and then slowly. When the field is switched off the intensity decreases rather quickly (Fig. 1 curve a). In curve (a) the application of the electric field and X-ray irradiation were simultaneous and the crystal was continously irradiated with the incident X-ray beam throughout the experiment. In Fig. 1 curve (b) the crystal was irradiated with the incident X-ray beam only at the time of recording the diffracted intensity. The intensity was recorded at intervals of 5 minutes in the first one

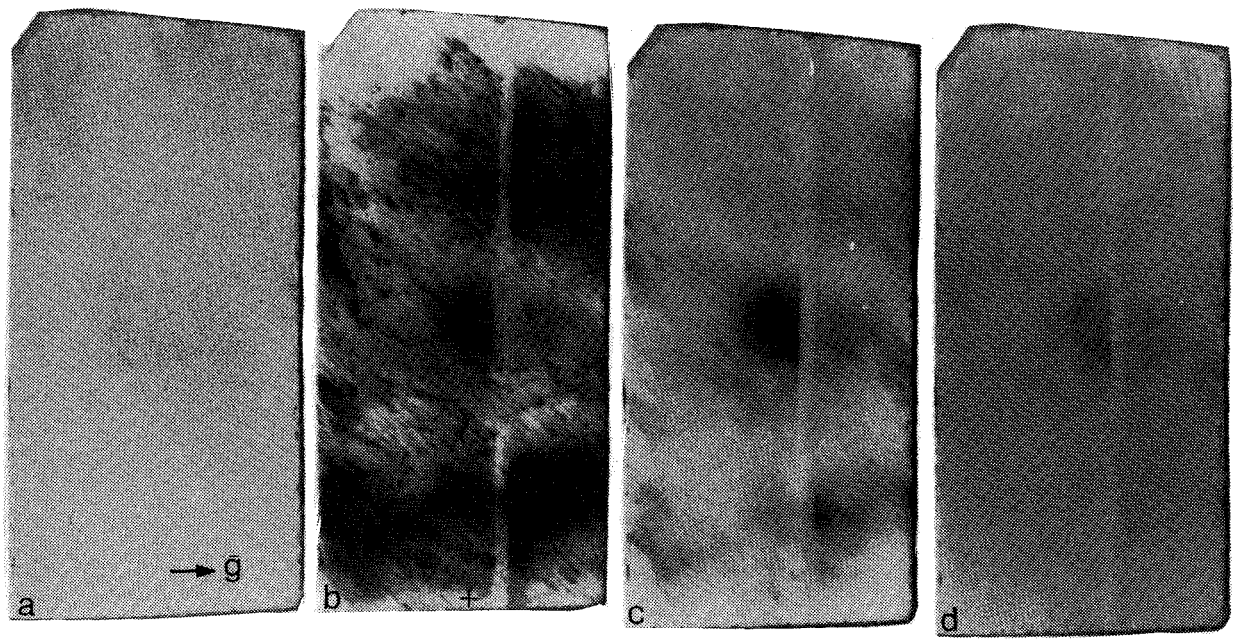


Fig. 3. 10.1 X-ray transmission topographs of a natural quartz crystal AT cut (17 X 9 X 1 mm) recorded with insitu electric field along the 10.1 direction. (a) No E (b) 2.5 KV/cm (c) no E, recorded immediately after (b); (d) No E; recorded immediately after (c). The crystal was irradiated with X-rays for several hours under 4 KV/CM after taking Fig. 3a for Fig. 1 and section topographs. The irradiated region is marked I.

hour and later 20 minutes. The marking S represents the time (30 minutes) during which the crystal was continously irradiated with the incident X-ray beam to record section topographs. A study of Fig. 1 curve (b) reveals that irradiation results in a decrease of diffracted intensity as evidenced by the lowering of the intensity after recording section topographs.

Fig. 2 shows a rocking curve taken from a natural crystal using the 10.1 reflection. The rocking curves in each case were recorded under electric field about one hour after applying the electric field. The peak intensity depends strongly on the applied electric field strength. The half-widths at full maximum (HWFM) of the peak intensity is found to increase with an increase in the applied voltage. There is no shift in the position of the peak maxima which is in contrast to the observation of Dousse and Kern /5/ using γ-rays.

The X-ray topographs recorded with insitu electric field showed a strong contrast (see Figs. 3 & 4). The topographs of some of the crystals contain dark striations parallel to c-axis (Figs. 3b, 4b). The contrast with striations is very strong near the anode side and appear at the opposite side on reversing the electric field (Fig. 4c). It is interesting to note that the striations are parallel to c even if the applied field is along the [10.1] direction (Fig. 3b). Both the h0.0 and 00.1 type topographs show the electric- field related intensity increase by extinction reduction and contrast with striations . The contrast gradually disappeared on recording the topographs successively in the absence of the electric field (Figs. 3, 4). Irradiation with X-rays using a wide slit and translating the crystals in the absence of an electric field for about 20 minutes completely removes the contrast. The bleaching of the contrast on removal of the field takes few hours to few days and can be expedited by intentional X-ray irradiation. Irradiation of stationary crystals with in-situ electric field for about 10

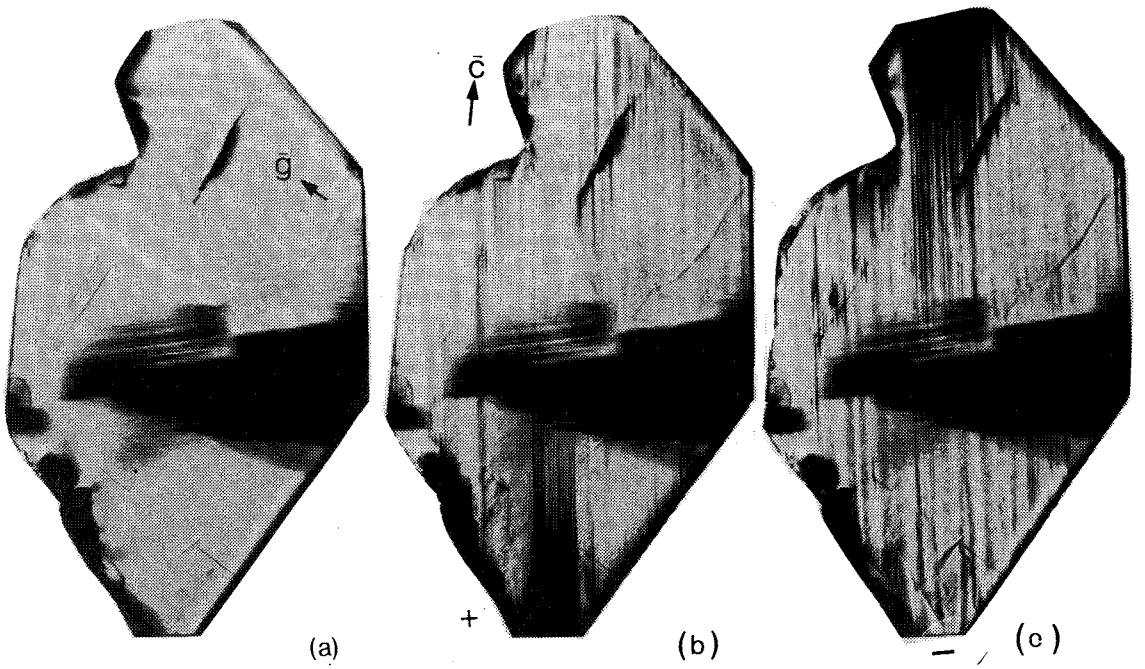


Fig. 4. 10.1 topographs of a Y-cut Herkimer diamond quartz crystal recorded with electric field along the c-axis. (a) No E (b) 5 KV/cm (c) - 5 KV/cm.

hours generated a smoky color at the irradiated region and a local decrease in the contrast on the topograph marked I in Fig. 3.

DISCUSSION

A study of the X-ray rocking curves recorded from several natural and synthetic quartz crystals with in-situ electric field show that there is considerable extinction reduction in agreement with earlier reports /2-8/. The enhancement in the diffracted intensity is directly related to the contrast on the topographs. This contrast is not due to the generation of dislocations as might be supposed.

The extinction reduction and the contrast on the topographs are attributed to the inhomogeneous lattice strain due to space charge collection in the crystals. Aluminium invariably exists as a substitute for silicon in both synthetic and natural crystals. The trivalent aluminium can make only three of the bonds normally made by a silicon and hence a negative charge exists on the site due to the fourth non-bonding oxygen. The charge compensation is carried out by by alkali ions (Li^+, K^+ or Na^+) and protons (H^+) or holes (e^+) trapped at oxygen ions/9,10/.

When a dc electric field is applied parallel to the c-axis, the weakly bound charge compensating monovalant impurity ions move towards the cathode through the open channels. If additional mobile positive ions are not available at the anode side the electrolysis will result in the collection of negative space charge at the 4 th non-bonding oxygen near the aluminium . This space charge polarisation deform the atomic planes inhomogeneously giving rise to an enhancement in the diffracted intensity. The

space charge collection is predominant at the anode side since the alkali ions are moving towards the cathode. This is in agreement with the experimental observation that a very dark contrast with striations is sprouting from the anode side and extending towards the cathode on the X-ray topographs (see fig. 4). The effect is present on all hkil topographs indicating that the lattice deformation is inhomogeneous and is not directional. In Fig. 3, even though the electric field was applied along the [10.1] direction, the striations appeared parallel to the c-axis indicating a lack of charge compensating ions along this direction i.e., the alkali ions moved along this axis.

It should be noted that the exposure time for the topographs recorded with insitu electric field is much less than that without electric field. The topographs recorded with the same exposure time as those without field were overexposed (very dark).

The electric field related contrast disappears on removal of the electric field. A few hours of irradiation of the crystals in the absence of the electric field completely destroy the contrast and restore the crystal to its original state in contrast. The disappearance of the contrast is due to the neutralization of the negative charges collected at the 4 th non-bonding oxygen near the substitutional aluminium during X-ray irradiation. It is well known /9-15/ that irradiation of quartz crystals produce Al-hole and Al-OH centres. Very recently electrolytic generation of Al-hole centres has also been reported /15/. During irradiation the fourth non-bonding oxygen may loose an electron from its outer P orbit producing a hole. Electron spin resonance (ESR) data of vacuum electrolysed crystals showed /13/ that the Al-hole concentration increases by a factor of four during irradiation.

Recently Sebastian et al./20/ observed a very strong contrast with striations parallel to the c-axis near the anode side of quartz crystal plates which were subjected to long time (about 10 days) solid state electroysis in vacuum at elevated temperatures (vacuum swept crystals) and the contrast disappeared during irradiation.

Berlinite ($AlPO_4$) which is homotypic with quartz having c-axis channels of nearly the same size does not show intensity increase or topographic contrast on any of the reflections under the influence of electric field. This is due to the fact that in berlinite there is no element (ion) with a difference in charge substituting for Al or P with associated charge compensators like Al-alkali centres in quartz.

Keerti and Lang /22/ observed fine lines parallel to [0001] on X-ray topographs in weakly electrolysed regions of a quartz plate. These lines were visible in reflections from planes of the major and minor rhombohedral forms and were invisible in the topographs obtained with reflections of the type 10.0. This indicates that the lattice displacements which produce the diffraction contrast are predominantly parallel to [00.1] These lattice distortions are attributed to the migration of silver ions. The silver formations build up inwards from the cathode surface and grow right across to the anode surface. In the present study the dominant contrast with fine lines appeared at the anode side of the crystal. All types of X-ray reflections (both h0.0 and 00.1) showed the contrast. Further the contrast slowly disappears on removal of the field, indicating that the contrast is not due to the migration of the electrode material. The fact that berlinite does not show any electric field related contrast, also indicates that the contrast is not due to the migration of the electrode material.

Calamiotou et al./7/ with BT cut oscillator crystals using silver electrodes found that the Bragg reflected intensity for the face under the positive potential along the [20.3]

is much higher than that for the face under negative potential and no reason was given for this difference. The present study shows that this is due to space charge polarisation and hence the lattice strain is predominant at the anode region.

It should be noted that Dousse and Kern /5/,Calamiotou et al./7/ and Kato et al. /3,4,6/ used thin AT or BT cut crystal plates and oscillators with vapour deposited metal electrodes and the electric field 10 to 300 KV/cm was applied along the thickness of the crystals. In the present study we applied the field (1 to 6 KV/cm) along the length and breadth of the crystals(15 to 30 mm), and so the electric field distribution may not be homogeneous. Hence our results as such cannot be directly compared with that of theirs.

Similar X-ray extinction reduction and contrast on the topographs have been observed in lithium iodate /23-25/, lithium hydrazinium sulphate and potassium titanyl phosphate /26/. All these crystals contain channels running parallel to the polar axis. The application of a static field leads to one-dimensional conduction by Lithium ions, protons and potassium ions respectively. In all these crystals the effect is present only when electric field is along the polar axis and visible on reflections normal to the polar axis. This indicates that the lattice deformation producing the effect is normal to the polar axis.

Acknowlegement

one of the authors (MTS) is grateful to the Alexander Von Humboldt Foundation for financial support, Prof. Th. Hahn for critically reading through the manuscript, R. A. Becker for technical assistance and Mrs. Delmhorst reproducing the photographs.

References

1. A. R. Lang. Diffraction and imaging techniques in materials Science, Eds. S. S. Amelinckx, R. Gevers and J. Van Landyut. North Holland Publishing Co. (1978).

2. E. Chrpa, H. Ihringer, H. Jagodzinski and A. Kneifel. Z. Naturf. a27(1972) 469.

3. S. Yamashita and N. Kato. J. Appl. Cryst. 8(1975)623.

4. K. Yasuda and N. Kato. J. Appl. Cryst. 11(1978)705.

5. J. Cl. Dousse and J. Kern. Acta Cryst. A36(1980)966.

6. K. Gouhara, Y. Bessho, K. Yasuda and N. Kato. Jap. J. Appl. Phys. 21(1982)L503.

7. M. Calamiotou, E. Anastassakis, V. Psicharis and S. Filippakis. J. Phys. D. C20(1987)5641.

8. J. Cl. Dousse and J. Kern. Phys. Lett. 59A(1976)159.

9. A. Kats. Philips Res. Repts. 17(1962)133.

10. J. J. Martin. IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control 35(1988)288.

11. J. A. Weil. Radiation Effects 26(1975)261.

12. S. P. Doherty, J. J. Martin, A. F. Armington and R. N. Brown. J. Appl. Phys. 51(1980)4164

13. A. Kahan, F. K. Euler, H. G. Lipson, C. Y. Chen and L. E. Halliburton. Proc. AFCS (1987)216.

14. M. A. Mondragon, C. Y. Chen and L. E. Halliburton. J. Appl. Phys. 63(1988)4937.

15. G. Lehman et al. Z. Naturf. 44(1989)278.

16. K. Nassau and B. E. Prescott. Miner. Mag. 41(1977)300.

17. A. J. Cohen and L. N. Makar. Phys. Stat. Sol. a73(1982)593.

18. O. F. Shirmer. Sol. St. Commun. 18(1976)1349.

19. R. H. D. Nuttal and J. A. Weil. Sol. St. Commun. 35(1980)789.

20. M. T. Sebastian, A. Zarka and B. Capelle. J. Appl. Cryst. 21(1988)326.

21. R. H. D. Nuttal and J. A. Weil. Sol. St. Commun. 33(1980)99.

22. S. Keerti and A. R. Lang. J. Appl. Cryst. 5(1972)72.

23. M. T. Sebastian, H. Klapper and S. Haussühl. Proc. Intl. Conf. on Materials for Non-linear and Electro-optics, Cambridge (1989). IOP publishing House, pp 53.

24. Y. Li. In: Advances in the Science of China. Phys. Vol.1. Beijing Science Press (1985).

25. J. Bouillot, J. Baruchal, M. Remoissenet, J. Joffrin and J. Lajzerowicz. J. Physique 43(1983)1259.

26. M. T. Sebastian et al. (to be published).

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-05 Number 3

SPHERICAL TRIGONOMETRY FOR QUARTZ BLANK FAST ROUTINE X-RAY CONTROL

H. MERIGOUX, J.F. DARCES, D. AUGUIE

Laboratoire de Cristallographie
Université de Franche-Comté
25030 BESANCON Cedex - France

ABSTRACT

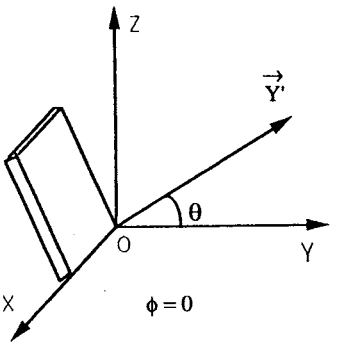
In this paper several methods are presented that give the value of the orientation of a blank cut in a quartz crystal. A fast method using spherical trigonometry calculations is introduced. Results given by this method are compared to those obtained by more sophisticated methods. Examples are given.

INTRODUCTION

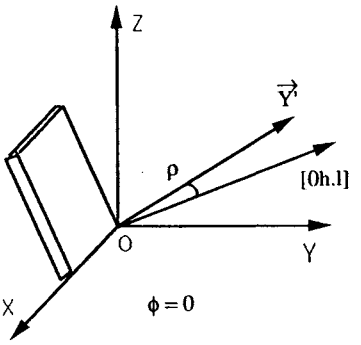
We must start saying that the determination of the orientation of a blank cut in a crystal is not a very easy task.

By definition a goniometer measures the angle between two directions. It works in a 2D space but crystals or the blanks we study are in a 3D space. So, from observations made in a 2D space, we have to rebuilt a 3D space. This needs generally several observations.

Fortunately a cut defined by only one rotation, as for instance an AT cut ($\Theta = \Theta_o$, $\Phi = 0$), may be controlled, on a classical goniometer, with only one observation. The chance we have with that kind of cut is that we can reduce the problem to a 2D space observation. In the quartz crystal space there are several lattice planes with their perpendicular on the same geometrical plane used for the definition of the cut.



The perpendicular to the cut is in the plane defined by the $\vec{Y}$ and $\vec{Z}$ axes. In this plane we have several(0k.l) lattice planes useful for an easy angle determination.



With doubly rotated cuts there are no simplifications. We must operate directly in a 3D space and new approaches giving an answer in a 3D space must be found.

THEORETICAL APPROACH

We suppose we are able to determine with a goniometer angles between $\vec{Ng}$ the perpendicular to the blank and the perpendicular to some lattice planes $(hk.l)_1$, $(hk.l)_2$, $(hk.l)_3$,... We name these angles ρ_1, ρ_2, ρ_3,... The positions of these lattice planes are known in the crystal space. They are expressed with the Θ and Φ angles recommended by the IEEE convention.

According to the number of different measured ρ angles a special method may be chosen.

With only two ρ angles spherical trigonometry is used.

With three ρ angles matrix calculation is needed.

With more than three ρ angles a statistic method giving in addition the accuracy on the results may be introduced.

MATRIX CALCULATION

An unambiguous result is obtained if three angles are measured. In that case an experimental axis system may be constructed. It is possible from it to obtain with matrix calculation the expected orientation expressed by the classical Θ and Φ angles.

It can be done because each of the three ρ measured angles is associated with one $[hk.l]$ direction in the crystal space. These three $[hk.l]$ directions may be considered as an experimental axis system. In this axis system the position of the $\vec{Ng}$ direction is perfectly known by the three measured ρ angles. Then using matrix calculation (appendix 1) (appendix 2) we pass from this special axis system to the classical $\vec{X}$ $\vec{Y}$ $\vec{Z}$ axis system.

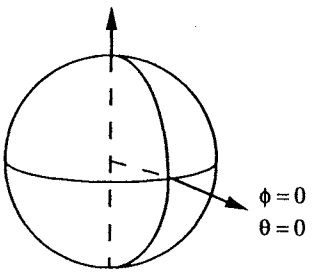
If more than three ρ angles are measured, it is possible to make statistics. The previous matrix calculation is done again taking all the ρ angles three by three. Then a standard deviation is associated to the calculated Θ and Φ angles.

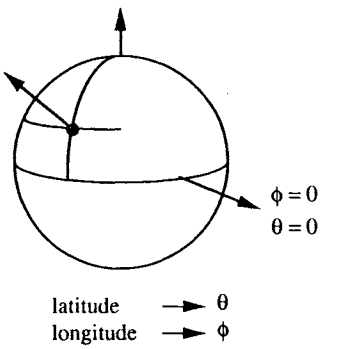
SPHERICAL TRIGONOMETRY

If only two ρ angles are measured it is still possible to obtain a result. But the mathematical solution of the problem is completely different. Several spherical triangles must be solved.

Instead of trigonometric formula it is more convenient to use figures showing how the calculation is made.

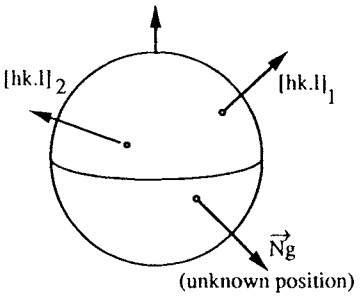
Supposed you are somewhere on a sphere. Your position is unknown. It must be expressed with longitude and latitude.

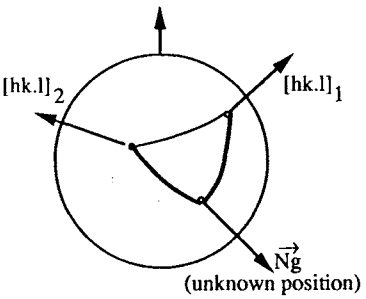




On this sphere, longitude and latitude correspond respectively to the Φ and Θ angles. The origin is the $\vec{Y}$ direction defined with $\Theta = 0$ and $\Phi = 0$.

The only two things you know are the positions of two special points on the sphere and your distance to them.

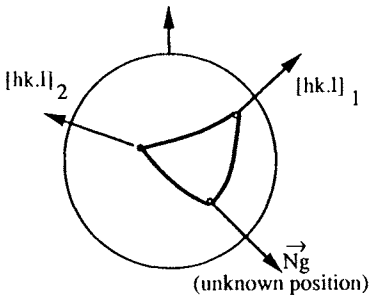




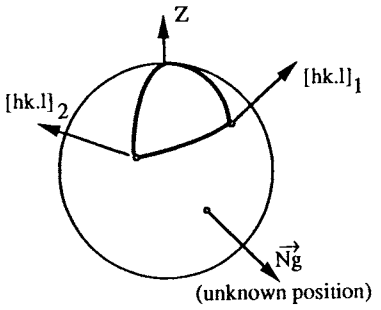
Your unknown position corresponds to the direction of the perpendicular to the Ng cut and the two special points are the two lattice planes $(hk.l)_1$ and $(hk.l)_2$. The distances are the measured ρ angles.

The unknown position is determined if several spherical triangles are solved.

The first one is built with the two $[hk.l]$ directions and the $\vec{Ng}$ unknown position.



The first angle is easy to solve because its three arcs are known. Two arcs are the measured angles : $\vec{Ng}$, $[hk.l]_1$ and $\vec{Ng}$, $[hk.l]_2$. The third, the arc $[hk.l]_1$, $[hk.l]_2$ is calculated from the crystal parameters and the value of the h, k, l indexes.



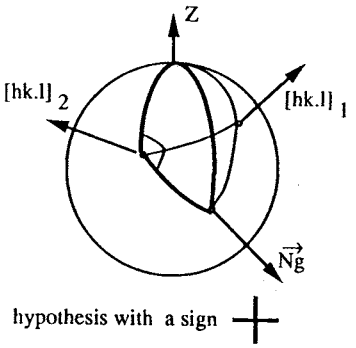
The second triangle is built with the same two $[hk.l]$ directions but now the $\vec{Z}$ direction replaces the $\vec{Ng}$ unknown position. The $\vec{Z}$ direction is the north pole on this imagined sphere.

All the arcs of this new triangle are known. Arcs Z, $[hk.l]_1$ and Z, $[hk.l]_2$ are the complementary value of the θ angle of the two observed planes.

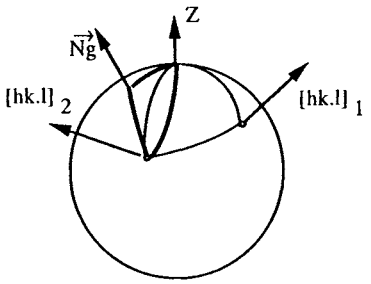
θ angle determination

When these two triangles are solved there is a last question. On a map of the world you must know if you are on the north or south hemisphere. The arc $[hk.l]_1$, $[hk.l]_2$ is equivalent to an equator and you have to know if you are above or under this line. As there is no mathematical answer, both hypotheses must be calculated, the first one above the line, the second one under the line.

With each of these hypotheses a special triangle is solved. Both triangles are constructed with the same summits : Z, Ng, $[hk.l]_2$.



hypothesis with a sign +



hypothesis with a sign —

But in the first triangle the angle $\vec{Z}$, $\vec{Ng}$, $[hk.l]_2$ is equal to the sum of the angles $\vec{Z}$, $[hk.l]_2$, $[hk.l]_1$ and $\vec{Ng}$, $[hk.l]_2$, $[hk.l]_1$.

In the second triangle that angle is equal to their difference.

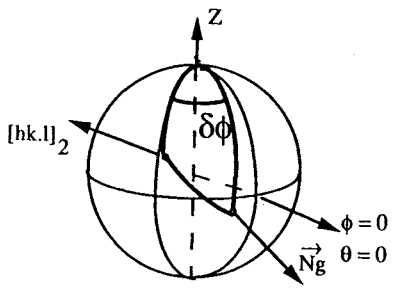
Two Θ angles are calculated. They are the complementary value of the arc Ng Z.

As it may be seen two different orientations are determined. Both of them are mathematically equivalent. A choice must be made either using a first matrix calculation or an approximate knowledge of the expected Θ value.

This method is particularly recommended during routine measurements.

Φ angle determination

The Φ angle is now easily calculated. In the triangles just solved, the angle $\vec{Ng}$, $\vec{Z}$, $[hk.l]_2$ is exactly the longitude difference between the measured direction and one of the well known positions in the crystal space : $[hk.l]_2$.



Remark

Methods we present are general and applied to any blank orientation of any crystal[1,2]. The goniometer must allow observations of any crystal lattice plane. The only theoretical limitation imposed by this refection method is :

ρ < Bragg angle.

RECOMMENDATIONS BEFORE A MEASURE

The choice of the Bragg angle is very important and will obey of two criteria.

First the Bragg angle must correspond to a low symmetry direction. If the symmetry of the direction is low, the number of equivalent lattice planes is greater. This increases the possibility to find them.

Second the Bragg angle must be of a great value. According to the previous inequality the highest value is always the best. A lattice plane is seen if its ρ angle is smaller than the chosen Bragg angle.

If these two criteria are fulfilled the probability to observer several lattice planes during a single run is the highest. That means several lattice planes may be observed on the goniometer without moving the crystal and the X-ray detector during a run[3].

Remark 1

It is possible to determine the sign of the θ angle of the blank if we choose a Bragg angle corresponding to enantiomorphous planes with very different scattering factors[2,4].

Remark 2

The sign of the ϕ angle may be obtained if we take into account the appearance order of the lattice planes during the observation on the goniometer[2,4].

RESULTS

In order to compare results given by this trigonometric method we have made a great number of observations on the same quartz blank.

The chosen lattice planes are : (33.0), (05.2), (32.4), (23.4), (12.4) and (42.0). We have studied 15 different lattice planes (table 1).

Indexes	$(\bar{2}6.0)$	$(\bar{4}6.0)$	(24.0)	$(\bar{6}4.0)$	$(\bar{5}3.\bar{4})$	$(\bar{5}2.\bar{4})$	$(\bar{2}5.4)$	$(\bar{2}3.\bar{4})$
Measured angles	32.15	35.125	57.015	56.95	43.383	56.68	69.77	18.21

Indexes	$(\bar{3}2.\bar{4})$	(05.2)	$(\bar{5}5.\bar{2})$	$(\bar{5}6.0)$	(33.0)	$(\bar{6}3.0)$	$(\bar{3}6.0)$
Measured angles	39.24	58.61	31.30	33.625	65.475	65.40	33.625

Table 1

The values of the indexes associated with the ρ measured angles are determined by a special trial and error method. A statistic is made on the different possible matrix calculations.

The obtained orientation is : $\theta = -33.61°$

$\phi = 30.03°$.

The standard deviations are respectively 0.02° and 0.03°.

Now spherical trigonometry is introduced. Values presented table 1 are introduced two by two in the programm. Some results are presented. (Table 2).

	θ	ϕ	or θ	ϕ
$(\bar{5}5.\bar{2})$ (05.2)	$\theta = -33.65°$	$\phi = 30.05°$	$\theta = +13.81°$	$\phi = 60.99°$
$(\bar{5}5.2)$ $(\bar{5}3.4)$	$\theta = -33.60°$	$\phi = 30.03°$	$\theta = +\ 7.36°$	$\phi = 79.27°$
$(\bar{2}3.\bar{4})$ $(\bar{3}2.\bar{4})$	$\theta = -33.64°$	$\phi = 30.01°$	$\theta = -55.88°$	$\phi = 12.11°$

Table 2

As expected two results are found. We must decide which one of the two results is the good one. The obtained orientation values are in the expected range.

During the computation the arc between the perpendicular two the introduced lattice planes is calculated. Generally the value of the crystal lattice parameters are used at this step of the calculation. If the introduced lattice planes have both indexes as (hk.0) the calculation of this arc is made without the knowledge of the lattice parameters. In that case the orientation is obtained without any introduction of a crystal description. With quartz crystals it may be very useful to be able to determine the orientation of a cut without any informations on the crystallographic parameters.

Some results are given (table 3).

	θ	ϕ	or θ	ϕ
$(\bar{2}6.0)$ $(\bar{4}6.0)$	$\theta = -33.61°$	$\phi = 30.04°$	$\theta = +33.61°$	$\phi = 30.04°$
$(\bar{2}6.0)$ (64.0)	$\theta = -33.62°$	$\phi = 30.02°$	$\theta = +33.62°$	$\phi = 30.02°$
$(\bar{4}6.0)$ $(\bar{6}4.0)$	$\theta = -33.60°$	$\phi = 30.01°$	$\theta = +33.60°$	$\phi = 30.01°$

Table 3

Two results are still found. They differ only with the sign of the Θ angle. We already must know this sign before the run. This simplification is found with crystals belonging to the trigonal, quadratic and hexagonal crystallographic systems.

This trigonometric method may also be used for single rotation cuts. The results for an expected orientation $\Theta = 36°$, $\Phi = 0°$ are presented for three different blanks, table 4. The observed lattice plane is (22.5). The two ρ angles measured are $\vec{Ng}$, $[4\bar{2}.5]$ and $\vec{Ng}$, $[\bar{4}2.5]$.

	θ	ϕ
$(4\bar{2}.5)$ $(\bar{4}2.5)$	$\theta = +35.96°$	$\phi = +0.005°$
$(4\bar{2}.5)$ $(\bar{4}2.5)$	$\theta = +36.015°$	$\phi = -0.005°$
$(4\bar{2}.5)$ $(\bar{4}2.5)$	$\theta = +35.95°$	$\phi = +0.02°$

Table 4

CONCLUSION

Introduction of spherical trigonometry in the orientation determination seems very useful with routine control. The number of the observations is reduced to the minimum. The measure of the position of only two reflection peaks is required.

With quartz crystals it is always possible to find a Bragg angle allowing the observation of two or more reflection peaks. Then it is possible to fix the position of the detector on the goniometer and to observe at least two lattice planes without changing the place of the blank during the measurements

If two lattice planes only are observed, spherical triangulation gives two calculated orientations. Both of them are mathematically equivalent and a choice must be made between them. If it is a routine control the expected values are known and the choice is easily made. However the accuracy is never exactly defined.

With this trigonometric method there is a new possibility. If the indexes of the observed lattice planes are (hk.0) the orientation is obtained without the knowledge of the lattice parameters.

In this paper only results about quartz crystals are presented, but the methods may be applied to all other single crystals.

ACKNOWLEDGEMENTS

We would like to express our thanks to the D.R.E.T. (Direction des Recherches Etudes et Techniques) of the Ministry of Defense who sponsored this research.

REFERENCES

1 - "Final X-ray control of the orientation of round of rectangular slides for industrial purposes", J.F. DARCES, H. MERIGOUX, Proceedings of the 32nd Annual Symposium on Frequency Control, Atlantic-City, U.S.A., 31 Mai - 2 Juin 1978, pp. 304-309.

2 - "Problèmes posés par la détermination de l'orientation absolue d'une lame monocristalline : application au quartz", J.F. DARCES, Thèse de Doctorat d'Etat es Sciences Physiques, Besançon, Septembre 1984.

3 - "A very fast orientation determination of doubly rotated quartz cuts", H. MERIGOUX, J.F. DARCES, Proceedings of the 11th Quartz Devices Conference and Exhibition, Kansas City, U.S.A., 29-31 Août 1989 - pp. 57-60.

4 - "Dissymmetry used for Θ, Φ angle sign determination of a piezoelectric crystal blank by a non destructive method", J.F. DARCES, J. LAMBOLEY, H. MERIGOUX, Ferroelectrics, 1982, vol. 40, pp. 245-248.

APPENDIX I

Determination of Θ and Φ angles of a blank when three lattice planes are observed on the goniometer

If three lattice planes $(hk.l)_1$, $(hk.l)_2$, $(hk.l)_3$ are observed on the goniometer, they are associated with the three angles ρ_1, ρ_2, ρ_3. The perpendicular directions to these observed lattice planes may be called $\vec{e}_1$ $\vec{e}_2$ $\vec{e}_3$.

These three directions form a new axis system and they are linked to the crystal reciprocal parameters $\vec{a}^*$, $\vec{b}^*$, $\vec{c}^*$ by :

$$\begin{bmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vec{e}_3 \end{bmatrix} = P \begin{bmatrix} \vec{a}^* \\ \vec{b}^* \\ \vec{c}^* \end{bmatrix} \qquad \text{with } P = \begin{bmatrix} h_1 & k_1 & l_1 \\ h_2 & k_2 & l_2 \\ h_3 & k_3 & l_3 \end{bmatrix}$$

or

$$\begin{bmatrix} \vec{a}^* \\ \vec{b}^* \\ \vec{c}^* \end{bmatrix} = P^{-1} \begin{bmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vec{e}_3 \end{bmatrix}$$

In the $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ axis system the position of $\vec{Ng}$ is defined by the ρ_1, ρ_2, ρ_3 angles.
Covariant coordinates of $\vec{Ng}$ in the $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ axis system are :

$$\begin{cases} y_1 = e_1 \cos\rho_1 \\ y_2 = e_2 \cos\rho_2 \\ y_3 = e_3 \cos\rho_3 \end{cases}$$

And their associated covariant coordinates in the $\vec{a}^*$, $\vec{b}^*$, $\vec{c}^*$ axis system are :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = P^{-1} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

If x_1, x_2, x_3 are covariant coordinates in the $\vec{a}^*$, $\vec{b}^*$, $\vec{c}^*$ axis system, they are contravariant coordinates in the $\vec{a}$, $\vec{b}$, $\vec{c}$ axis system.

From these numbers we obtain the three ρ_a, ρ_b, ρ_c angles between $\vec{Ng}$ and the three $\vec{a}$, $\vec{b}$, $\vec{c}$ vectors of the crystal cell.

$$\cos\rho_a = x_1/a$$
$$\cos\rho_b = x_2/b$$
$$\cos\rho_c = x_3/c$$

The blank orientation is now defined by three angles in the crystal space.

O angle of the blank is the complementary angle of ρ_c.

Φ angle is directly calculated from the angle between $\vec{Ng}'$, the projection on $\vec{Ng}$ on the XOY plane, and the $\vec{Y}$ direction.

APPENDIX 2

Determination of the actual value of the h, k, l indexes

The position of the X Ray detector on the goniometer commands the value of the Bragg angle and then the choice of the observed lattice plane. But with a given Bragg angle several lattice planes may be observed. Their number is imposed by the crystal symmetry. These planes differ with their indexes and their positions in space. For instance, with a copper X ray radiation the Bragg angle of 51.10° correspond to the (40.2) plane family. There are 12 equivalent planes such as : (40.2), (40.$\bar{2}$), (04.2), (04.$\bar{2}$),...
During a measurement the value of the h, k, l indexes of the observed planes is not known. A previous trial and error method gives their actual values before they are introduced in the matrix calculations.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-05 Number 4

RECENT RESULTS IN STUDY OF PIEZOELECTRIC RESONATORS BY X-RAY TOPOGRAPHY USING SYNCHROTRON RADIATION

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

Lab. de Minéralogie-Cristallographie, Universités P. & M. CURIE (Paris VI) et Paris VII, CNRS UA09, 4 place Jussieu, 75252 Paris Cedex 05, France

*Centre National d'Etudes des Télécommunications, PAB/BAG/MCT, 92220 Bagneux, France

Summary : Experimental settings of synchrotron X-ray topography are briefly recalled here and different aspects of vibrational states in resonators revealed by the present technique are described. Observations of vibration modes, especially on coupled piezo-electric components in quartz AT resonators, are reported. These observations revealed several complex mode coupling mechanisms.

Introduction

X-ray topography is an usual technique to assess the quality of crystals, namely to characterise vibration modes in quartz [1]. The use of synchrotron X-ray radiation from a storage ring has brought a number of new facilities to this technique in comparison with laboratory X-ray topography. The high flux of synchrotron X-ray sources and the large range of X-ray wavelengths permit to get simultaneously different diffraction patterns, thus to analyse the spatial structure of vibration modes. The pulsed synchrotron radiation allows the use of stroboscopic technique and the study of the time structure of vibration modes. Vibration modes in quartz have been analyzed with these advantages of synchrotron radiation at the D.C.I. storage ring (Orsay, France) and specific results, made available only by the present technique, have been obtained concerning coupled piezo-electric components.

I. Experimental techniques

Using synchrotron radiation several experimental settings, concerning either the resonator driving set-up or the topographic set-up, can be used :

Resonator driving set-up :

For classic (time-integrated) topography resonant vibrations are simply driven by a frequency generator and different resonances can be analyzed.

In the case of stroboscopic (time-resolved) topography only one resonant vibration whose frequency is n-times of that of pulsed X-ray beam can be analyzed. The driving electrical signal on resonators is directly piloted by the pulsed X-ray signal, so that the exact synchronization between X-ray beam and vibrations is assured [2].

Topographic set-up :

Laue patterns of resonators are obtained with white X-ray beam. When diffraction vectors **g** are suitably chosen, the spatial structure of the piezo-electric displacement **u** is visualized on different diffraction patterns according to the X-ray sensibility criterion **g.u**, where the extinction occurs for **g.u**=0.

Monochromatic X-ray beam can be extracted with a monochromator (for instance, with an asymmetric reflection on Ge(111)). In this case only one reflection occurs and the X-ray sensibility to the deformation field is higher than for Laue setting.

In this paper topographs shown are of Laue setting, excepted mentioned cases.

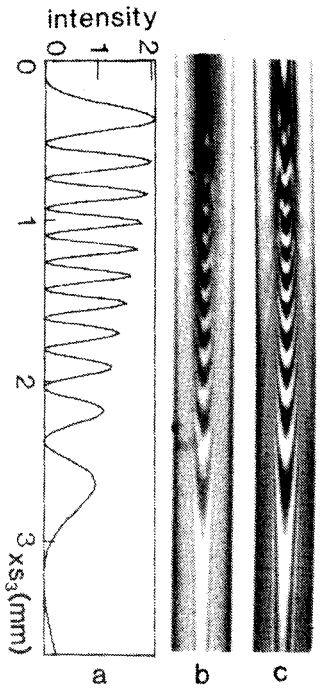


Fig.1. Experimental and simulated stroboscopic X-ray section topographs of a quartz AT resonator showing Pendellösung fringes with increasing vibration amplitudes. a) intensity profile according to Kato; b) experimental image; c) simulated image according to Takagi.

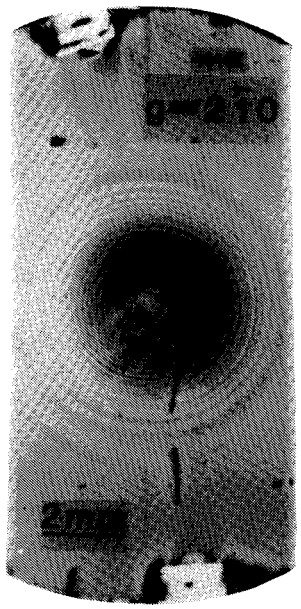


Fig.2. Stroboscopic X-ray traverse topograph of a quartz AT resonator showing elliptical fringes.

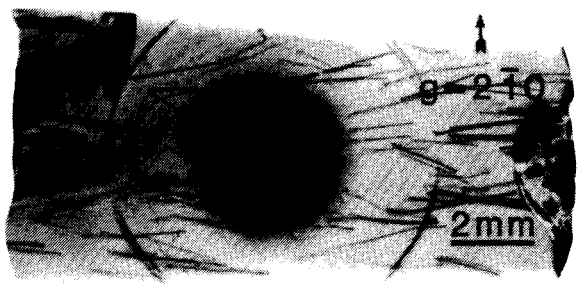


Fig.3. Stroboscopic X-ray traverse topograph of a quartz AT resonator containing dislocations.

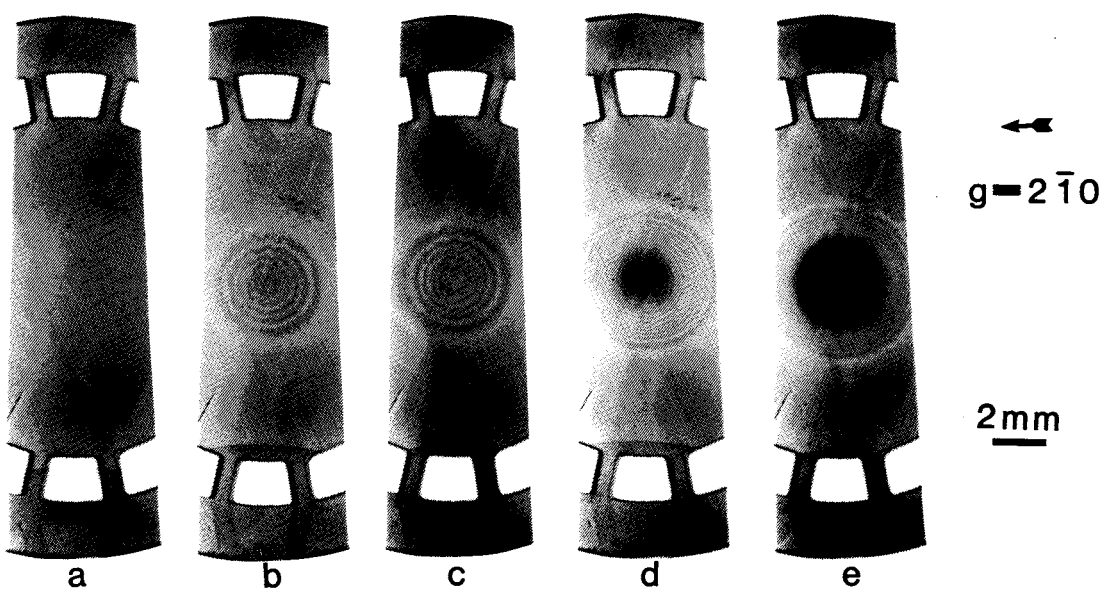


Fig.4. Stroboscopic X-ray traverse topographs of a quartz AT resonator containing a growth band. a) without vibration; b)-e) with increasing vibration amplitudes.

II. Vibrational aspects revealed by synchrotron X-ray topography

X-ray dynamic contrasts were evidenced by stroboscopic X-ray topography in connection to vibration amplitudes in resonators. Figures 1b et 2 are two examples of stroboscopic (respectively) traverse and section topographs in a high quality quartz AT resonator operating at fifth overtone thickness shear mode. On the section topograph *Pendellösung* fringes are visible with increasing vibration amplitudes and on the traverse one *elliptical* fringes, resulted from the integration of the previous *Pendellösung* fringes over the whole resonator surface, constitute equi-level curves of vibration amplitudes. These contrasts can be analyzed finely as shown by a numeric simulation according to Takagi's dynamic X-ray theory (Fig.1c) or by an intensity profile calculation according to Kato's geometrical X-ray theory (Fig.1a) [3; 4]. It was established that the fringe number is related to vibration amplitudes within a 0.3Å precision in usual cases [5; 6].

These X-ray contrasts characterise vibrational states in resonators and their modifications in presence of crystal defects provide indications on the defects influences on acoustic vibrations [7]. In figure 3 it is shown a stroboscopic X-ray traverse topograph of a quartz AT resonator containing dislocations. It can be noticed that *elliptical* fringes are mostly disturbed in the vicinity of dislocations and the form of the central vibrating zone which is seen by a black contrast is affected in the present case by dislocations. Figures 4a-4e show a series of stroboscopic X-ray traverse topographs of a growth band in function of vibration amplitudes. It can be clearly seen that *elliptical* fringes are also perturbed by the presence of the growth band and it suggests that the spatial coherence of acoustic waves is probably affected by growth bands.

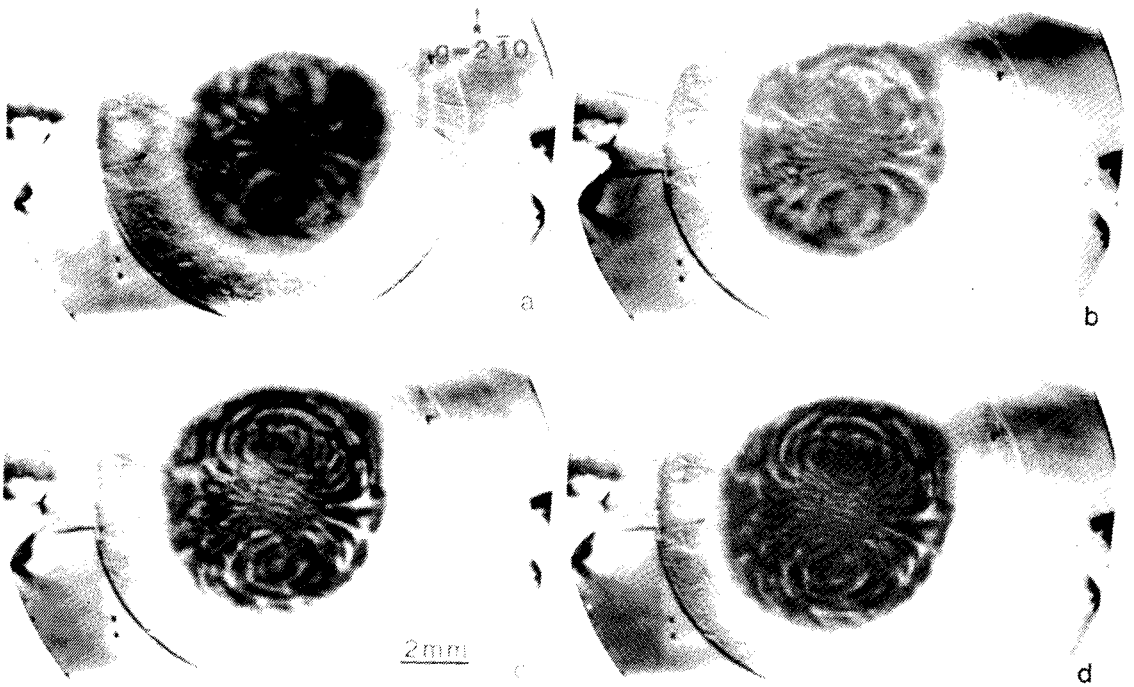


Fig.5. Monochromatic stroboscopic X-ray topographs of a highly excited quartz AT-cu plano-convex resonator imaged with different relative time phases between the X-ray beam and the vibration.

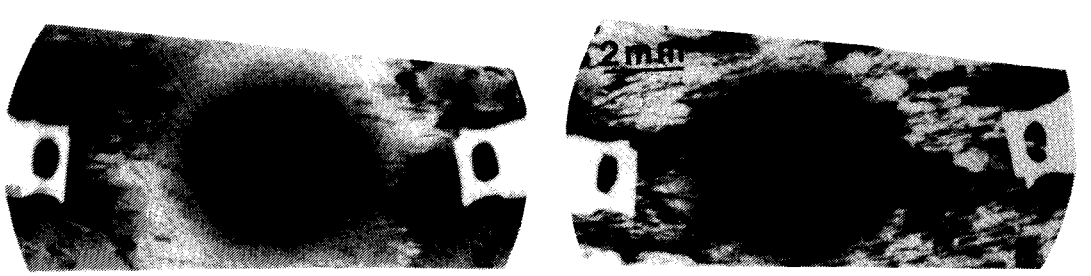


Fig.6. Classic X-ray topographs of two quartz AT resonators presenting different electric phase noises. a) sample of lower electrical phase noise; b) sample of higher electrical phase noise.

Observations by X-ray topography revealed several interesting aspects of resonators which need still further investigations in order to get a better understanding. In highly excited quartz AT resonators complex X-ray contrasts were observed by stroboscopic X-ray topography using the monochromatic setting. In figure 5a-5d four stroboscopic topographs are presented with increasing values of the relative phase between X-ray beam and vibrations. The symmetry of X-ray contrasts obtained in this case with respect to the crystallographic axis is to be noticed and these contrasts are probably related to the non-linear effects in highly excited resonators. Two samples presenting different electric phase noises were analyzed by classic X-ray topographs (Fig.6a-6b). A correlation between dislocations densities and electrical phase noises might be drawn from these observations. The sample of lower phase noise shown in figure 6a has a low density of dislocations and its vibration shape seems to be regular. In contrary for the sample of higher phase noise shown in figure 6b, the dislocation density is higher and the vibration shape is disturbed by dislocations. This point has to be confirmed by further observations.

III. Vibration modes observed by synchrotron X-ray topography

Identifications of three components u_1, u_2 and u_3 of the piezo-electric displacement **u** were carried out using Laue setting. In the case of quartz AT resonators, the main component u_1 is visualized by the reflection $\mathbf{g}=2\bar{1}0$ and the two secondary components u_2 and u_3 can be approximately seen by $\mathbf{g}=0\bar{1}2$ and $\mathbf{g}=011$ respectively. For the cases shown below where the distinction between u_2 and u_3 is unnecessary, the reflection used was $\mathbf{g}=0\bar{1}3$ and the component observed is noticed as u_2/u_3.

Coupled standing and untrapped components :

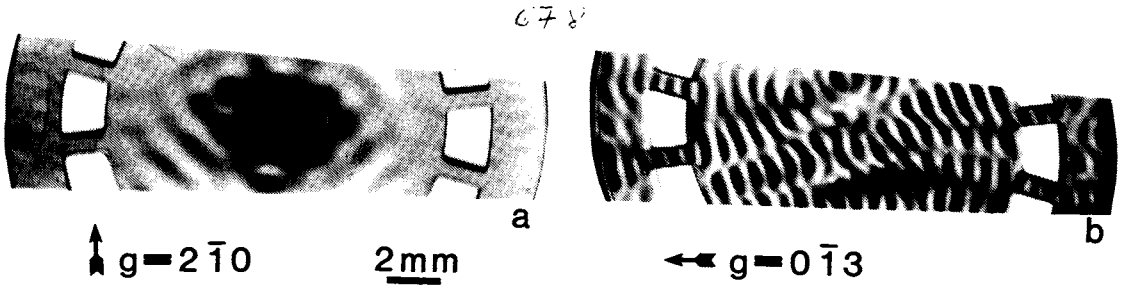


Fig.7. Coupled standing and untrapped u_2/u_3 components observed by classic X-ray topographs in a QAS AT-cut resonator operating at the fundamental thickness shear mode. a) u_1 component imaged with $\mathbf{g}=2\bar{1}0$; *b) u_2/u_3 components imaged with* $\mathbf{g}=0\bar{1}3$.

For untrapped modes, in plano-plane resonators or in plano-convex resonators operating at low overtone modes, u_2/u_3 components coupled to the main u_1 one have always been observed. Coupled components are temporally standing and spatially extending ones. They are visibly due to lateral boundary conditions of resonators and have a very complex spatial form. Figures 7a-7b show a case of a QAS AT-cut resonator operating at the fundamental thickness shear mode. It is to be noticed that coupled components extends up to the external part of the resonator and have high values in comparison with the main u_1 one.

Coupled progressive and untrapped components :

Temporally progressive components were observed in quartz by stroboscopic X-ray topography. It is to be recalled here that temporally progressive components are not visible on classic X-ray topographs. Figures 8a-8b are two stroboscopic topographs showing respectively the u_1 and u_2/u_3 components. In figure 8a the main u_1 component is seen by the central black contrast (standing component) with, in addition, extending irregular contrasts (progressive component) over the whole resonator surface. The u_2/u_3 components seen in figure 8b are entirely progressive and their extensions up to the resonator edges have to be considered in connection to the acoustic energy losses in resonators. This type of progressive and extending piezo-electric components was also observed in crystals of lithium tantalate [8].

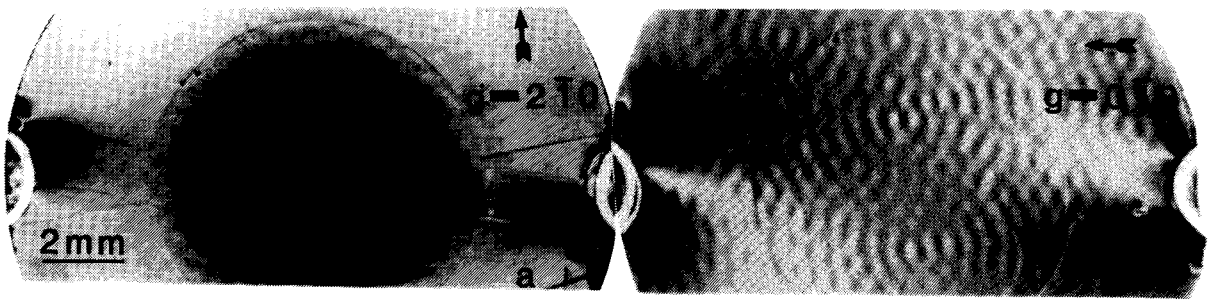


Fig.8. Stroboscopic X-ray topographs of a quartz AT-cut plano-convex resonator showing temporally progressive components. a) u_1 component imaged with $\mathbf{g}=2\bar{1}0$; *b) u_2/u_3 components imaged with* $\mathbf{g}=0\bar{1}3$.

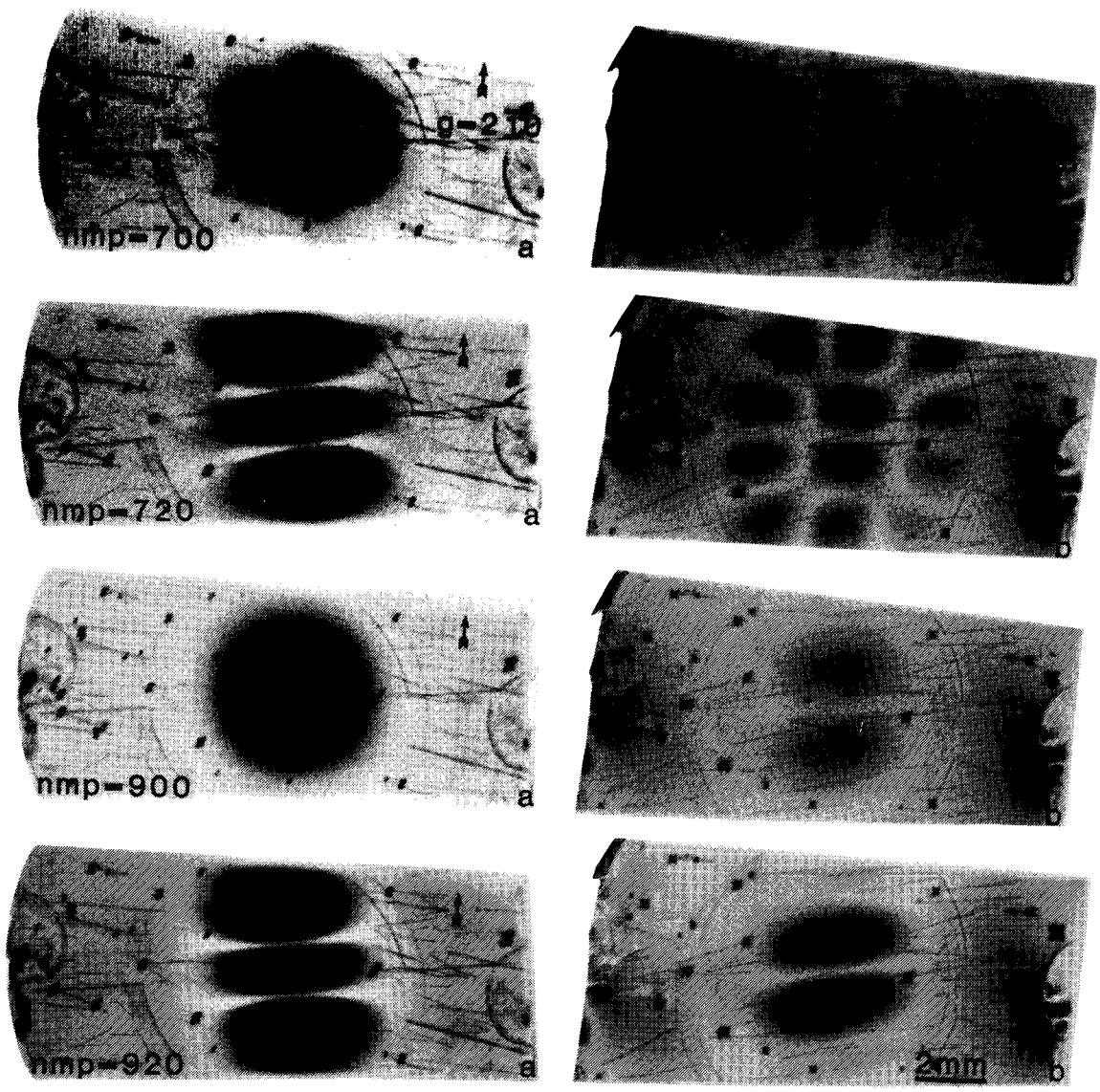


Fig.9 - Fig.12. Coupled u_1 and u_3 modes observed by classic X-ray topographs in a quartz AT-cut plano-convex resonator operating at respectively 7th and 9th overtone thickness shear modes. a) main u_1 modes imaged with $\mathbf{g}=2\bar{1}0$; *b) coupled u_3 modes imaged with* $\mathbf{g}=0\bar{1}3$.

Coupled standing and trapped u_1 and u_3 modes :

From the analysis of quartz plano-convex resonators, it was estimated that the u_2 and u_3 components are approximately of one and two orders of magnitude lower than the main u_1 one for usual cases [9; 10]. Cases of strongly coupled u_1 modes were evidenced by classic X-ray topography and observations indicate that the assumptions made previously in the analysis may fail in some particular cases. Figures 9-12 are topographs of a quartz plano-convex AT resonator operating at respectively 7th and 9th overtone thickness shear modes (nmp=700, 720, 900 and 920). In the column a it is shown the usual u_1 modes and in the column b corresponding coupled u_3 modes can be seen. It is interesting to notice that the coupled u_3 modes have perfect geometric forms and they are trapped in the central part of the resonator in the same way than the u_1 component.

The last point indicates that the coupling mechanism is not due to lateral boundary conditions, but rather intrinsic. The most probable reason of this coupling is that a secondary u_3 mode, usually not excited for the AT-cut geometry, is undirectly driven because its resonance frequency is close to the u_1 one. A very simplified calculation according to Tiersten's procedure can be used to check this point in estimating u_1 and u_3 resonance frequencies for this case. In treating u_1, u_2 and u_3 independently and in neglecting coupling coefficients c_{24}, c_{34} and c_{56} and electrode effects, uncoupled u_1 and u_3 resonance frequencies can be estimated using formula similar to Tiersten's ones in replacing $\bar{c}_{66}$, M_n and P_n by corresponding values :

- for u_1 modes

$$f_{nmp}^2 \approx \frac{\bar{c}_{66} n^2/4}{\rho (2h_o)^2} \left\{ 1 + \frac{\sqrt{2h_o/R}}{n\pi} \left[(2m+1) \sqrt{\frac{c_{11}}{\bar{c}_{66}}} + (2p+1) \sqrt{\frac{c_{55}}{\bar{c}_{66}}} \right] \right\}$$

with n odd, m and p even;

and

- for u_3 modes

$$f_{nmp}^2 \approx \frac{c_{44} n^2/4}{\rho (2h_o)^2} \left\{ 1 + \frac{\sqrt{2h_o/R}}{n\pi} \left[(2m+1) \sqrt{\frac{c_{55}}{c_{44}}} + (2p+1) \sqrt{\frac{c_{33}}{c_{44}}} \right] \right\}$$

with n even, m and p different parities.

Table I. Observed coupled u_1 and u_3 modes and their estimated resonance frequencies

resonator parameters : $2h_o$=1.3270mm, R=200mm $2\rho'h'/\rho h_o$ adjusted for f 500 (u_1) = 2 x 3.169280MHz			
observed modes		**calculated f_{nmp} (MHz)**	
u_1	**u_3**	**u_1**	**u_3**
700	612	8.803	8.821
720	632	8.914	8.918
900	810	11.303	11.568
920	810	11.414	11.568

In such a way, uncoupled u_1 and u_3 frequencies are estimated and the comparison with the experimental observations is summarized in Table I. It can be noticed that coupled modes observed have effectively close resonance frequencies. According to this hypothesis and by varying values of curvature radius and maximal thickness, u_1 and u_3 resonance frequency spectrum shall be consequently modified so that coupled u_1 and u_3 modes shall change. Resonators with different values of curvature radius and maximal thickness were checked and expected experimental results were obtained in those cases. A complete theoretical modelling has to be carried out in order to predict this kind of coupled modes, to calculate precisely amplitudes of both u_1 and u_3 components and to get more realistic piezo-electric parameters in these cases.

Coupled u_1 and u_1 modes

An interesting case of mode coupling was observed in highly excited resonators. It concerned a coupling between two u_1 modes which have a relative time phase :

$$u_{1\,(total)} = u_1^{(1)} e^{i\omega t} + u_1^{(2)} e^{i(\omega t + \phi)}$$

where $u_1^{(1)}$ represents the main mode, $u_1^{(2)}$ a secondary one and ϕ the relative time phase between two modes. Using the stroboscopic technique and by varying the relative time phase between X-ray beam and vibrations in such a way that a mode disappears from the observation, two modes were separated. While the main $u_1^{(1)}$ mode excited was the fifth overtone thickness shear one, the secondary $u_1^{(2)}$ mode observed might be assigned to the closest antisymmetric anharmonics. This observation, available only by the stroboscopic technique, shall be instructive concerning non-linear effects in highly excited resonators. For the present case it was observed a correlation between the detection of the secondary mode and the electrical hysteresis phenomenon related to the anisochronism effect.

Conclusion

It has been shown through this study that different vibrational aspects of resonators can be finely analyzed using synchrotron X-ray topography. Particularly interesting facts on mode coupling in quartz were evidenced, namely concerning the existence of temporally progressive components and coupled u_1 and u_3 modes in plano-convex AT resonators. Similar investigations are being carried out on other piezo-electric materials such as berlinite and lithium tantalate and further experimental informations are expected.

Acknowledgement : The authors would like to acknowledge SICN and CEPE companies and Dr. Euler of RADC (USA) for kindly providing samples studied here. This work was supported by D.R.E.T. (France) through the contract no. 87 34 13 80 04 70 75 01.

References :

[1] W.J. SPENCER (1972). Phys. Acoust., **9**, 167.

[2] A. ZARKA, B. CAPELLE, J. DETAINT and J. SCHWARTZEL (1988). J. Appl. Cryst., **21**, 967-971.

[3] S. TAKAGI (1969). J. Phys. Soc. Jap., **26**, 1239.

[4] N. KATO (1963; 1964). J. Phys. Soc. Jap., **18**, 1785; **19**, 67; **19**, 971.

[5] Y. ZHENG, A. ZARKA, B. CAPELLE, J. DETAINT and J. SCHWARTZEL (1989). Acta Cryst., A**45**, 275-285.

[6] B. CAPELLE, A. ZARKA, Y. ZHENG, J. DETAINT and J. SCHWARTZEL (1989). Proc. of the 43rd Annual Symp. on Freq. Control, p.470-476, 2 May - 2 June 1989, Denver, USA.

[7] Y. ZHENG, A. ZARKA, B. CAPELLE, J. DETAINT and J. SCHWARTZEL (1989). Proc. du 3ème Forum Européen Temps et Frequence, p.301-304, 21-23 Mars 1989, Besançon, France.

[8] J. DETAINT, J. SCHWARTZEL, H. CARRU, R. LEFEVRE, C. JOLY, B. CAPELLE, Y. ZHENG and A. ZARKA (1990). In this Proceedings.

[9] H.F. TIERSTEN and R.C. SMYTHE (1979). J. Acoust. Soc. Am., **65**, 1455-1460.

[10] D.S. STEVENS and H.F. TIERSTEN (1986). J. Acoust. Soc. Am., **79**, 1811-1826.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session ROM-05 Number 5

THE USE OF ANNULAR DIAMOND SAWS FOR CUTTING OF QUARTZ PLATES TO HIGH ANGLE ACCURACY

J Dowsett
R B Spencer
P Morley

STC Components, Quartz Crystal Division

Harlow, Essex, England, CM20 2DE

Summary

For many years the Quartz Crystal Industry has used as a standard method the multiblade slurry saw for the cutting of Quartz resonator plates. Within the last few years STC Crystal Division has commissioned a large annular saw with a thin stressed blade for the cutting of tight tolerance blanks. Two previous papers given in 1987 and 1988 described the commissioning results. This paper is a report of the results now being achieved in full production from a number of machines of AT cut blanks. It also describes the extension of its use into cutting of doubly rotated cuts such as the SC cut.

The Annular Saw

Recent advances in the Semi-conductor industry have lead to the development of large annular saws using a thin stressed blade with an inner diamond load edge, capable of cutting large diameter silicon bars. We have modified the machine to handle the cutting of large dimension quartz bars typically 43 x 24 mm cross section. Figure 1 shows a view of the machine construction. The block is supported on a slide mechanism which is capable of correcting the ZZ and XX directions in the block. The target angle can be achieved by obtaining corrections from the first slice which is sufficiently flat and parallel and of good finish to obtain correction to 0.25 minutes of arc by X-ray orientation.

The machine is capable of indexing each feed step to maintain thickness control to better than 15 uM. Coupled with the cut feed mechanism is a blade distortion sensor which detects blade deflection on a continuous analogue plot and gives warning of blade wear and the need to redress the diamond cutting edge.

The quartz block is lumbered in the normal way, a reference face is prepared on the -X, this face is prepared with an angle tolerance of better than 5' to minimise machine resetting between block loads. The block is cemented to a glass plate to ensure a complete cut through the quartz block. See figure 2.

The diamond loaded cutting blade is shown in figure 3. This blade has a 9" internal diameter and a thickness of 175 uM, the diamond cutting edge produces a Kerf of 300 uM and blank thickness of less than 200 uM have been achieved. Work on the blade has resulted in a blade life with a capability to cut quartz slices from which 27,000 blanks of 8mm diameter can be produced.

Use for Cutting Doubly Rotated Cuts

The extension of the application of the annular saw to cutting of doubly rotated cuts has produced a further significant application of this type of saw. With the capability to adjust the cutting angle in two planes it was evident that it should be possible to use it in this mode. The results achieved have made possible the economic cutting of SC blanks to close angle tolerances thus eliminating angle correction on the cut blank.

The method used was to produce a quartz block cut to include the first orientation. This was then mounted in the usual way and the second orientation set by the controls on the machine, however it was also possible to adjust for the first orientation.

The results from these were analysed, see figures 4 - 6 and it was found that the Frequency/Temperature characters were within the target specification for all resonators produced. The estimated angle distribution from this Frequency/Temperature data was found to be within 1.25' of arc. The electrical parameter showed excellent control.

Results

A comparison of results achieved from the annular saw initially on commissioning and in production with those from multiblade slurry saw are given in figures 7 - 9. Figure 7 shows the angle distribution now being achieved in production. Figures 8 and 9 give a comparative summary of the economy and performance

of the saws. Significant improvements are still being achieved by improved blade life and blade cutting accuracy with the consequent improvements in the efficiency and yield of the method. It is particularly notable that the losses of blanks for poor XX angle control have been practically eliminated and the yield of blanks within an angle tolerance of ±0.5′ on ZZ angle is now running at 3.5 times that previously achieved on the slurry saw.

Conclusion

The advantages of the annular saw are:-

1) Ease of setting to target angle

2) Improved yield to ±0.5 angle tolerance

3) Improved thickness and parallelism control

4) Significant improvement in an operating economy

5) Capability to cut doubly Rotated cut crystals into angle tolerance.

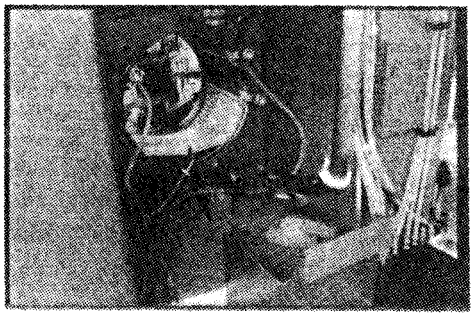


fig 1

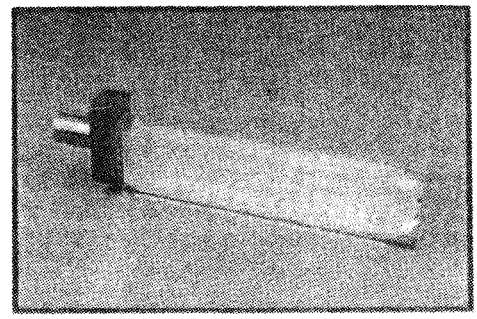


fig 2

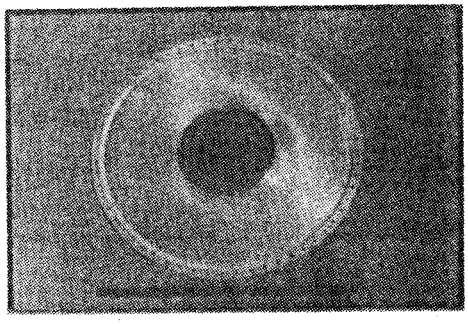


fig 3

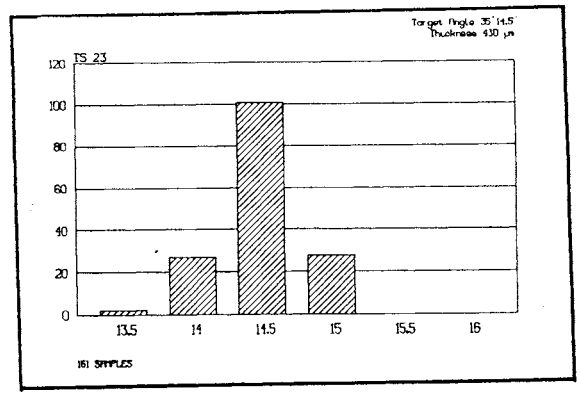


fig 4

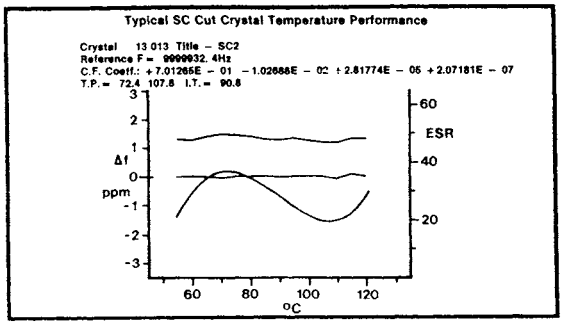


fig 5

SC Cut Crystals

Crystal code	fs (kHz)	drive (uW)	R1 (ohms)	C1 (fF)	Q	CO (pF)
1	9999.799	23.84	45.82	.27	1297978	4.41
2	9999.839	23.25	49.65	.26	1213929	4.53
3	9999.781	23.81	45.97	.26	1351175	4.40
4	9999.759	23.72	46.58	.26	1308417	4.50
5	9999.782	23.34	49.06	.25	1302103	4.51
6	9999.802	23.95	45.07	.24	1458610	4.49
7	9999.805	23.80	46.06	.25	1401493	4.57
8	9999.781	23.69	46.76	.26	1293214	4.50
9	9999.779	23.75	46.41	.25	1372580	4.50
10	9999.801	23.95	45.07	.25	1403294	4.54
11	9999.782	23.85	45.71	.26	1335856	4.52
12	9999.796	23.87	45.60	.25	1395860	4.54
13	9999.795	23.67	46.89	.26	1314931	4.40
14	9999.823	23.05	50.93	.24	1291696	4.48
15	9999.804	23.54	47.74	.24	1378587	4.58
16	9999.790	22.11	57.07	.26	1052626	4.49
17	9999.793	18.63	82.51	.25	762271	4.54
18	9999.816	23.60	47.39	.27	1253322	4.45

Complete batch Electrical parameters

fig 6

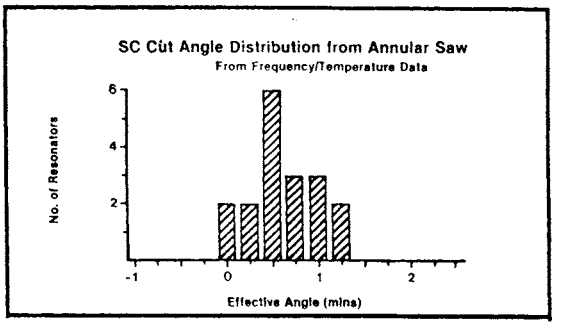


fig 7

Summary of Results (Blanks)

	Multiblade Saw	Annular Saw Initial	Annular Saw Production
Angle Control Z Z angle			
Target Angle ± 0.25'	75%	95%	95% +
Angle Spread	3-4'	2'	1.5'
Blank yield ± 0.5'	10%	30%	35%
Angle Control XX Angle ± 10'			
Yield	10-15%	95%	98%
Blank Parallelism	6.0μm	5-10μm	5-10μm
Blank Thickness Control	100μm	15μm	15μm

fig 8

Summary of Results (Sawing Economy)

	Multiblade Saw	Annular Saw Initial Results	Annular Saw Production Results
— Sawing rate	3.2 cm^2/min	3.5 cm^2/min	4.2 cm^2/min
— Blade life	8,515 cm^2	18,920 cm^2	35,000 cm^2
— Quartz utilisation/block	150 slices	175 slices	180 slices
— Blade change time	0.75 hrs	1.5 hrs	1.5 hours

fig 9

Applications

APP-05 Time Transfer in Geodynamics and Radioastronomy

Chairman: F. Walls

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-05 Number 1

APPLICATIONS OF HYDROGEN MASERS TO DEEP SPACE TRACKING AND RADIO SCIENCE

Richard L. Sydnor*
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109 U.S.A.

Abstract

This paper describes the applications for Hydrogen Masers in the National Aeronautics and Space Administration (NASA)/Jet Propulsion Laboratory (JPL) Deep Space Network (DSN). These application include tracking and navigation of spacecraft at extreme distances and special experiments based on the propagation characteristics of planetary and solar atmospheres, diffraction and refraction of the signals and the effect of gravity waves on the signals. A set of requirements to perform these tasks has been developed. These requirements are based on the detection of gravity waves while tracking spacecraft at distances exceeding several AU, studying planetary atmospheres or rings and precise station location using VLBI. The requirements for gravity wave detection are 2 x 10^{-15} frequency stability for 1000 to 3600 seconds. This requirement will become more stringent in the next few years. Investigation of planetary atmospheres and rings requires spectral purity of -54 dBc one Hertz from the carrier at 8.1 GHz, -64 dBc at ten Hertz and -75 dBc at 100 Hertz and higher. This performance must be obtained under field conditions with temperature and vibration environmental characteristics very different from laboratory conditions. The combination of performance requirements requires the use of the best available atomic and crystal oscillators in this environment. System and component design to meet these requirements will be discussed as will the extensive testing and packaging necessary that the requirements be met. Future needs will be outlined and the research program under way to meet these needs will be described.

Introduction

The DSN consists of three complexes located in Goldstone, California; Robledo, Spain; and Tidbinbilla, Australia as shown in Figure 1. Each of these complexes consists of three or more antennas including one 70 meter antenna and one or more 36 meter antennas. All the complexes are linked by a system of land lines and satellite communications. They are all controlled from the Space Fight Operating Facility at the Jet Propulsion Laboratory in Pasadena, California. Synchronization of the stations in time and syntonization in frequency are performed using the GPS satellite system.

Figure 1
DSN Complex Locations

The entire complex uses one Frequency and Timing System (FTS). A typical FTS is shown in Figure 2. The important features are: (1) the redundant frequency standards (2 Hydrogen Masers as the first two prime standards and 2 Cesiums for the next two levels of backup); (2) triply redundant clocks driven from a flywheel crystal oscillator (Disciplined Oscillator) to maintain continuous clock operation in the event of frequency standard switching; (3) Time Insertion/Deletion System (TIDS) which individually addresses the Time Code Translators (TCT) to run on UTC or on 'Simulation Time', a time whose epoch is controllable from either the built-in keyboard or from the DSS main control computer to be offset from UTC in integral seconds by as much as one year for practise and check-out of mission software; 4) Time Code Generator and Time Code Translators which operate from the 5 MHz distribution system and are synchronized by a modified IRIG-G signal, the time code translators are located near the user equipment and are adjustable in time delay in steps of 100 nsecond to compensate for the different cable delays and to keep the entire station synchronized within 50 nsecond; 5) Phase Calibration Generator which actively stabilizes the cable between the control room and the antenna feed area and injects a comb spectrum into the antenna horn so as to calibrate the receiver delays in near real time during the actual tracking of the spacecraft and 6) GPS receiver to maintain synchronization and syntonization between the different DSS's and with NIST.

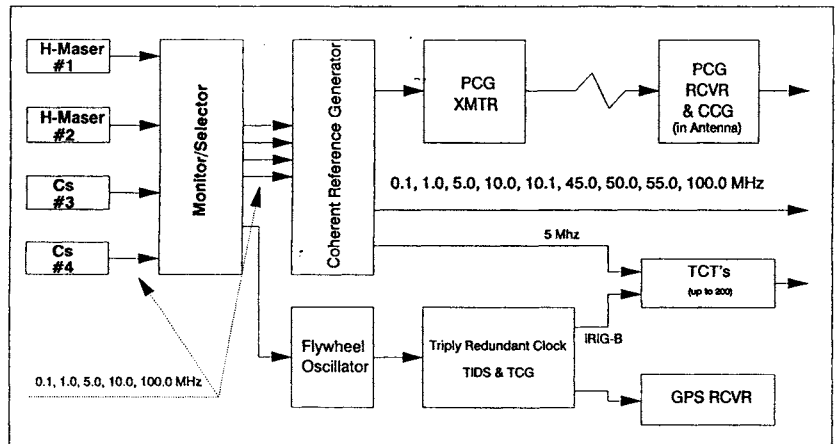


Figure 2
DSN Frequency and Timing System

Navigation Requirements

There are two sets of requirements that have to do with navigation of spacecraft.

For the classical navigation, which performs ranging by measuring the round-trip-light-time (RTLT) of a signal from the station to the spacecraft and return, the requirement is based on 1) minimizing the noise on the signal as received at the station which gives an Allan Variance requirement for an averaging time of 1 second and 2) minimizing the inaccuracy due to the change in frequency between the time of transmittal of the signal and the reception of the signal for times (i.e. the RTLT) as great as 8½ hours (≈ 30000 seconds) for Neptune. These numbers are given in Table 1.

Since the classical navigation solution has low angular resolution near the ecliptic plane, a Differential VLBI (Very Long Baseline Interferometry) technique is used wherein the position of the spacecraft is measured relative to a set of radio stars which have been mapped very precisely over the last ten years. The requirements for this navigation method are just those of ordinary VLBI, i.e. small enough noise at the short term (≈ 1 to 10 seconds) and good enough stability over the VLBI run (≈ 8 hours) to not degrade the position information. These requirements are given in Table 2.

Table 1

FTS Requirements for Classical Deep Space Navigation

Averaging time τ, seconds	Allan Deviation, parts in 10^{15}
1	200.0
1000	1.5
30000	10.0

Table 2

FTS Requirements for VLBI Deep Space Tracking

Averaging Time, τ, seconds	Allan Deviation, parts in 10^{15}
1	300
1000	2
3600	2

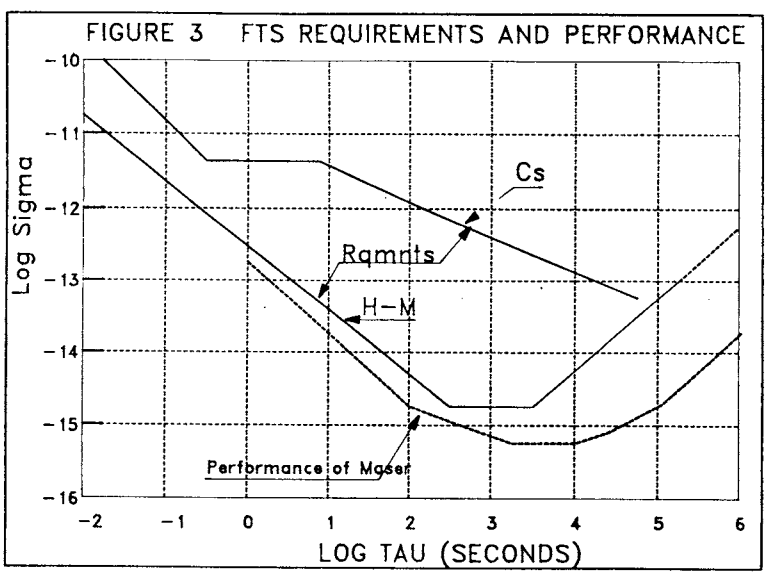
In addition to the stability requirements in Tables 1 and 2, There are requirements on accuracy and reliability given in Table 3. Included in this table are the time-to-repair numbers which are very important for deep space missions since there may only be a few minutes or hours to get data after a flight of 12 years. An FTS failure is catastrophic for the operation of the station and must be repaired or replaced in a very short time. The Monitor/Switching System shown in Figure 1 monitors many of the operating parameters of all the frequency standards and measures the differential phase between selected pairs of the standards for performance assurance. In event of failure of a standard a manual command switches the system to the next standard in the hierarchy.

In Table 3, the numbers given under 'Requirements' are for the entire FTS and are the ones dictated by the navigation problems. Internal requirements to the FTS are needed because of the necessity of accurately monitoring the absolute and relative time and frequency between the different standards and between the DSS's as shown in the column labeled 'Achieved Performance'. This explains the great disparity between some of the numbers in the two columns.

Table 3
FTS Timing and Miscellaneous Requirements

Parameter	Requirements	Achieved Performance
Temperature Variations		
In Frequency Standards Room	±0.1°C	
In Control Room	±3 °C	
In Antenna Room	±5 °C	
External Cable Runs	-20 °C to +55 °C	
Magnetic Field Variations	± 20 mg	
Humidity Variations	5% to 100%	
Time to Repair and/or Substitute Unit	15 minutes	15 minutes
Maximum Offset from NIST	±3 parts in 10^{13}	±3 parts in 10^{13}
Traceability to NIST	±1 part in 10^{13}	±1 part in 10^{13}
Timing offset from NIST	±20 μseconds	±5 μseconds
Knowledge of Time Offset from NIST	±10 μseconds	±20 nseconds
Knowledge of Internal Time Offsets	±10 μseconds	±100 nseconds

The Allan Variance for the FTS is shown in Figure 3. There are two sets of graphs in this figure. One is the performance of the entire FTS as guaranteed to the flight projects, the other is the performance of the frequency standards alone.



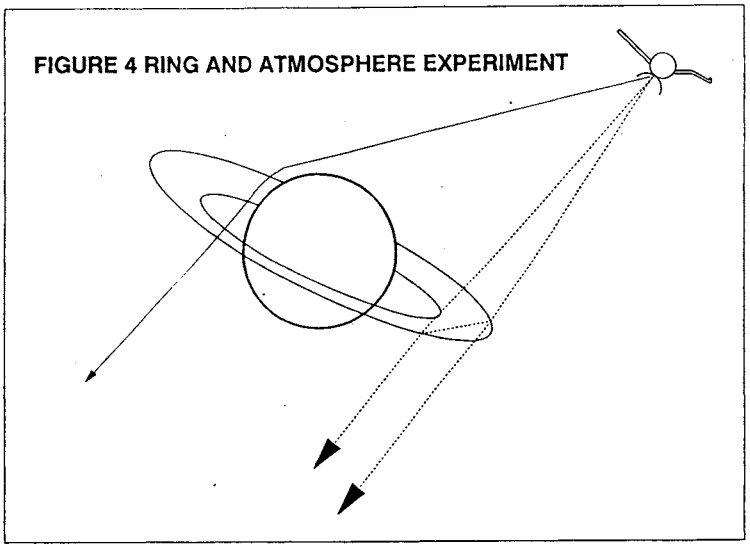
Radio Science Requirements

There are several Radio Science experiments that have more stringent requirements than the navigation requirements given above. These are 1) gravity wave detection and 2) ring and atmosphere research.

The search for gravity waves consists of using the RTLT to the distant spacecraft as a large gravity wave antenna. When the ground-spacecraft-ground signal path is perturbed by an impinging gravity wave, the Doppler is affected in a distinctive way. In order to detect this perturbation, the noise on the ground transmitter signal must be minimized. The spacecraft transponder is phase locked to the received signal from the ground transmitter and the ground receiver is locked to the received spacecraft signal, so the most critical element is the stability of the ground transmitter. The amplitude of the Doppler perturbation depends on the size of the vent that causes the gravity wave. The more stable the ground transmitter is, the smaller event that can be detected. Preliminary experiments were performed with the performance outlined above, but the size of the event that could be seen with that performance was very large and, as a result, very rare. In order to find gravity waves more often the capability to find smaller ones must be developed. The goals for the frequency standards in the future are given in the section on future improvements.

The ring and atmosphere research utilizes the (open loop) signal from the USO (Ultra-Stable-Oscillator) on the spacecraft which generates the down-link signal. As shown in Figure 4, there are two modes of operation.

For examination of the rings, the signal propagating through the rings suffers reflection and diffraction such that the signal received on the ground is modulated in phase and amplitude be the interference pattern of the various rays as the spacecraft passes behind the planet. Multiple Fourier Transforms of the received signal yield information about the sizes and density of the material in the rings.



The atmosphere refracts the signal passing through it as the spacecraft goes behind the planet. The effect is to bend the rays and make the signal continue slightly after the spacecraft disappears and to make the signal appear before the spacecraft reappears. Analysis of the phase variations and amplitude variations *vs.* time gives information about the atmospheric depth and density.

Both of these experiments depend on the spectral purity of the signal from the spacecraft and the local oscillator on the ground. The specification for the local oscillator for the last Neptune encounter and for future Mars Observer are given in Table 4. In order to meet the Neptune requirements the Hydrogen Masers in the critical stations were modified with improved quartz oscillators and frequency multipliers.

Table 4

Spectral Purity Requirements for Radio Science

Offset frequency	$S_\phi(f)$ Neptune (1989) at 5 MHz		$S_\phi(f)$ Mars Observer (1993) at 5 MHz
	REQUIRED	ACHIEVED	
1 Hz	-118 dBc	-124 dBc	-124 dBc
10 Hz	-124 dBc	-148 dBc	-132 dBc
100 Hz	-124 dBc	-153 dBc	-139 dBc
1 kHz	-124 dBc	-154 dBc	-139 dBc
10 kHz and up	-124 dBc	-154 dBc	-139 dBc

Research and Developments for Future Requirements

1) Cooled Sapphire 'Whispering Gallery' oscillators for improved short term (<200 seconds) stability, -146 dBc 1 Hz from the carrier at 5 Mhz, $\sigma_y(\tau) = 10^{-14}$ at $\tau = 1$ second

2) Improved Hydrogen Masers for improved medium term (200 to 10000 second) stability, $\sigma_y(\tau) = 3 \times 10^{-16}$ at $\tau = 1$ second

3) Trapped Ion Standards for improved medium to long term (>3000 seconds) stability, $5 \times 10^{-13}/\sqrt{\tau}$

4) Improved Fiber Optic Distribution for low degradation of the frequency standards performance, $\sigma_y(\tau) < 10^{-17}$ at $\tau = 1000$ seconds

5) Two-way time transfer for better synchronization and syntonization of the DSN and between the DSN and NIST, $\Delta t \approx 1$ nsecond

Efforts to achieve these goals are being pursued by people at JPL and by other research organizations which are supported in these efforts by JPL. The work on the sapphire resonator oscillators is being done at JPL and the Applied Physics Laboratory of Johns Hopkins University; the Hydrogen Maser work is being done at the Smithsonian Astrophysical Observatory and The Applied Physics Laboratory; the Trapped Ion work is being done at JPL as is the Fiber Optics work. The work on two-way time transfer has not started yet, but will be done at JPL and NIST.

Session APP-05 Number 2

Time and Frequency Facilities at the Applied Physics Laboratory

Joseph J. Suter
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20707
U.S.A.

In this paper we will discuss the performance of frequency standards in APL's Time and Frequency Laboratory. The Johns Hopkins University Applied Physics Laboratory operates a Time and Frequency Laboratory in support of space flight programs. This laboratory operates three APL-developed NR hydrogen masers and two HP 5061 (A and B) cesium standards with high performance beam tubes. Time comparison with other time and frequency laboratories is accomplished with two NIST-type GPS receivers using established common view techniques. Performance comparison between the hydrogen masers and cesium standards is done by using an HP 5335 counter-based data acquisition system. This system measures the relative drift of the one PPS signal from each of the standards on an hourly basis. The drift rates of the atomic hydrogen masers is measured with a ten Hz phase comparison system. The resolution of this system is approximately 0.05 ns. The drift rate of the masers is also continuously monitored as a function of ambient temperature, relative humidity and barometric pressure.

The hydrogen masers in the Time and Frequency Laboratory have, dependent on the configuration of their resonant cavity, different temperature coefficients. The NRX maser has a temperature coefficient of 1.8 kHz/°C and its cavity is of so-called integral design (Ref. 1). The NRB-2 maser, which has a quartz liner, has a temperature coefficient of 6.65 kHz/°C. The earlier type NR maser (NR-6), which has an all aluminum cavity, has a temperature coefficient of 33 kHz/°C (Ref. 2). This maser will be upgraded with a silver-coated quartz liner in the near future.

The different in temperature coefficients of each maser affects their thermal time constants and hence their susceptibility to temperature variations. Figures 1 and 2 present data on the stability of the NRB-1 and NR-6 masers over the periods indicated. Recall that NRB-1 has a silver-coated quartz liner placed in an aluminum cavity. It is apparent that the NRB-1 maser has a smaller sensitivity to temperature variations. Variations in temperature, humidity, and barometric pressure can limit the stability of atomic frequency standards and crystal oscillators (Ref. 3). A long term test has been planned at APL to attempt to examine these effects.

The hydrogen masers at APL are all of the so-called NR-type and have typical line Q's of about 1.2×10^9. Table I summarizes some of the more important characteristics of these masers. A modest program is presently underway at APL to improve the line Q, IF power output, and hence the stability of the maser (Ref. 4).

The hydrogen masers and cesium standards are mainly used as clocks. The stability of the one PPS signal is compared daily, via GPS common view, with USNO and NIST. The drift rates of the standards are monitored hourly and analyzed by BIPM on a monthly basis. The typical drift rates of the standards is shown in Table II and Figure 3. Note that the HP 5061B cesium standard is a new standard. Also, the HP 5061A received a new beam tube on MJD 47879. The drift rates of these cesium standards have been steadily improving since their first reported performance to BIPM. The performance of APL's atomic standard is reflected in BIPM's calculation of UTC-UTC[k]. Figure 4 presents the data of UTC-UTC[k] for APL, USNO, and NIST as calculated by BIPM. The period covered is MJD 47809 to MJD 47889.

FIGURE 1.
Temperature Stability of NRB-1 Maser

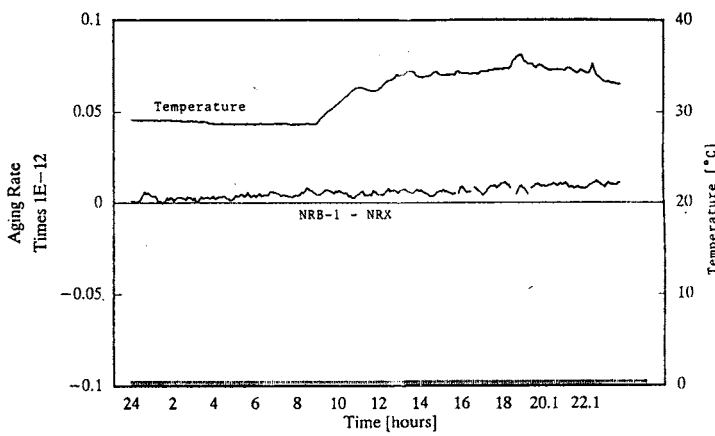
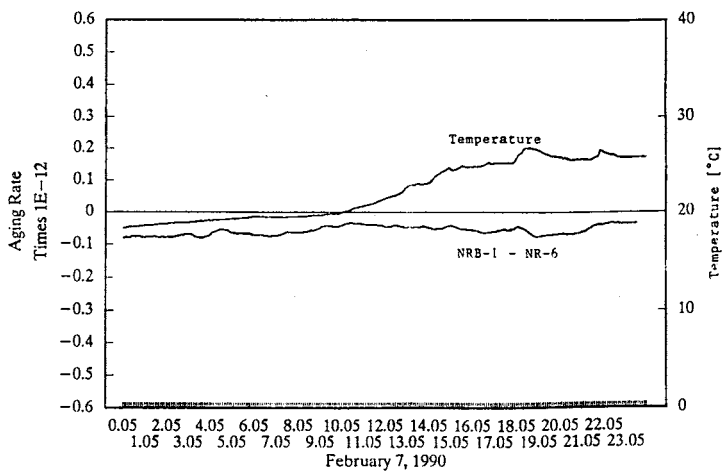


FIGURE 2.
Temperature Stability of NR-6 Maser



References

1. R. Kunski, "Integral Cavity Hydrogen Maser," Proceedings PTTI, 631-651 (1986).

2. L. J. Rueger, "Characteristics of the NASA Research Hydrogen Maser," J. Inst. of Electron. Telecommun. Eng., 27, 493-500 (1981).

3. J. R. Norton, "Ultrastable Quartz Oscillator for Spacecraft," To be published in 1989 PTTI Proceedings.

4. J. J. Suter, "Advances in the Development of Piezoelectric Quartz-Crystal Oscillators, Hydrogen Masers, and Superconducting Frequency Standards," Johns Hopkins APL Technical Digest, 9, 212-219 (1988).

Table I.

NR Maser Characteristics

$Q_{loaded} \cong 39500$

$Q_{unloaded} \cong 48500$

Zeeman Frequency: 380 Hz

Cavity Time Constant ≅ 0.6 sec.

Line Q ≅ 1.2×10^9

Cavity RF Output ≅ -93 dBm

State Selector: Quadrupole

Storage Bulb: 12.5 cm (diameter)
40.0 cm (long)
Teflon coated (FEP-120)

Table II.

Drift Rates+

(typical)

Atomic Standards	Model/Type	Drift Rates [ns/day]
Hydrogen Masers	NR-6	-1.6
	NR-X	0.5
Cesium Standards	Hp 5061A	-12.7
	Hp 5061B	15.9

+versus Hydrogen Maser NRB-1
50 days of observations
reliability of regression: $0.96 < R < 0.98$

FIGURE 3.

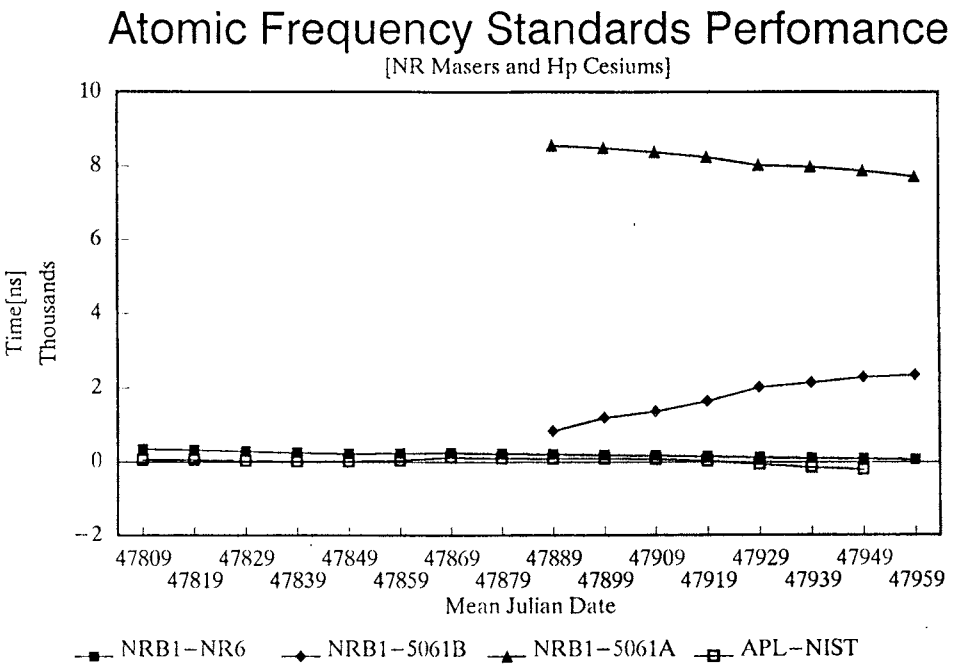
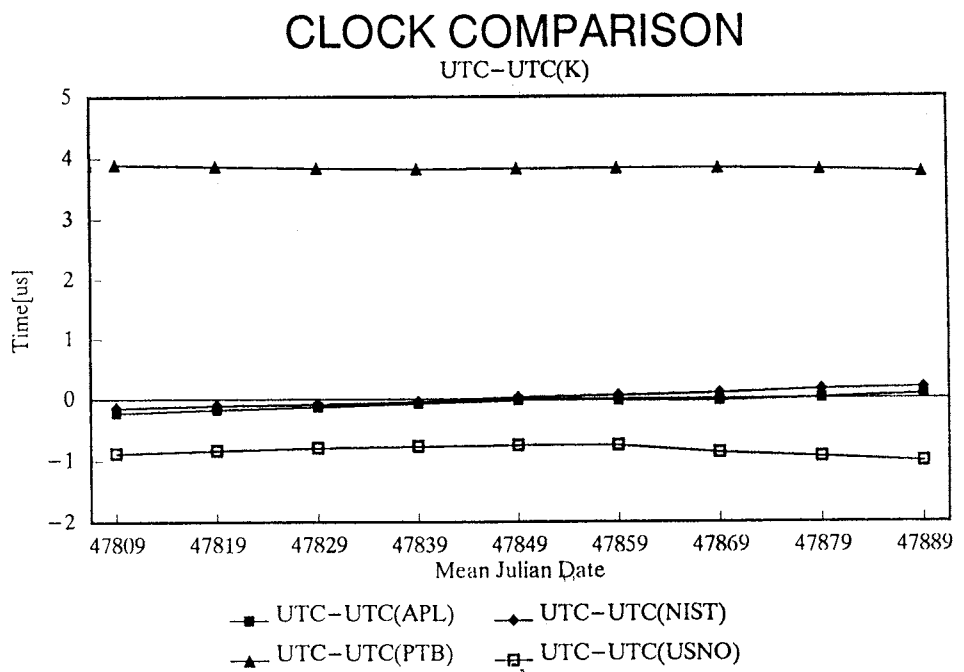


FIGURE 4.



Session APP-05 Number 3

Universal Time (UT1) Determination: Present and Future

W E Carter and D S Robertson, (both at National Geodetic Survey, Charting and Geodetic Services, National Ocean Service, NOAA, Rockville, MD 20852; telemail NGS.VLBI)

Abstract

Very-Long-Baseline Interferometry (VLBI) observations currently yield the highest accuracy and time resolution determinations of UT1. Multi-station 24-hour observing sessions yield accuracies of 0.03 to 0.05 millisecond, while single interferometer sessions of only 45 minutes duration attain 0.05 to 0.10 millisecond. The IRIS UT1 time series contains nearly continuous daily values since 1985. The sampling frequency will increase as additional VLBI facilities become operational, but VLBI becomes less and less attractive for monitoring UT1 as the sampling rates increase beyond a few per day. It would be desirable to have a more easily automated "continuously" operating system. The use of superfluid helium as the working medium for a gyroscope, exploiting the frictionless flow of the helium fluid and the quantized nature of its allowed angular momentum states, could yield a system with comparable accuracy, higher time resolution, and better operating characteristics than VLBI. Initial calculations indicate that the superfluid gyroscope could achieve improvements in sensitivity of six to nine orders of magnitude over ring-laser gyroscopes. The NGS and the AFGL in the United States and the IFAG in the FRG have agreed to jointly support the development of the first prototype operational superfluid gyroscope at the University of California, Berkeley. If no unanticipated problems are encountered a small scale superfluid gyroscope could be operational in 3 years and a full scale unit in five years. By the mid 1990's the combination of VLBI and superfluid gyroscope observations could provide an essentially continuous record of UT1 accurate to the few hundredths of a millisecond level.

Introduction

The rate of rotation of the Earth varies in a complex manner on time scales ranging from days to geological ages. Tidal variations resulting from the changes in the inertia tensor of the Earth as the Earth, Moon and Sun change their relative positions cause highly systematic variations over the full range of time scales, but are amenable to modeling. Improving the accuracy with which they can be computed is largely a matter of refining the Earth model used. For example, the Yoder *et al.* [1983] tidal series currently in general use includes the effects of core-mantle coupling and equilibrium oceanic response, but not those of anelasticity of the mantle and non-equilibrium oceanic responses.

From Rosen and Salstein's [1983] work, as well as that of others during the past decade [Carter *et al.*, 1984; Rosen *et al.*, 1984], we know that the exchange of angular momentum between the atmosphere and the solid earth causes large quasi-periodic and stochastic variations in Universal Time (UT1) on time scales of weeks to years. However, it is unlikely that complex weather related global atmospheric variations can be modeled in the foreseeable future. Even if this were possible, the formidable task of understanding the interactions between the atmosphere and the surfaces of the oceans and land masses sufficiently well to compute the changes in Earth rotation from the atmospheric data would remain. For the present at least, there is no choice but to observe the changes in UT1 as frequently and as accurately as technology and resources allow.

Current State-of-the-Art

The most accurate UT1 time series currently available are derived from VLBI observations. In geodetic VLBI, networks of radio telescopes located thousands of kilometers apart simultaneously track extragalactic radio sources. Each station records the microwave signals received in digital form on magnetic tape, along with precise time from Hydrogen MASER frequency standards. Typically 20 to 30 sources are tracked for periods of 3 to 6 minutes several times each over the course of a 24 hour observing session. The data tapes are sent to a correlator center where the signals are processed to determine the differences in arrival times (delays) and the changes in the delays with time (delay rates) between each pair of stations. The delays and delay rates are then analyzed by least squares techniques to estimate a variety of parameters including such nuisance parameters as the offsets and rates between station clocks and atmospheric refraction corrections, important global parameters such as the coordinates of the radio sources which define the celestial reference frame and of the stations which define the terrestrial reference frame, and parameters which describe the temporal variations in the orientation of the Earth, including precession, nutation, polar motion, and UT1.

The International Radio Interferometric Surveying Atlantic (IRIS-A) network of VLBI observatories is currently the primary source of high accuracy time series for the full suite of Earth orientation parameters, *i.e.*, precession, nutation, polar motion and UT1. This network consists of a set of core stations (Wettzell, Federal Republic of Germany; Westford, Massachusetts, Richmond, Florida, and Mojave, California) that participate in 24-hour observing sessions each 5 days, and

additional stations (Onsala, Sweden, and Medicina, Italy) that participate about once per month. Using the nutation, polar motion, and station and radio source positions obtained from these multi-station observing sessions it is possible to determine UT1 accurately with a single baseline interferometer in observing sessions of less than one hour [Carter and Robertson, 1985]. The Wettzell-Westford interferometer is used in this fashion to produce daily values of UT1.

The IRIS-A 5-day Earth orientation time series is continuous since January 1984. There is a three month segment of daily UT1 values beginning in April 1984, and the series is continuous since April 1985. Based on the agreement between the 5-day and daily time series and intercomparisons with length-of-day determinations from satellite laser ranging, we estimate the accuracy of IRIS-A UT1 time series to be typically about 0.05 milliseconds, perhaps reaching 0.03 milliseconds for the best 5-day values and deteriorating to 0.10 milliseconds for a fraction of the daily points.

Figure 1 shows the UT1 determinations since the daily observation series began in April, 1985. The plot shows both the daily 1-hour determinations and the 5-day 24-hour determinations. Long period drifts have been removed by fitting a parabola. The residuals contain several periodic signatures, including annual, semi-annual and bi-annual.

Figure 2 shows an amplitude spectrum of the data plotted in figure 1. In addition to the periods noted above, there is a shoulder on the strong annual peak, possibly indicative of another periodic component present in the data. To clarify the finer structure in UT1, four sinusoids were estimated and removed from the data. Both the periods and amplitudes were adjusted in the estimation process. Table 1 gives the periods (in days) and amplitudes (in milliseconds) of the sinusoidal variations removed.

Figure 3 shows the variations in UT1 after removing the sinusoids. The amplitude of the variations has decreased from about 35 milliseconds to 5-10 milliseconds, yet much structure remains. This is not unexpected: Most of the structure in UT1 at harmonics and sub-harmonics of one year is produced by meteorological effects, which are almost never well modeled by a small number of simple sinusoids. We would have to remove a large number of additional terms to begin to approach the noise level of the data.

Table 1: Periods and Amplitudes of sinusoids fit to the UT1 data shown in figure 1.

Period	Amplitude
183.121	10.667
361.109	21.109
525.310	11.886
659.018	18.691

To examine the stability of the sinusoidal components, we have extrapolated the sinusoids fit to the data from 1985 to the present backwards, to cover the period from 1980-1985, for which a small amount of VLBI UT1 determinations are available. Those determinations, again with a parabola removed, are displayed in figure 4 together with a line indicating the extrapolated sinusoids. It is clear from the figure that the extrapolation is not very successful in "retro-dicting" the behavior of UT1 during this time period. Again, this is not unexpected--the meteorological effects that are driving these variations are not stationary in time, and are not well modeled by a small number of sinusoids. Many more sophisticated techniques have been attempted to characterize these variations and extrapolate their effects for short periods away from the actual measurement, without conspicuous success.

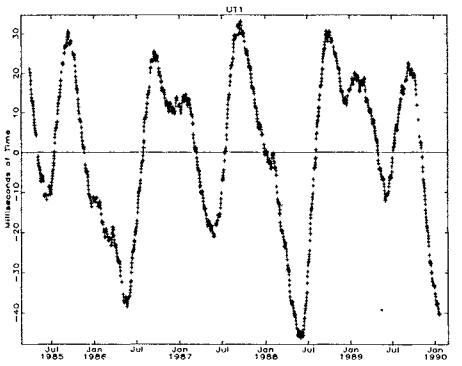


Figure 1. UT1 determinations for a 4.5 year period starting in 1985

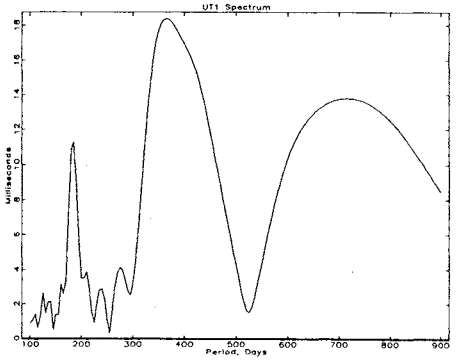


Figure 2. Amplitude spectrum of the UT1 data shown in figure 1.

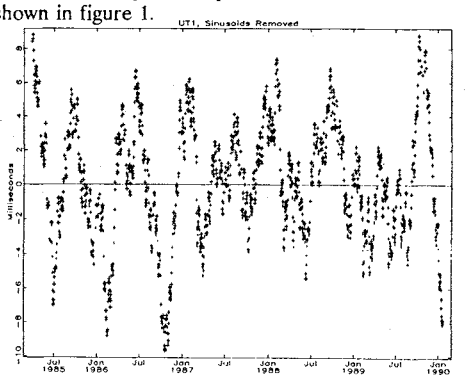


Figure 3. UT1 data after removal of four sinusoids, as described in the text.

Future Improvements in VLBI Determinations Of UT1

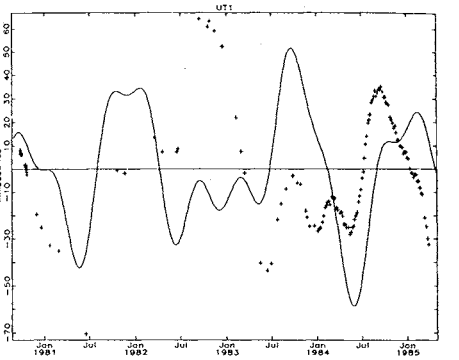


Figure 4. UT1 determinations from the 1980-1984 time frame, along with periodic components extrapolated from the later UT1 determinations.

In principal the temporal resolution of the UT1 time series could be improved by more than an order of magnitude simply by increasing the number of Wettzell-Westford (or other nearly equatorial interferometer) observing sessions conducted each day. There is no technical development required, just additional funds. In order to justify the additional expenditure of funds we require some evidence that the rate of rotation might vary significantly on time scales of a few hours and an application for the higher frequency information. Carter and Robertson [1985] found some evidence of changes in UT1 of several tenths of milliseconds over time periods as short as 6-8 hours, but the observations used were not optimized for measuring UT1. More recently Herring [personal communication] has processed data from 24-hour observing sessions using a Kalman filter to follow high frequency variations in UT1 and has detected an apparent approximately semi-diurnal variation (perhaps a lunar tidal effect) with an amplitude of a few tenths of a millisecond. During October-November 1989 the NASA Crustal Dynamics Project organized an intensive observing campaign to test improved VLBI instrumentation (wider bandwidth receivers, improved calibration) that collected nearly continuous data for a period of weeks. When these observations are reduced and analyzed they may shed more light on high frequency UT1 variations.

The National Earth Orientation Service (NEOS), operated jointly by the National Geodetic Survey and the U.S. Naval Observatory, is currently performing a one year observational program to develop unified procedures, reference frames, and schedules which will allow the IRIS and NAVNET (Navy Network, consisting of observatories at Richmond, FL, Greenbank, WV, Kauai, HI, and Fairbanks, AK) data to be fully integrated. Once it has been demonstrated that systematic differences do not exist or can be adequately removed the two observing programs will yield the entire suite of Earth orientation parameters at three-and-one-half day intervals and UT1 at 12-hour intervals. It is likely that the frequency of UT1 only determinations could be increased to 6-hour intervals if studies such as those already described confirm that significant variations occur on that time scale. But even with projected improvements the feasible temporal resolution of the UT1 time series achievable with VLBI will be limited and a totally different approach such as a gyroscope may prove more accurate and cost effective for monitoring high frequency variations in UT1.

Superfluid Gyroscopes

A team of researchers at the University of California at Berkeley led by R. E. Packard has proposed to develop a new type of gyroscope, using superfluid helium, that could ultimately achieve a sensitivity of 10^{-13} radians per second in a few seconds [Packard, 1989]. Such an instrument could provide a continuous record of length-of-day, and by integration UT1, with accuracies comparable to VLBI, at significantly lower operating costs. The Air Force Geophysical Laboratories (AFGL), NGS, and the Institut fur Angewandte Geodesie, FRG, have agreed to jointly support the development of a small scale prototype instrument to demonstrate the basic physics. Contractual arrangements for the 3-year project were completed by AFGL and work began on the first phase early this year. There has not yet been sufficient time to obtain experimental results from this new effort, and the physics are far too complex to present in detail here, but we will briefly introduce the basic concepts behind a superfluid gyroscope.

Superfluids, *e.g.*, ^{4}He below 2 degrees Kelvin and ^{3}He below 2 millidegrees Kelvin, have the unique ability to flow without dissipation. Within the superfluid a finite fraction of all the particles is in the same quantum state, and that state can be described by an order parameter which is in essence a macroscopic wave function, *i.e.*, a quantum mechanical wave function describing a macroscopic object. The detailed form of the wave function depends on the specific superfluid but always involves complex numbers having a well defined phase. A uniform velocity produces a gradient in the quantum phase of a superfluid, and this property provides the basis for developing a superfluid gyroscope.

Consider for example what happens when a blocked toroidal enclosure (Figure 5) filled with a superfluid is rotated about an axis perpendicular to the plane of the toroid with an angular velocity ω. The resulting phase gradient will create a phase difference δ across the blockage, which will be proportional to ω. If δ can be measured, ω can be computed. Deriving the precise mathematical relationship between δ and ω is non-trivial, but it turns out that just as with a Sagnac "ring-laser" interferometer, the phase shift per unit change in velocity, *i.e.*, the sensitivity of the gyroscope, increases as the area enclosed by the toroid. However, because the phase shift in the superfluid is dependent on the mass of the ^{4}He atom (or, in the case of ^{3}He the Cooper pair mass, which is twice the mass of the ^{3}He atom), while the phase shift in the ring-laser gyroscope depends on the much smaller relativistic mass of a photon, a superfluid gyroscope is potentially some

10^9 times more sensitive. The Earth's rotation would produce a phase difference of more than 2π in a superfluid gyroscope enclosing an area of only 10 cm^2.

The principles of detecting rotation with a superfluid gyroscope have been known for decades, and have been discussed in publications dating back several years (see, for example, Chiao [1982] and Anandan [1984]). However, until recently there was no known method for accurately measuring the phase difference δ. The situation changed dramatically when Avenel and Varoquaux [1988] published two new observational results. First, they found that in ^{4}He a well defined critical phase difference can be detected across a small orifice in a thin wall. When the velocity induced phase difference across the wall exceeds this critical value a 2π phase slip occurs. This phenomenon appears to be highly reproducible, and may provide the means to develop a gyroscope using ^{4}He, which Packard refers to as a phase slip gyroscope.

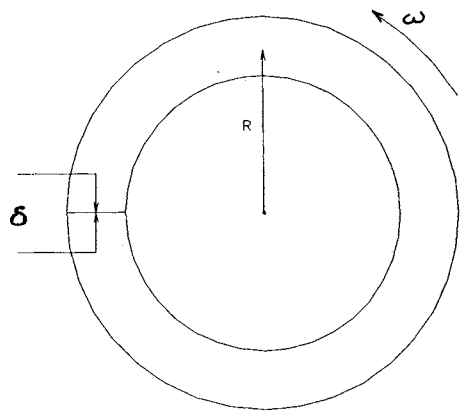


Figure 5. Schematic of a rotating blocked toroidal enclosure containing superfluid helium.

The second important observation reported by Avenel and Varoquaux was that in ^{3}He superfluid the phase difference across a very small orifice (of order 70 nanometers in length and breadth) is related to the supercurrent, I, through the hole by the simple relationship

$$I = I_c \sin \delta,$$

where I_c is the critical current for the particular orifice. This suggests the possibility of developing a ^{3}He superfluid gyroscope based on a quantum interference effect. Packard has named this device a Superfluid Helium Interference Gyroscope (SHIG) because of its similarity to a superconducting D.C. SQUID (superconducting quantum interference device).

There are some clear advantages to be gained if ^{4}He, rather than ^{3}He, can be used in a superfluid gyroscope. Working with ^{4}He has the obvious advantage that it reaches superfluidity at 2K and the phase slip events detected by Avenel and Varoquaux occurred at temperatures below about 500 mK. These temperatures can be reached with a simple dilution refrigerator, which makes it feasible to build a larger gyroscope and simplifies temperature regulation. A second potential major advantage of using ^{4}He is that it does not appear to require the manufacture of the very small orifice required for the ^{3}He SHIG. It is not yet completely clear just how the size of the orifice will affect the reproducibility of the phase slip events in the ^{4}He gyroscope, but is believed that relatively large orifices will work. Early priorities of the current research are to confirm the relative insensitivity to the orifice size in using ^{4}He and to determine if the phase slip gyroscope can achieve the same angular velocity resolution as the SHIG.

By the conclusion of the current 3-year program we expect to have a small scale working prototype superfluid gyroscope, and to have developed a set of specifications for the construction of a full scale operational gyroscope for monitoring variations in the rotation of the Earth. The construction of an operational system will take another two to three years, meaning that it will be the latter half of the 1990's before we can hope to have continuous monitoring of UT1. In the interim, VLBI will almost certainly continue to be the best available source of UT1 values.

References:

Anandan, J.H., *Journal of Physics A*, **17**, 1376, 1984.

Avenel, O.and E. Varoquaux, *Phys. Rev. Letters*, **60**, 416, 1988.

Carter, W.E., D.S. Robertson, J.E. Pettey, B.D. Tapley, B.E. Schutz, R.J. Eanes, and Miao Lufeng, *Science*, **224**, 957-961, 1984.

Carter, W.E. and D.S. Robertson, *IEEE Transactions on Geoscience and Remote Sensing*, **GE-23**, #4, 369, 1985.

Chiao, R.Y., *Phys. Rev. B*, **25**, 1655, 1982.

Packard, R.E., Proposal for the Development of a Superfluid Helium Interference Gyroscope, Private Memorandum, 1989.

Rosen, R.D., and E.A.Salstein, *J. Geophys. Res.*, **86**, 5451, 1983.

Rosen, R.D., D.A. Salstein, T.M. Eubanks, J.O. Dickey, and J.A. Steppe, *Science*, **225**, 411, 1984.

Yoder, C.F., J.G. Williams, and M.E. Parke, *J. Geophys. Res.*, **86**, 881, 1983.

Session APP-05 Number 4

The European VLBI Consortium

Giancarlo Setti
University of Bologna, Italy
European Southern Observatory, Garching bei München, F.R.G.

Introduction

Very Long Baseline Interferometry (VLBI) is a well known and extremely powerful interferometric technique which has been developed by radio astronomers. It connects radio telescopes which can be widely separated over continental or even Earth size dimensions. In this way one can achieve angular resolutions at typical cm wavelengths of 10^{-3} to 10^{-4} arcsec, by far the best spatial resolving power yet available to astronomical instruments at all wavelengths for the detailed structural studies of celestial objects. At the same time point-like, extremely distant radio sources can be located with high precision and, conversely, one can determine the rotation parameters of the Earth with respect to the inertial frame, while the possibility of detecting small changes of the baselines between the interferometer elements provides geophysicists and geodesists with an accurate tool for studying continental drifts and regional crustal movements. For this technique to be viable it is required that the phase coherence of the signals be maintained to a high degree of accuracy and this is achieved by the employment of hydrogen maser frequency standards which are now available to almost all VLBI stations.

Having reached the size of the Earth an expansion of the physical size of the interferometer is only possible by placing a VLBI element into space orbit. As will be seen there are of course extremely good scientific reasons why this costly step towards space must be undertaken. The point to keep in mind is that an orbiting radio telescope working in conjunction with the ground based VLBI network will not only permit to increase the spatial resolving power, but will provide much better imaging and improved sensitivity. Two VLBI space missions are now under way: the Japanese VLBI Space Observatory Project (VSOP) which will be launched in 1994-95 and RADIOASTRON (USSR). Since the subject of the following paper deals specifically with RADIOASTRON I shall not discuss further the questions related to these space missions. Let me just add that a further step forward in space VLBI is a new proposal, called International VLBI Satellite (IVS), which is under feasibility study at the European Space Agency. If this mission will be approved it will provide among other things an angular resolution of ~ 10 microarcsec at cm wavelengths.

The European VLBI Network (EVN)

The EVN is administered by the European consortium of VLBI composed of five member institutes and four associate members. The member institutes are: Institute for Radioastronomy CNR (Bologna, Italy), Max Planck Institute for Radioastronomy (Bonn, F.R.G.), Netherlands Foundation for Radioastronomy (Dwingeloo, The Netherlands), Nuffield Radio Astronomy Laboratories (Jodrell Bank, U.K.) and Onsala Space Observatory (Onsala, Sweden). The associate members are: Geodetic Institute (Bonn University, F.R.G.), Paris Observatory (Meudon, France), Space Research Institute (Moscow, U.S.S.R.) and Torun Radio Observatory (Torun, Poland). The institutes and the radiotelescopes of the network are summarized in Table 1, while Fig. 1 shows their geographical distribution. A distinctive characteristic of the EVN is its sensitivity due to the very large collecting surface of some of the radiotelescopes. It is rewarding and fortunate that the location of the network elements is such as to give good image quality, despite the fact that most radiotelescopes have been built to meet scientific objectives different from VLBI applications, and even before the introduction of this technique. The coverage of the u-v plane starts to deteriorate toward low latitudes where the Very Long Baseline Array (VLBA) now being completed in the U.S., and which has been optimized for VLBI work, becomes superior. Only four antennae of the network have been built primarily for VLBI, those at Medicina, Noto, Wettzell and Matera (the last two primarily for geodynamic research), but it is expected that the number of network telescopes with more and more observing time dedicated to VLBI will significantly increase in the course of this decade. There are 8 network frequencies, although not all of them available at all telescopes, comprised in the interval 327 MHz - 23 GHz.

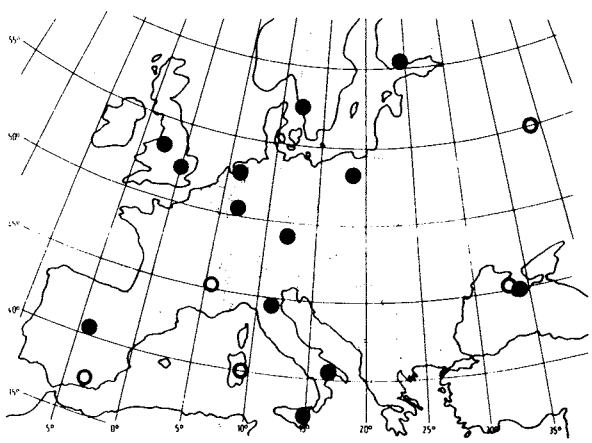


Fig. 1. - The locations of EVN telescopes. Open circles are proposed projects.

One major problem now facing the EVN consortium is the procurement of an adequate capability to process VLBI data. At present these are

Table 1

EVN Telescopes

Country	Station	Diameter (metres)	Recording mode	Atomic standard	Planned/Proposed Telescopes Site	Diameter (m)
Germany	Effelsberg	100	II, III	H		
	Wettzell	20 (geod. priority)	III	H		
Italy	Medicina	32	II, III	H	Sardinia	> 32
	Noto	32	II, III	H	Teramo	32
	Matera	20 (geod. priority)	III	H		
Netherlands	Westerbork	93 (equiv.)	II, III	H		
Sweden	Onsala	26 20	II, III	H	Onsala	70
U.K.	Jodrell Bank	76 25	II, III	R_b, H		
	Cambridge	32 (under construction)				
France	Nançay	94 (equiv.)	II	R_b	?	32
Poland	Torun	32 (under construction)	II	R_b		
U.S.S.R.	Simeiz	22	II	H		
	Suffa	70 (under construction)				

processed at the Max Planck Institute for Radioastronomy in Bonn which has built a 5-station correlator for the Mk 3 recording mode. Also a 3-station Mk 2 correlator is available at the same institution, while the 5-station Mk 2 correlator of Caltech has been transferred to the Institute for Radioastronomy in Bologna and is now operational. In view of the completion of the VLBA project and of the extension of VLBI to space, it had become clear early on that to remain scientifically competitive a new, much more powerful data processing facility is required. Consequently the EVN Consortium has submitted a proposal to the Commission of the European Communities (EEC) already in June 1988. Unfortunately up to now it has not been possible to accommodate the financing of this European project within the current EEC programmes, but discussions are still taking place. It would be a great pity if the excellent quality of the EVN network, this unique European instrument, could not be fully exploited because of lack of funding.

Table 2

Summary of VLBI Main Research Topics in Astronomy

Galactic Astronomy

Study of interstellar molecular clouds, in particular the structures and detailed motions of OH and H_2O maser sources, and the physics of star-forming regions.
Small scale structure of the interstellar medium.
Quiescent and flaring emission from radio stars.
Radio structures of interacting and X-ray binaries.

Extragalactic Astronomy

Young supernova remnants in external galaxies.
Studies of OH and H_2O extremely luminous maser sources found in association with star burst and other active galaxies.
Radio galaxies, quasars and other active galactic nuclei.
Gravitational lenses in distant objects and the dark matter problem.
The distance scale of the Universe.
The geometry of the Universe.

Astrometry

Determination to a great accuracy of the inertial reference frame.

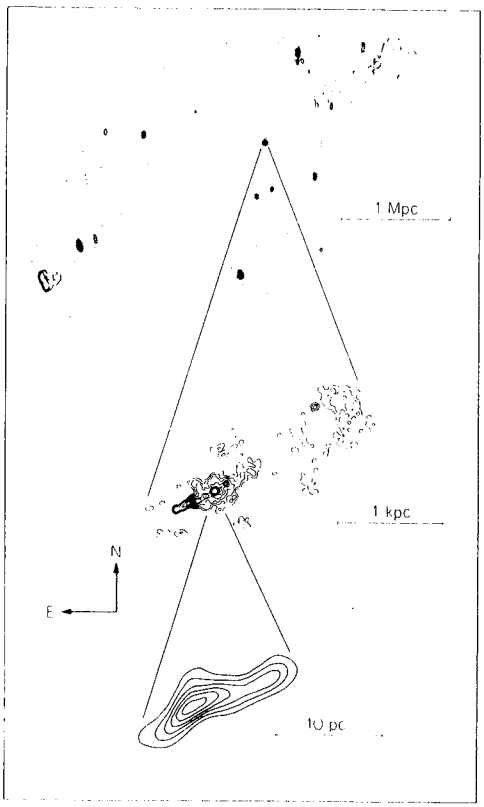


Fig. 2. - The radio structure of the giant radio galaxy 3C 236 (Barthel et al., 1985, Astron. Astrophys. **148**, 243). The angular resolution goes from just over 10 arcsec (top) to approximately 10^{-3} arcsec of the VLBI structure of the nuclear source (bottom). The point-like unresolved nuclear source has been removed from the map at the top. Note the very accurate alignment of the structures on the different angular scales which encompass four orders of magnitude.

Science with VLBI

The scientific research which can be done by using VLBI is diverse and encompasses a wide variety of fields. For the sake of illustration I have summarized in Table 2 the main research subjects which can be tackled by using VLBI in the field of astronomy. Obviously it is not possible in the limited time available to even briefly discuss all of them. As a consequence I have selected the topic of radio-galaxies and quasars as representative of the most spectacular results already obtained by VLBI.

Fig. 2 shows a composite map of the radio morphology of 3C 236, a classical double lobed radio galaxy which possesses gigantic dimensions. The separation between the outer lobes projected on the celestial sphere is larger than 10 million light years. By comparison the associated optical galaxy, which lies between the outer radio lobes and is perfectly aligned with them, has a size which is about 100 times smaller. The radio emission is due to synchrotron radiation of ultra-relativistic electrons accelerated in magnetic fields. The energy content in relativistic particles and magnetic fields is at least equivalent to the rest mass energy of 1 million stars like our Sun. The conjecture is that highly energetic phenomena are taking place in the nuclear region of the galaxy from where highly collimated jets of energy are ejected in opposite directions to feed the outer lobes. This has been dramatically confirmed by the VLBI linear scales corresponding to only several light years (bottom of Fig. 2). Because of the large distances to be traversed and of the high collimation of the jets it has also been postulated that the energy in particles and radiation must propagate outward with relativistic speeds. This has found an indirect confirmation in the studies of the radio loud quasars, which can be thought of as a related class of objects where the galactic nuclei are so bright even at optical wavelengths that they outshine the surrounding galaxy. VLBI studies of compact radio quasars have led to the astonishing discovery of apparent superluminal speeds associated with the motion transverse to the line of sight of radio blobs ejected by a central core. The most accredited interpretation of this phenomenon is that matter is ejected at relativistic speeds along a direction close to the line of sight so that the time intervals are significantly compressed. Obviously the main goal of these studies is to find out finally what is going on in the central core of these active galactic nuclei. A popular picture is that of a gravitational machine in which matter is accreted onto a massive black hole and the liberated gravitational energy is transformed and ejected along the rotation axis. The accretion disks have dimensions of ~ 1/100 light year, to be compared with sizes of no less than several light years which are typically explored with ground-based VLBI. While RADIOASTRON and VSOP will certainly improve our knowledge of these "hot" surroundings, one will probably require the angular resolution of a space mission like IVS to start probing the physical conditions of the central engine for a reasonable number of objects.

A real advancement in our knowledge for most of the topics listed in Table 2 requires the extension of VLBI to space based elements.

Concluding remarks

The VLBI technique provides a unique tool to investigate fundamental problems of modern astronomy and cosmology. The EVN represents a very powerful network that allows European astronomers to be competitive in VLBI research. To this end it is essential that a powerful state-of-the-art data processing facility is funded and made available to researchers on a very short time scale. The cost of this central facility is small compared to the large investments already made by the European countries in radiotelescopes and related instrumentation. The competitiveness of the EVN may be further strengthened if plans for new radiotelescopes get final approval. Coming back to the subject of this conference, this means that new and perhaps even more accurate frequency standards will be required.

References

The interested reader may wish to consult a recently published comprehensive book edited by M. Felli and R.E. Spencer, Very Long Baseline Interferometry (Techniques and Applications), NATO ASI Series C, Vol. 283, Kluwer Academic Publishers, 1989.

4 TH EUROPEAN FREQUENCY AND TIME FORUM NEUCHÂTEL - 13 - 14 - 15 MARCH 1990

Session APP-05 Number 5

THE "RADIOASTRON" SPACE VLBI PROJECT

Esko Valtaoja

Metsähovi Radio Research Station,
Helsinki University of Technology, Otakaari 5A, SF-02150 Espoo, Finland
and
Turku University Observatory,
Tuorla, SF-21500 Piikkiö, Finland

Abstract. The first dedicated space VLBI mission, "Radioastron-1", will produce images of quasars and other compact radio sources with a resolution ten times higher than earthbound observations. Technological, operational and scientific aspects of this Soviet-led international mission are described.

1. Introduction

The main goal of Very Long Baseline Interferometry (VLBI) is to increase the resolution in observing radio sources. The resolution obtained is λ/D, where λ is the observing wavelength and D the separation between the telescopes. Global VLBI with D ≈ 10 000 km at 22 GHz (1.3 cm) is becoming routine, giving a beamsize (angular resolution) of 0.3 mas. In March 1988, a six station VLBI experiment at 100 GHz (3 mm) resulted in the first 50 μas resolution maps of a few brightest sources (Bååth et al., preprint). First successful 230 GHz VLBI observations are expected in the near future.

The use of much shorter wavelengths is prohibited by technological limitations, such as the accuracy of present-day hydrogen maser frequency standards. Fundamental limitations for ground-based VLBI are set by the atmosphere, which absorbs virtually all radiation between one millimeter and optical wavelengths, and by the size of the Earth, which limits the length of the baselines. For still higher resolution, required for many interesting astronomical objects, one of the telescopes must be placed onboard a satellite.

For space VLBI, sensitivity becomes the chief limiting factor. To create a VLBI map of a source, a flux 3-5 times the noise level must be detected within a beam corresponding to the maximum resolution. With very long baselines, only a few of the brightest and most compact sources can be detected. The sensitivity presently achievable with a space radio telescope (typical values: antenna area ≈100 m^2, noise temperature ≈100 K, recording bandwidth ≈ 100 MHz, coherence time ≈ 100 s) limits the longest useful baselines to $<10^5...10^6$ km, if one wishes to map a significant number of sources.

The feasibility of space VLBI has been demonstrated with experiments at 2.3 GHz using the 4.9-m antennas of the geosynchronous TDRS (Tracking and Data Relay) satellites together with large ground radio telescopes (Levy et al. 1986). Four dedicated space VLBI missions have been proposed: the European QUASAT (Schilizzi 1987), which lost to other ESA missions in the final selection; the Soviet Radioastron (Kardashev and Slysh 1987), the Japanese VSOP mission (Hirabayashi 1987), with planned launch in 1995, and the second generation international space VLBI mission IVS, which is in the proposal stages (Schilizzi et al. 1989). Radioastron-1 is already in the advanced planning and constructing stages, with the earliest possible launch date in late 1993.

2. The Radioastron-1 mission

The Radioastron-1 satellite will carry to Earth orbit a deployable 10-m parabolic reflector antenna and four receivers at the standard VLBI frequencies 327 MHz, 1.665 GHz, 4.830 GHz and 22.235 GHz. It is not yet technically feasible to place the necessary hydrogen maser frequency standard onboard the satellite; instead, all the local oscillator frequencies in the satellite are derived by a two-way radio link from hydrogen masers at the ground tracking stations. The satellite carries a rubidium oscillator, which will maintain a continuous approximate on-board time during the periods when no ground stations are in contact.

The data is continuously transmitted by another down-link to the ground station, where it is recorded with standard VLBI techniques. The magnetic tapes from the satellite data down-link and from the ground radio telescopes observing simultaneously are later correlated in the normal manner.

Selecting the recording system for Radioastron has been a longstanding problem. Wideband (≈64 MHz) system is necessary for sufficient sensitivity, but the VLBI recorders in use, Mark III, Mark IIIA, VLBA, and the Canadian system, are in various degrees incompatible, and the situation is further complicated by considerations of cost, availability of sufficient number of recording units to ground stations by 1993, existence of correlation facilities, and trade restrictions. Furthermore, the Japanese VSOP satellite will use yet another standard.Since it will be impossible to have all participating ground stations use the same system as the Radioastron satellite, compatibility will probably be achieved by rerecording and copying the tapes.

The mission uses a standard Soviet astrophysical module (SPECTR), which has also been used in several other scientific missions. The total weight of the scientific payload is 1450 kg, of which the parabolic antenna takes 900 kg and the focal package holding the receivers, power supplies, LNAs, synthesizers and control units, 110 kg (Figure 1). The weight of the whole satellite, carried to orbit by a Soviet launcher, is 3-4 tons.

The orbit of Radioastron-1 is a compromise between high resolution, imaging capability, and sufficient sensitivity. The "nominal" 24-h orbit has an apogee radius of 75600 km, a perigee radius of 8400 km and an inclination of 65°; the semimajor axis points towards the ecliptic pole. This orbit gives a formal resolution of

30μas with the highest observing frequency (although smaller details can be discerned with sufficient S/N ratio), providing a tenfold increase in resolution as compared with ground-based VLBI.

Alternative orbits are also still being considered, mainly to increase the resolution. One interesting possibility is that the satellite could be lifted from its initial 1-day orbit to a higher 4-day orbit with apogee ≈ 200 000 km. However, the sensitivity of Radioastron may not be sufficient to take full advantage of so large an orbit. The quality of the maps also decreases with higher orbits.

The planned minimum lifetime of the satellite is two years, of which time at least 50% will hopefully be available for VLBI observations. The exact percentage will depend on the extent of support from ground stations.

3. International cooperation and mission management

The Radioastron program was initiated by N. Kardashev of the Space Research Institute (IKI) of the U.S.S.R. Academy of Sciences, who also continues to lead the project. The program has expanded into a broad international collaboration, in which both the magnitude and the depth of involvement of Western participants in a Soviet space mission from the very beginning is probably unprecedented: scientists from over ten countries are constructing the instruments, planning the mission profile, or securing ground support for Radioastron-1.

The Soviet Union will provide the satellite and most of the hardware. Several other countries will contribute to the scientific payload. The 327 GHz receiver onboard the satellite is being built by India (TIFR) and the USSR, the 1.6 GHz receiver by Australia (CSIRO), the 5 GHz receiver by European VLBI Consortium, and the 22 GHz receiver by Finland (Helsinki University of Technology). Other technical or scientific cooperation includes institutes from Bulgaria, Canada, Czechoslovakia, Hungary, Japan, United States and West Germany. The agreements have usually been made directly between the participating institutes and IKI.

Ground support from radio telescopes is particularly important for a space VLBI mission, since the quality of the resulting map depends directly on the number of participating telescopes. Although the three primary VLBI and telemetry stations will be in USSR (Evpatoria, Suffa and Ussurijsk), foreign tracking stations are needed when Soviet telescopes either are not in operation due to other commitments, or when the satellite is invisible from the Soviet Union. The scarcity of suitable radio telescopes in the Southern hemisphere remains a problem.

At present, all the major components of the mission, the spacecraft system, the science payload, and the ground systems, are progressing according to a schedule aiming for a late 1993 launch. The project development is followed in regular semiannual meetings.

The management of Radioastron will be through Radioastron International Scientific Council, with members selected from participating countries and institutes. Although all the details have not been worked out yet, it seems likely that after an initial period of a few months, during which a core science program is executed, observing time will be allocated by an international program committee.

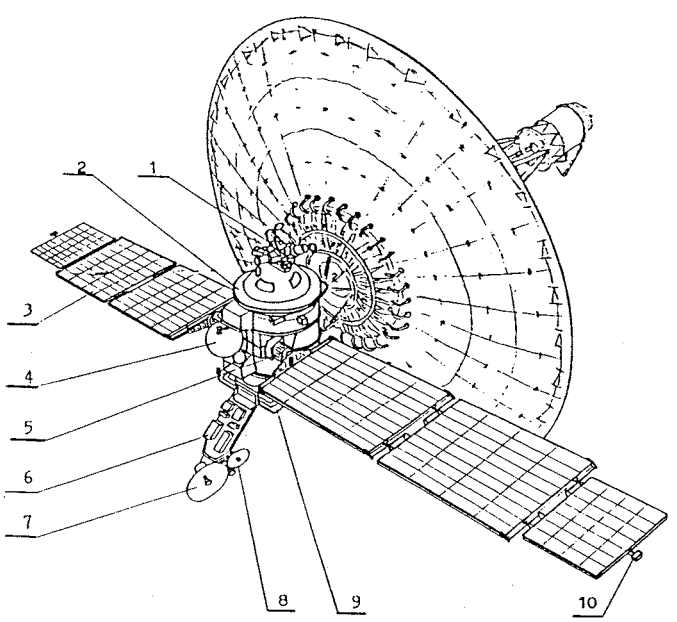


Figure 1. The Radioastron satellite (SPECTR astrophysical module). The diameter of the radio telescope antenna is 10 meters. The receivers and other scientific payload are in the focal package. 1 – Optical/Electronic unit, 2 – Equipment compartment, 3 – Solar panels, 4 – Propulsion subsystem, 5 – Solar panels yoke, 6 – Platform for the high bit rate communications subsystem, 7 – Antenna for scientific data transmission, 8 – Antenna for reference signal reception, 9 – Radiator-cooler, 10 – Sun sensor for solar panel pointing.

4. Science with Radioastron

The most important goal for Radioastron-1, as for all proposed space VLBI missions, will be extragalactic radio sources. Earthbound observations have given us a general framework for models of quasars and other active galactic nuclei: in the center of a galaxy hides a supermassive black hole, which swallows matter from a surrounding accretion disk. Part of the gravitational energy of the infalling matter is released, somehow collimated into a narrow relativistic jet, and transported to large distances from the core. Most of the evidence, however, remains indirect or purely theoretical: even in the best VLBI maps, the mysterious cores of quasars remain unresolved, featureless points.

The tenfold increase in resolution provided by Radioastron-1 may make possible direct imaging of the central regions of active galaxies; theoretical estimates place the angular sizes of the accretion disks in nearby active galactic nuclei close to the resolution limit. Relativistic jets, apparently superluminal motions, growth and evolution of shocks in the jet, and the possible existence of a quiescent black hole in the center of our own galaxy can all be studied with Radioastron.

The impact of tenfold increase in resolution is well demonstrated in the imaging simulations shown in Figure 2. A milliarcsecond-size model source with smaller internal structure remains an indistinct blob using an "Earth only" VLBI array. With Radioastron, most of the structural details are recovered, and if the Japanese VSOP satellite can join the observations, the image is virtually identical to the original model.

At 22 GHz the sensitivity of Radioastron is such that a flux of about 0.2 Jy within the 30 μas beam (corresponding to a brightness temperature ≈4 10^{11} K) is required for mapping the source with maximum resolution. There are perhaps 200 quasars bright enough for such observations; from these, the Radioastron Science Objectives Committee is selecting a smaller sample of the most interesting sources for the core science program.

In numerous other fields of study the microarcsecond resolution provided by space VLBI will have a decisive impact. The Radioastron Science Objectives Committee has listed a number of such goals:

- flare activity of nearby stars
- measurements of distance to pulsars and other compact galactic sources through their parallax and proper motion
- detection of envelopes of ejected relativistic particles near pulsars
- structure of radio emission regions in galactic relativistic objects (neutron star or black hole binary systems)
- structure and dynamics of maser radio sources in regions of star formation
- turbulent structure of interstellar medium through the scintillations of compact radio sources
- cosmological evolution of compact extragalactic radio sources
- determination of fundamental cosmological parameters and of the nature of hidden mass

For astrometry and geodesy with Radioastron, a very precise (≈1 m) orbit determination is required. This may be achievable with an on-board PRARE system.

The probable distribution of observing time for different goals has been estimated to be the following:

(1) Brightness temperature survey (quasars)	30%
(2) Continuum imaging (quasars)	20%
(3) Monitoring of superluminal motion (quasars)	20%

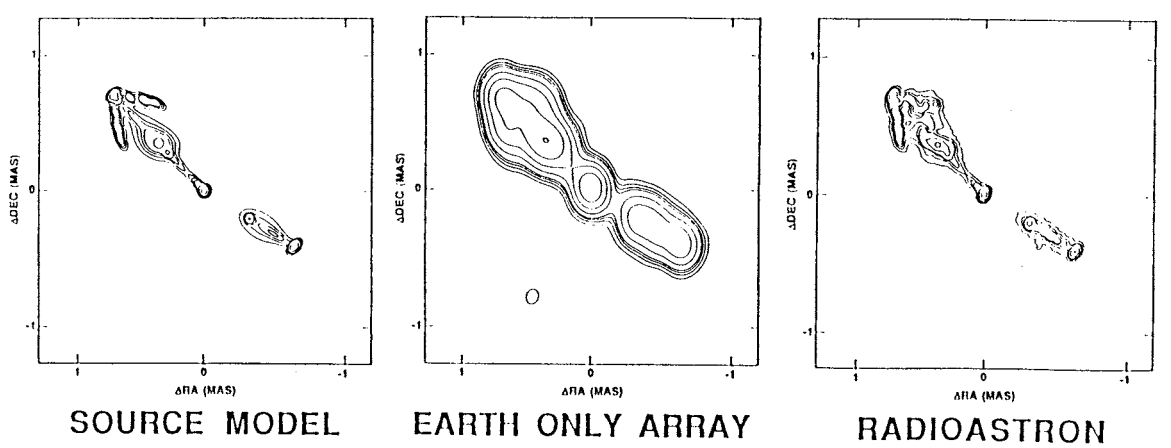


Figure 2. The influence of tenfold increase in resolution. Left: the milliarcsecond model source. Center: a simulated VLBI map of the model source at 22 GHz using only Earth-based radio telescopes. Right: a simulated map using the Radioastron satellite together with ground VLBI arrays. From the report "*The Imaging Potential of RADIOASTRON*" by D.W. Murphy, presented at the 8th Radioastron Review Meeting in Green Bank, USA, May 1989.

(4) Mapping of water maser regions	5%
(5) Monitoring of water maser motions	5%
(6) Mapping of OH-maser regions	5%
(7) Interstellar scattering	5%
(8) Observations of pulsars	5%
(9) Astrometric and geodetic measurements	5%

5. Future of space VLBI

Radioastron-1 represents the first pioneering effort in space radio astronomy, to be followed by other experiments already in planning stages. These future radio telescopes in space will be larger, have wider bandwidths and more sensitive receivers, use higher frequencies, and be at ever-increasing distances from the Earth. One central problem is how to increase the coherence time of the interferometer, which is potentially the best way to increase the sensitivity. Possibilities include sub-meter accuracy in the orbit determination and an onboard frequency standard, either a hydrogen maser or some better future alternative; for submillimeter wavelengths the present-day masers will not be accurate enough.

The Soviet Radioastron program includes plans for placing several 30-m telescopes, working up to 230 GHz, at distances up to 1.5 million km from the Earth by 2005, and a huge 400-m diameter telescope in space, larger than any on Earth. Among other things, such space arrays would make possible to obtain true three-dimensional images of objects in our own galaxy, and to determine the shape, size and future of our Universe.

The large costs of space missions may also result in resources of space nations being pooled into common projects. One such project, a proposed follow-up for Radioastron-1 and VSOP missions, with tenfold sensitivity and increased resolution, is presently in the ESA Assessment Study phase: International VLBI Satellite (IVS). Joint participation of ESA, Soviet Union, Japan and United States would make possible an effort too expensive for any single space agency.

References

Bååth, L., et al., preprint (1990)

Hirabayashi, H., *The Impact of VLBI on Astrophysics and Geophysics*, eds. M.J. Reid and J.M. Moran, Kluwer Academic Publishers, Dordrecht, p. 449 (1988)

Kardashev, N.S., Slysh, V.I., *op.cit.*, p. 433 (1988)

Levy, G.S., et al., Astrophys. J. **336**, 1098 (1989)

Schilizzi, R., *op.cit.*, p. 441 (1988)

Schilizzi, R., et al., *A proposal for IVS (International VLBI Satellite)*, submitted to ESA on 30 November, 1989

Postdeadline posters

Session Postdeadline poster 1

Some Results on the Line Q Degradation in Hydrogen Masers

Laurent-Guy Bernier
Giovanni Busca
Observatoire Cantonal, Neuchâtel, Switzerland

1.0 Introduction: A Pathological Case

As a rule, the Teflon coating of hydrogen masers ages very slowly. If the coating is good, the storage bulb does not need a re-coating before many years of operation. There are some pathological cases, however, when the coating process is unsuccessful without failing completely. The Teflon coating passes the "water drop" test and the initial value of the line Q is reasonably high. However the line Q degrades excessively fast causing a fast amplitude decay of the signal. A similar case was reported in [1]. This paper reports observations of a maser behavior under this kind of situation.

2.0 The Effect of Air on a Bad Teflon Coating

In the past, we observed once that any intervention on a certain maser would temporarily cure its amplitude decay problem, whatever the intervention was. It was finally suspected that the simple fact of exposing the storage bulb to air, during the intervention, would temporarily improve the quality factor. More recently this past observation lead us to deliberately try the experiment on a maser that suffered a fast degradation of the line Q. The only intervention made on the maser between the 2 measurements of the quality factor reported below consisted of filling the maser with air and then pumping down again. The quality factor Q_L was 1.20×10^9 before and 1.44×10^9 after exposure to air.

3.0 The Time Behavior of a Bad Teflon Coating

The effect of a bad Teflon coating on the maser behavior was observed for several months. Figure 1 shows the evolution of the line Q starting just after the re-coating of the storage bulb. The origin of the time axis is the first day of oscillation. It can be observed that there is a decelerating degradation, during the first days, and then the degradation rate stabilizes. Figure 2 shows a linear regression of the linear part of the line Q degradation. The equation of the linear regression is as follows: $Q_L = 1.9528\times10^9 - 3.7052\times10^6$ t [day]. The relative degradation rate is 0.19 [%/day].

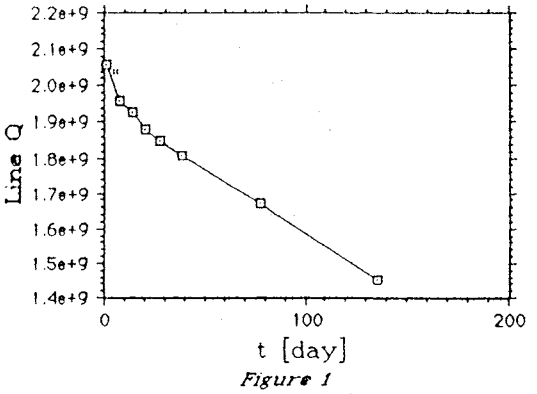


Figure 1
Line Q versus Time

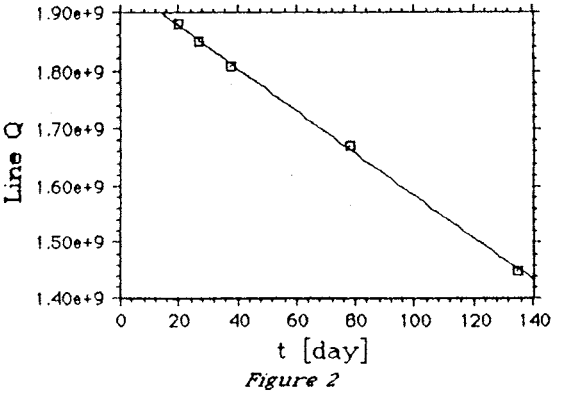


Figure 2
Linear Part of Line Q versus Time

Figure 3 shows the amplitude of the maser signal versus time. Again there is a decelerating decay, at first, and then the rate of decay stabilizes. The amplitude steps are due to hydrogen pressure resettings. Figure 4 shows the linear part of the curve with the amplitude steps corrected by translation of the constant hydrogen pressure segments. A linear regression of the data points yields: A [V] = 11.741 - 4.0735×10^{-2} t [day]. The relative rate of decay is 0.35 [%/day]. Figure 5 shows the normalized frequency $y(t) = \delta\nu(t)/\nu_0$. The beat frequency $\delta\nu(t)$ was measured by comparing the maser frequency ν_0 with the frequency of a reference maser.

A periodical spin-exchange tuning of the cavity showed that the microwave cavity was not drifting. The big frequency step on the curve is due to the initial spin-exchange tuning of the microwave cavity.

The small steps are due to the imperfect resetting of the varactor voltage after each verification of the correct spin-exchange tuning. Figure 6 shows the y versus t curve with the frequency steps corrected by translation of the constant cavity tuning segments. The residual frequency change is approximately logarithmic. The analytic curve shown on the figure is given by: $y(t) = -1.7693 \times 10^{-12} + 2.8426$ LOG(t [day]). If the frequency change were due exclusively to the line Q degradation, one would expect a linear drift. Thus the observed non-linear drift seems to indicate a wall shift evolution associated with the Teflon degradation.

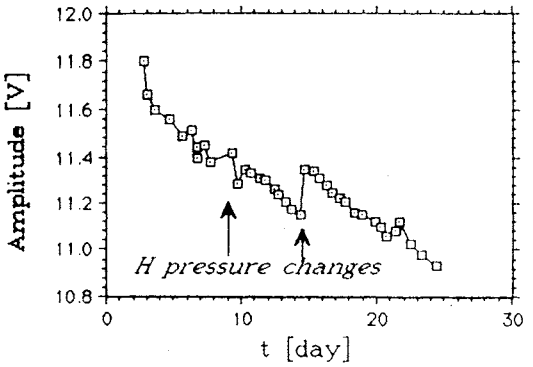


Figure 3
Signal Amplitude versus Time

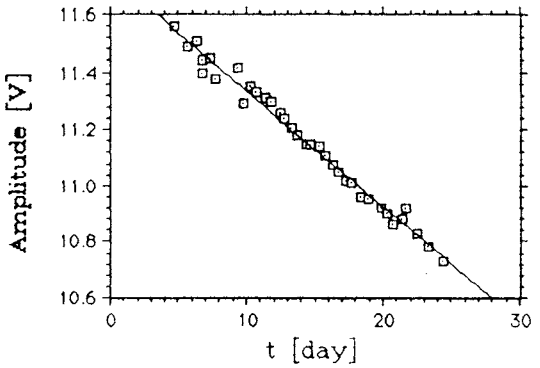


Figure 4
Signal Amplitude versus Time
Corrected for Constant Hydrogen Pressure

4.0 Conclusion: What Makes a Good Teflon Coating ?

The storage bulb of the maser described here was re-coated again and put back into operation with the same hydrogen dissociator as before. The signal is now higher (12.4 [V]) for a smaller hydrogen flux (V_p = 6,5 [V]). The quality factor is now $Q_L = 2.5 \times 10^9$ for V_p = 6,5 [V] and there is no measurable line Q degradation or amplitude decay up to now. This shows that this particular hydrogen dissociator has a stable efficiency and that the line Q degradation was entirely responsible for the amplitude decay observed before. We are now in the process of evaluating the value of ESR (Electron Spin Resonance) as a technique for testing Teflon coating properties. Preliminary results show that the amplitude of the ESR **is a function of the free spins density in the tested material, varies considerably between samples made from different Teflon batches.**

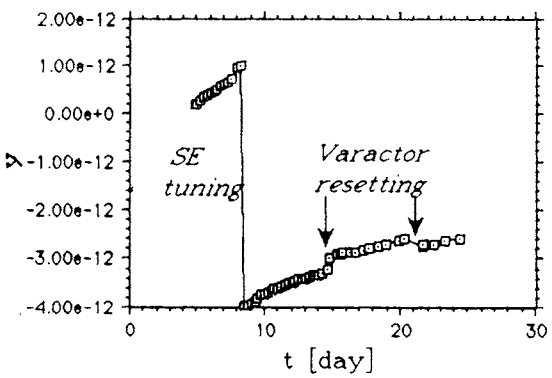


Figure 5
y versus Time

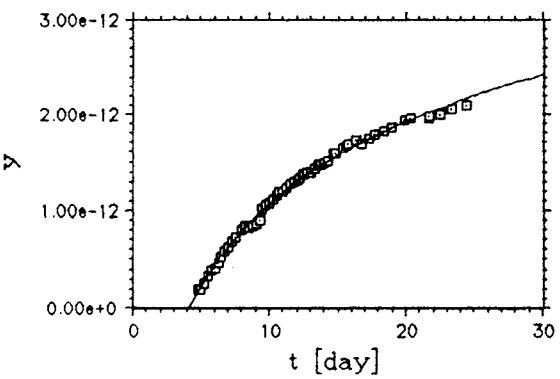


Figure 6
y versus Time
Corrected for Constant Cavity Tuning

From a practical point of view, we believe that it is equally important to control both the properties of the initial Teflon coating solution as well as the conditions of the drying and curing process if reliable coatings are to be produced. On the other hand, it is clear that the "water drop" test does not mean much as a measurement of a Teflon coating actual performance. Therefore further investigation about both the basic Teflon coating properties and the techniques that could be used in order to check those properties must be pursued.

[1] Mattison E. M., Vessot R.F.C., Bain C., Wasserman S., Whiteside G. "Surface Interaction of Atomic Hydrogen with Teflon," Proc. 41st Annual Symposium on Frequency Control, Philadelphia, May 1987, pp. 95-98.

Acknowledgements

We want to thank Prof. Schweiger of the Laboratorium für Physicalisch Chemie, ETH Zürich, for his help in our ESR experiments.

Session Postdeadline poster 2

THE TIME AND FREQUENCY SYSTEM AT THE MATERA SPACE GEODESY CENTER

A. Cenci, L. Garramone.
Telespazio SpA, Via Tiburtina 965,Rome, Italy.

G. Bianco, B. Pernice, G. Sylos Labini.
ASI, Space Geodesy Center, Matera, Italy.

1. INTRODUCTION

The Matera Space Geodesy Center started its activity in 1983 with the installation of a Satellite Laser Ranging System (SLR), under an agreement between National Aeronautics and Space Administration (NASA), Italian National Space Plan (CNR/PSN, now Italian Space Agency - ASI) and Regione Basilicata.
The center is devoted to Geodesy and Geodynamics, such as study of crustal deformations of the Earth, plate tectonics, polar motion and earth rotation, as well as to improve the knowledge of the earth gravity field and earth/ocean tides.
The Matera SLR station is operated by Telespazio under ASI management.
A new Very Long Baseline Interferometry station (VLBI), wich includes a 20 m antenna and the MARK III Data acquisition Terminal, has been completed. A new Time and Frequency system is going to be installed, in order to meet the requirements for both SLR and VLBI measurements. This system, realized by Telespazio in cooperation with Politecnico di Torino, could also be used in the future as prime reference for a Time and Frequency distribution station, to be possibly installed near Matera. Furthermore, in connection with other Time Laboratories (e.g. IEN - Istituto Elettrotecnico Nazionale), it will also contribute to maintain the Italian National Time Scale.
The characteristics of this system are described in this report, starting with a general description of the SLR and VLBI techniques and they requirements in terms of frequency and time.
An overview of the system configuration is given, together with a description of the special room dedicated to host the frequency standards.

2. THE NEW TIME AND FREQUENCY SYSTEM

The Time and Frequency system at Matera was based up to fall 1984 on the use of two Rubidium vapour standards to generate the local time scale (average stability $\approx 3\times10^{-12}$, T>1 day) and a LORAN-C receiver to keep it synchronized to UTC at microsecond level. Since 1985, the main oscillator has been replaced with a Cesium beam HP 5061A standard (average stability $\approx 3\times10^{-13}$, T>1 day) and since 1986, a GPS FTS 8400 receiver has been installed /1/. For SLR, timing of the epoch of observation must be an accuracy that is comparable with the range measurement /2/. The average velocity of a satellite (e.g. LAGEOS) is of the order of 5 km/s, so that absolute timing of better than 1 microsecond is required to fix the sa-

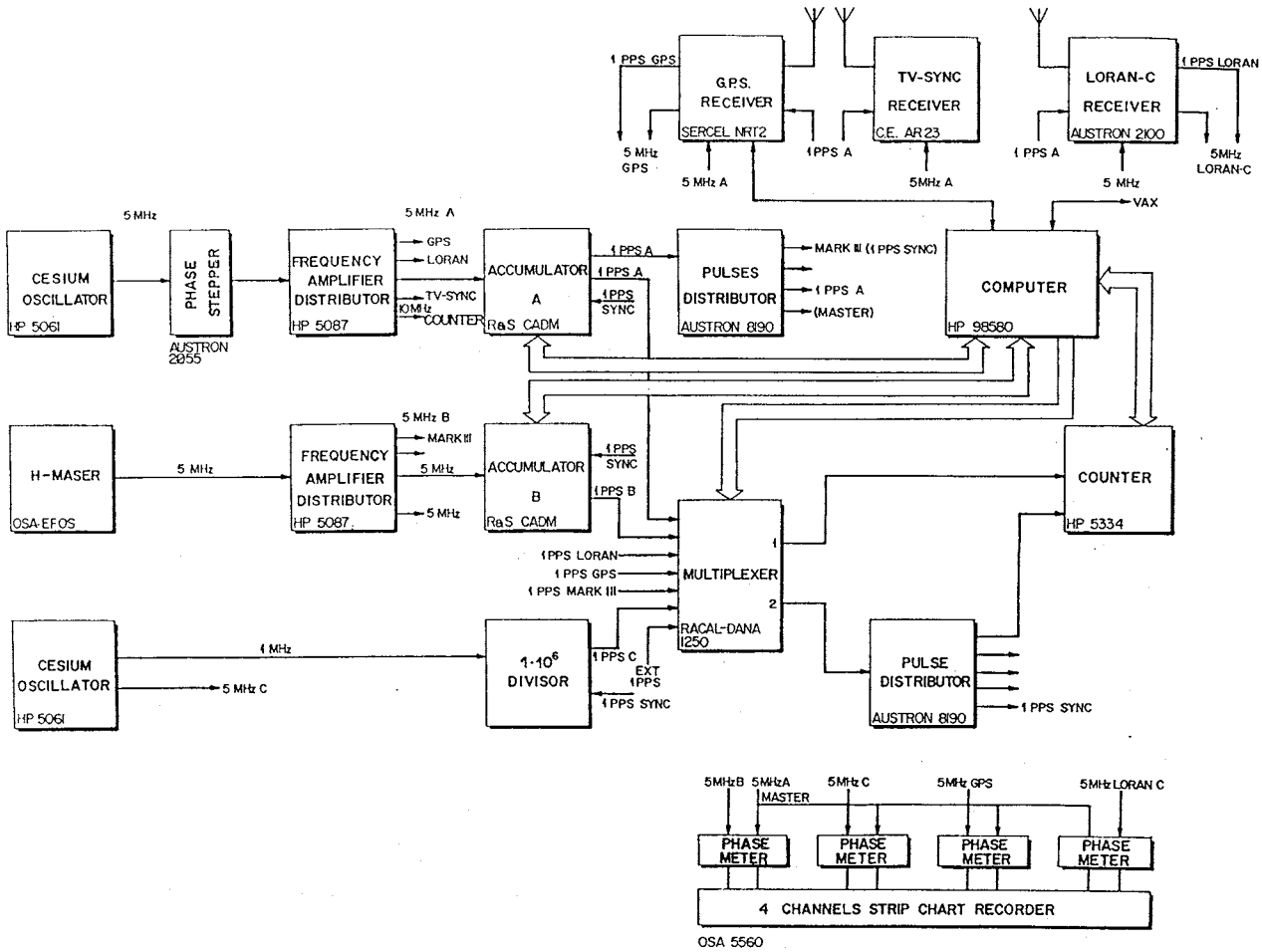


Fig. 1 The Frequency and Time System

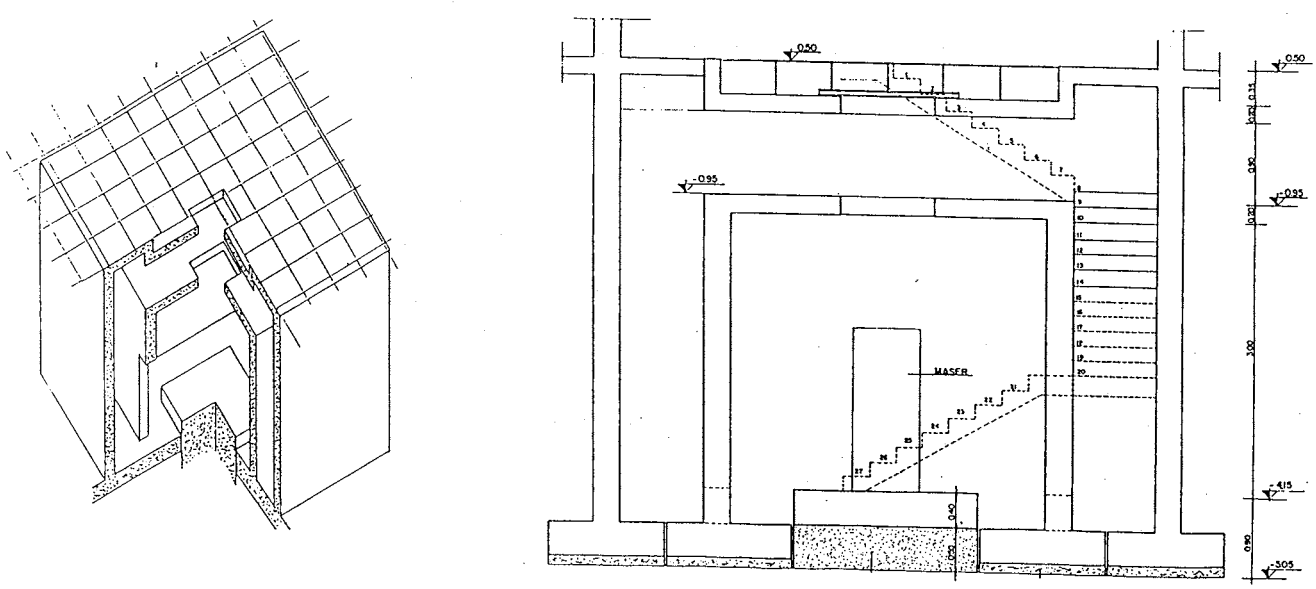


Fig. 2 Axonometry and cross-section of the frequency standards room

tellite in its orbit with,say, 5 mm accuracy. For VLBI /3/, let us suppose a duration of 1000 seconds in reception of signals from a given radio source at, say, 10 GHz frequency. To avoid ambiguity errors, the maximum phase rotation must not exceede 90°, i.e. 25 picoseconds, corresponding to a maximum variation of the frequency used for the synthesis $\Delta f/f = 2.5 \times 10^{-14}$. The stability of each local oscillator must be therefore in the order of 1×10^{-14} for periods of 1000 seconds. In order to meet these requirements, a new time and frequency system is going to be installed at Matera. The overall block diagram of this system is shown in fig.1. It consists of three frequency sources and three independent clock chains. The OSA EFOS-8 H-MASER is used as frequency source for VLBI. This system shows excellent performance in terms of stability ($\Delta f/f < 2 \times 10^{-15}$, $1000 > \tau > 10000$ seconds), spectral purity (>110dB, 1Hz), temperature sensitivity ($<2 \times 10^{-14}$/°C). The prime time scale is generated by a HP 5061/opt 004 Cesium standard (a second Cesium is intended as back-up). Synchronization to UTC to better than 100 nanoseconds is guaranteed by the use of a GPS time receiver. A dedicated computer system manages all the control and monitoring operations, such as clock comparisons, time and/or frequency adjustement, synchronization, environmental control, etc. Absolute continuity of operations is guaranteed by the main power inverter. An additional battery back-up is dedicated to the frequency standards. To minimize the effect of environmental conditions, in particular for the H-MASER, the frequency standards are hosted in a dedicated room, as described in the next section.

3. THE FREQUENCY STANDARD ROOM

A special underground room has been built into rock to host the frequency standards (see fig.2). It consist of a 4x4x3.4 meter "box" located 5 meters (floor) below the equipment room. Particular criteria were adopted in the construction of this room, following the suggestions of Prof. Leschiutta. An 1 meter air space (aisle) around the room and above the ceiling, with double access door, has been designed to improve the termal insulation. A 2x2x0.4 meter concrete pedestal on the floor, insulated from the rest of the building to reduce vibrations, will host the standards, allowing also easy maintenance. Active air conditioning is present in the above equipment room. With no active air conditioning neither in the room nor in the air space, temperature variations less than 1°C/month have been measured (with the room empty) from july 1989 to january 1990. Diurnal temperature variations are very well absorved at a level not measurable with the instrument used (≈0.1°C). The room is equipped with temperature and humidity sensors and monitored by the computer.

4. REFERENCES

1. A. Cenci: Matera Laser Station Performance Analysis from July 1986 to June 1987, Telespazio, CNR/PSN contract,1987.
2. Proceedings of the 6th International Workshop on Laser Ranging Instrumentation, Antibes, 1986.
3. J.B. Thomas: An analysis of Long Baseline Interferometry, JPL Technical Report 32-1526, 1972 (Vol. VII, VIII, XVI).

Session Postdeadline poster 3

PHASE ERRORS DISTRIBUTION OBTAINED FOR A CLASS OF CLOCKS.

M.-O. HONGLER and C.V. QUACH thi
Dpt. of Theoretical Physics
University of Geneva
CH-1211 GENEVA 4 (Switzerland).

The mathematical model of a clock is in fact the representation of the time process $\theta(t)$ observed at the output of a clock. A clock can generally be viewed as a dynamical device which consists of an oscillator and a counter. The oscillator is used to generate a cyclic motion whose amplitude is independent on the initial conditions and which is generically stable under weak perturbations.

Clocks are therefore non-linear, dissipative (dissipation is required for the stability requirement), self-oscillating devices. Of course, actual clocks operate in various environments and are therefore always subject to random disturbances. This external noise introduce errors and the time process $\theta(t)$ is in fact a stochastic process. The fluctuations about the average of a time process are essential to characterize its quality.

In this paper, we discuss the stochastic process $\theta(t)$ for a rather general class of clocks. The devices we study are characterized by dynamical equations which admit a limit cycle solution; (strictly speaking, the limit cycle occurs only in the noiseless limit).

The fluctuations of the environment are idealized in terms of Gaussian White Noises. It is a remarkable feature that for such a class of dynamics which belongs to the diffusion processes, the time process $\theta(t)$ is spreaded around the ideal time according to a Cauchy distribution and hence no finite moments exists. While this result crucially depends on the use of White Gaussian Noises, it is however qualitatively not much sensitive on the detailed nature of the drift fields provided that the system admits a finite stationary probability density (i.e.invariant measure).

We introduce the class of Langevin Eqs. in the form :

$$dx = (H_y - aV'(H)H_x)dt + \mu a\, d\beta_{1,t}, \tag{1a}$$

$$dy = (-H_x - bV'(H)H_y)dt + \mu b\, d\beta_{2,t}, \tag{1b}$$

where μ, a, b are constants,

$$V'(H) = \frac{d}{dH}V(H). \tag{2}$$

and

$$lim_{h\to\infty}V(H) = \infty, \tag{3}$$

and $d\beta_{\nu,t}$ are White Gaussian Noise processes ($\nu = 1, 2$), with the properties :

$$\langle d\beta_{\nu,t}\rangle = 0, \tag{4a}$$

and

$$\langle d\beta_{\nu,t}\, d\beta_{\mu,t'}\rangle = \delta_{\nu,\mu}\, \delta(t-t'). \tag{4b}$$

We require the function H = H(x, y) to be such that H(x, y) = E defines a closed curve L_E in R^2 and $L_E \cap L_F = \emptyset\ for E \neq F$

The Langevin Eqs. (1a,b) belong to the class of mixed canonical-dissipative dynamics discussed in [1]. The canonical evolution is governed by the Hamiltonian drift part, namely $(H_y, -H_x)$ and the dissipative forces are described by the drift :
$(-aV'(H)H_x, -bV'(H)H_y)$. The Fokker-Planck Equation for the transition probability density $P(x, y, t|x_0, y_0, t_0)$ of the diffusion process reads as :

$$\frac{\partial}{\partial t}P(x, y, t \mid x_0, y_0, t_0) = F\ P(x, y, t \mid x_0, y_0, t_0), \tag{5}$$

where the Fokker-Planck operator F in this case is :

$$F(.) = -\frac{\partial}{\partial x}([H_y - aV'(H)H_x](.)) - \frac{\partial}{\partial y}([-H_x - bV'(H)H_y](.))$$
$$+\frac{1}{2}\frac{\partial^2}{\partial x^2}(.) + \frac{1}{2}\frac{\partial}{\partial y^2}(.), \tag{6}$$

The stationary probability density $P_s(x,y)$ can be found in the explicit form :

$$P_S(x,y) = N^{-1} exp(-\frac{2}{\mu^2} V(H(x,y))). \tag{7}$$

The potential V(H) is chosen such that the normalization constant N exists, (the necessary condition is already given by the asymptotic property of V(H) Eq. (3)). From the class of models Eqs. (1a,b), we shall now construct a class of clocks by assuming that we have :

$$\frac{d}{dH} V(H)_{H=E_0} = 0, \tag{8}$$

$$\frac{d^2}{dH^2} V(H)_{H=E_0} \geq 0. \tag{9}$$

Indeed, when V(H) in Eqs.(1a, b) fulfills the conditions Eqs. (8) and (9), then in the limit $\mu = 0$, the system admits a stable limit cycle solution L_0 defined by the equation $H(x,y) = E_0$; [2].

Let us now introduce the polar coordinates :

$$r(t) = \sqrt{x^2(t) + y^2(t)}, \tag{10a}$$

and

$$\theta(t) = Arctg(\frac{y(t)}{x(t)}). \tag{10b}$$

The time process $\theta(t)$ is defined as the winding angle of the diffusion procees defined by Eqs.(1a, b) and Eqs. (8) and (9).

With the above definitions, we can establish the following result [3,4] :

Theorem. For weak noise, namely $\mu \approx 0$ and for the initial conditions distributed according to $P_S(x_0, y_0)$, we have for asymptotically large time :

$$P(\frac{\theta(t)}{t}) \in [x, x+dx]) = \frac{\eta dx}{\eta^2 + (x-\rho)^2}, \tag{11}$$

where the constants η and ρ read as :

$$\eta = -2\sqrt{\pi}(a+b)\frac{\mu}{[L_0]} exp(-\frac{2}{\mu^2}[V(0) - V(E_0)])\sqrt{V''(E_0)}, \tag{12}$$

where $[L_0]$ denotes the length of the limit cycle L_0 and

$$\rho = \int_0^{2\pi} d\theta \, (G(H)\mid_{r=\infty}) - G(H)\mid_{r=0}), \tag{13}$$

with :

$$G(H) = N^{-1} \int_0^H exp(-\frac{2}{\mu^2} V(z)) \, dz. \tag{14}$$

It is easy to observe that in the noiseless limit $\mu = 0$, we obtain $\eta = 0$ in Eq. (12). Hence Eq. (11) reduces to the simplest form :

$$P(\frac{\theta(t)}{t}) \in [x, x+dx]) = \delta(x-\rho)\, dx, \tag{15}$$

and therefore we have :

$$\theta(t) = \rho t \tag{16}$$

which characterizes a perfect clock.

A Cauchy distribution is expected when the fluctuations are Gaussian White Noises and the drift fields yield the existence of a stationary probability density in the phase space, [4]. Howewer, it is remarkable that for the dynamical systems proposed here, we are able to give the explicit characterization of the waist of the distribution in terms of the various parameters which enter into the dynamical equations of the motion. As a consequence, we observe that, the larger the size of the limit cycle and the deeper the valley which control the dissipation near this cycle, the more accurate are the clocks. These observations enable us to identify η as a quality factor.

REFERENCES.

[1] C. P. ENZ. PHYSICA 89*A*, (1977), 1.

[2] A. ANDRONOV, E.A. LEONTOVICH, I. GORDON and A.G. MAIER "Theory of bifurcations of dynamical systems on the plane." Reprinted edition, Halsted Press. (1973).

[3] M.-O. HONGLER and C.V. QUACH thi. "On the time process delivered by a class of clocks". Preprint UGVA-DPT/12-646 (Univ. of Geneva) 1989. Submitted to J. Math. Phys.

[4] G. GONG and M. QUIAN." On the winding of stationary diffusions." Preprint Dpt. of Probability and Statistics, Beijing University. To appear in Adv in Math , China (1990).

Session Postdeadline poster 4

SWEEPING OF THIN AT AND SC CUT QUARTZ CRYSTALS UNDER VACUUM AND AIR ATMOSPHERE

J.P. BACHHEIMER and P. PALLEAU
Université Joseph Fourier, Laboratoire de Spectrométrie Physique (associé au CNRS), Boite postale n° 87 - 38402 Saint-Martin-d'Hères Cedex - France

Abstract : A series of thin (~ 1 mm) AT (medium quality) and SC (good quality) cut crystals was electrolysed (swept) at different temperatures (500-550-650°C) and electric fields under air and vacuum atmosphere. We present the results of our measurements of sweeping currents, activation energies, infrared spectra, ohmicity tests. Sweeping process is effective on the AT cuts in air (best 500-3500 V/cm) and vacuum (at 650° it changes the I.R. medium quality spectra in a good one) and also for the SC cuts in air but it is questionable for the SC cuts in vacuum in α phase. Current-tension curves (ohmicity tests) are highly non linear showing possible strong space charge effects.

Introduction : The electrolysis (sweeping) of quartz is an industrial high temperature (500°C) process which is used to improve the quality of piezoelectric devices (radiation hardness) by means of both displacement and exchange impurities obtained with an electric field applied along the Z axis of thick (1-2 cm) synthetic bar. The most important (i.e. recognize up to now for resonators performance) impurities present in synthetic quartz are alkali ions (M^+) associated to aluminium (Al-M^+ center) which exists as a substitution for silicon and hydrogen related as grown OH defects. Schematically sweeping in an ambiant air atmosphere replaces the alkali ions by hydrogens to form Al-OH (e_1 - e_2 peaks in I.R. spectra) while sweeping in vacuum form Al-holes (1). Ohmicity tests and activation energies can be useful (2) to appreciate the achievement of the process which lasts usually several days. Indeed we hope thus a linear relationship between current and tension which can be observed with thick samples (3). In the final stage of the process we can hope an activation energy for the conduction by proton of ~ 1.6 - 2,1 eV (2) instead of the conduction by alkali impurities which gives values near 1,1 eV (4). We present the results(sweeping currents, activation energies, infrared spectra, ohmicity tests) obtained from electrolysis of thin AT and SC cuts with air and vacuum (10^{-6} mm Hg).

Experimental : The samples are AT and SC cuts (thickness 0,96 mm, area ~ 3 cm^2) crystals made of two synthetic bars (medium quality for AT, good quality for SC). Table I gives the average infrared absorption coefficients before sweepings.

Table I :

Samples	α3500	α3410
AT(n° 1...10)	0,08	0,145
SC(n°11...20)	0,02	0,080

Electric fiels were applied by means of Pt pressed foil electrodes (gold evaporated electrodes for n° 4) on the optically polished faces which are cleaned with H.F. They were fully applied (in ~1 minute) once the temperature was reached but for 650°C a small field (a few V/cm) was maintained since the α phase to locate the α-β transition.

Results : Figures 1, 2, 3, 4 show the sweeping currents. They are nearly the same for all the cases at 500°C (except n° 10 which displayed a special behavior : the dashed line is an average inside a factor 2) and the final current is close to that obtained with thick (~ 1 cm//Z) vacuum sample (5). At 550° initial currents are close and final current are smaller under vacuum than under air. At 650°C the behaviors are more scattered perhaps due to the fact that low electric fields where applied while heating. An overall comparison with thick samples (5) shows clearly a gain of about an order of magnitude on the time scale.

Table II gives the activation energies after sweeping : they are pratically between 1,5 and 2 eV except for the SC cuts under vacuum in α phase. Infrared spectra (room temperature) are given by figures 5-6-7-8. For the AT cuts in α phase, all the samples display the e_2 peak which proves the sweeping process but only the electrolysis in air at 550°C-3500 V/cm improves the α_{3500} and α_{3410} absorption coefficients. At 650°C vacuum sweeping changes the medium quality spectrum into a spectrum similar to that from the best quality crystal, a result already found with thick Z samples (5). For the SC cuts in air (figure 7) the sweeping action is also evident (e_2 peak) but in vacuum (figure 8) no peak appears in α phase. It weakly appears at last at 650°C where we also find an increase of activation energy. Therefore it is likely that the sweeping process takes place. At 500° - 1500 V/cm the low (1,24eV) activation energy favors a major conduction by alkali impurities and thus a poor sweeping efficiency.

Ohmicity tests before cooling were not performed systematically. Figure 9 gives some examples which show a highly non linear behaviour particularly under vacuum where the curve 18 appears which looks like electrical conduction by dissociation and injection in dielectric liquids (6)). In contrary to air sweeping (where protons can be supplied by water vapor), both low as grown OH contain and vacuum for the 16-18 samples could favor a strong space charge (7) which could reduce the sweeping efficiency and homogeneity.

Conclusion : Under usual (// Z) electric fields sweeping process is effective in thin AT (medium quality) cut in ambiant air and vacuum and also in thin SC (good quality) cut in air but it is questionable for thin SC cut in vacuum at 500 - 550°C. Ohmicity tests give non linear relation which could point out strong space charge effects particularly under vacuum so that the role of water vapor in ambiant air is again emphasized.

References :
(1) H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci. NS 31, 1223(1984)
(2) J.J. Martin, Trans. on Ultr. Fer. and Freq. Cont. 35, 288/1988
(3) W.P. Hanson, 38th A.F.C.S. 50-53 (1984)
(4) J. Verhoogen, Am. Mineralogist 37, 637 (1952)
(5) J.P. Bachheimer and G. Dolino 3rd EFTF Besançon, France (1989)
(6) R. Fournié "Les isolants en électrotechnique" (1986) - Eyrolles
(7) M.T. Sebastian, A. Zarka, B. Capelle J. Appl. Cryst. 21, 326 (1988)

Acknowledgements : The authors thank J.P. AUBRY (CEPE) for samples, G. DOLINO for discussions. This work is supported by a DRET contract n° 87-126

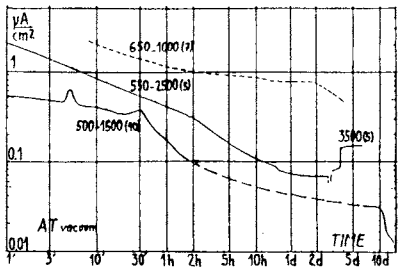


Fig. 1 : Sweeping currents versus time (AT vacuum)

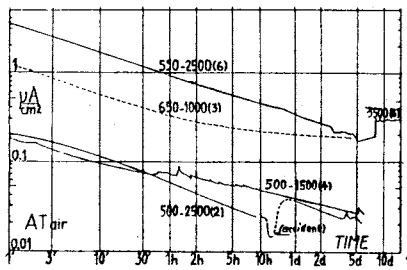


Fig. 2 : Sweeping currents versus time (AT air)

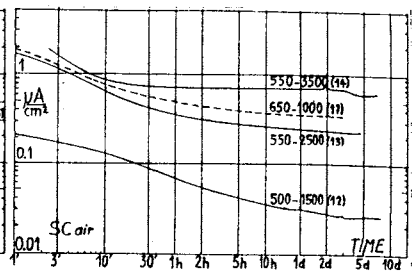


Fig. 3 : Sweeping currents versus time (SC air)

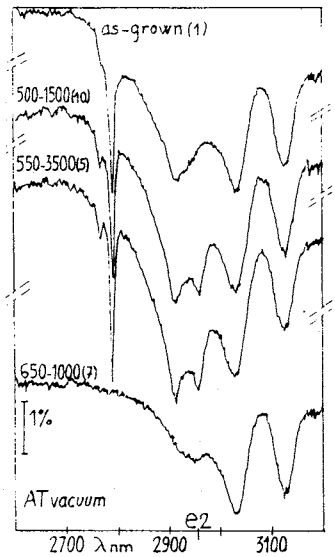


Fig. 5 : I.R (r.t) Spectra for AT cut (vacuum)

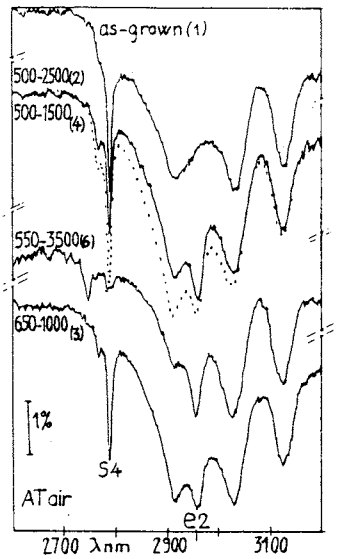


Fig. 6 : I.R. (r.t.) spectra for AT cut (air)

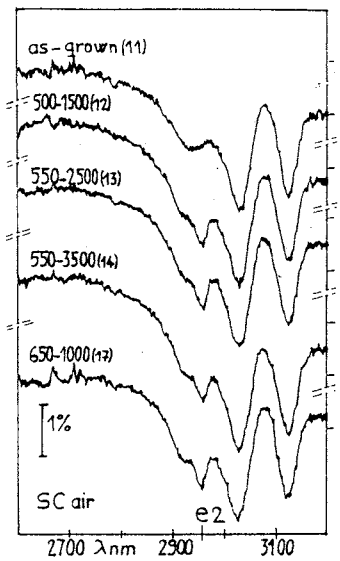


Fig. 7 : I.R. (r.t) spectra for SC cut (air)

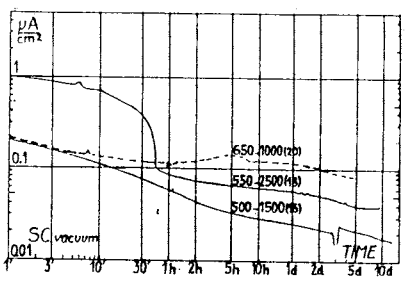


Fig. 4 : Sweeping currents versus time (SC vacuum)

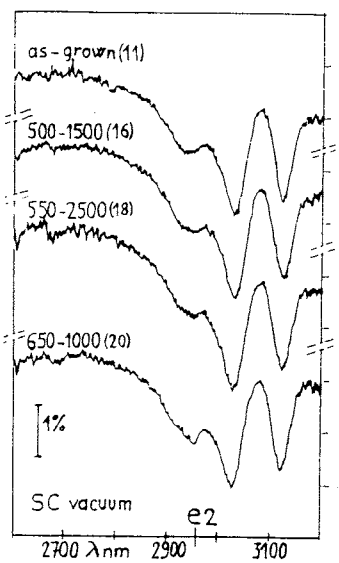


Fig. 8 : I.R. (r.t.) spectra for SC cut (vacuum)

Table II

Activation energies E after sweeping measured during the cooling (temperature interval).

* swept 9 days in addition

	Process t°C-V/cm//Z		Vacuum SW n° E(eV)	Air sweeping n° E(eV)
AT	500	1000	10 1,73(500-400)	4 1,77(500-340)
		2500		2 1,88(500-400)
	550	2500	5 1,58 (550-500)	6 2,06(550-500)
		3500	5 1,65(550-400)	6 2,01(550-430)
	650-1000		1,63(β)	1,65 (β)
			7 1,63(Tc-425)	3 1,80 (TC-450°)
SC	500-1500		16 1,24(500-360)	12 1,56(500-420)
	550	2500	18 1,43(500-350) 1,45*(550-400°)	13 1,56(550-410)
		3500		14 1,63(550-370)
	650-1000		1,78(β)	1,64 (β)
			20 2,06(Tc-550)	17 1,49(Tc-400)

Transition α-β at 573°C

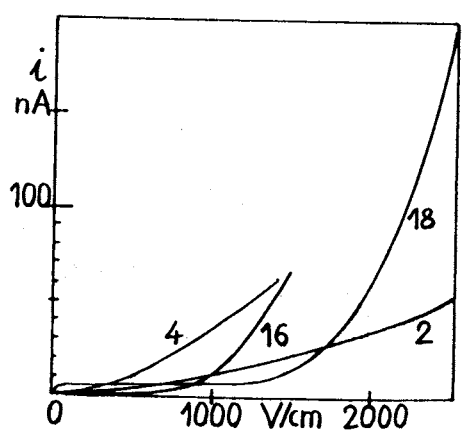


Fig. 9 : Examples of ohmicity tests after electrolysis
air : 2-4 : 500°
Vacuum : 16 : 500° ; 18 : 550°

AUTHOR INDEX

R. Agarwal 245
D.W. Allan 619
D.R. Allen 643
C.O. Alley 153
R. Ambrosini 181
K. Anemogiannis 405
R. Arnaud 577
J.P. Aubry 601
C. Audoin 133, 169, 543
D. Auguie 667

J.P. Bachheimer 723
F. Baillieu 351
S.M. Bass 451
S. Berek 405
O. Bergene 49
J.C. Bergquist 267
L.G. Bernier 195, 713
B. Bertotti 151
R.J. Besson 99, 359
G. Beutler 335
N. Beverini 381
G. Bianco 715
L. Bidart 259, 495
O. Bignon 609
J.J. Bollinger 267
H.J. Bosshard 439
M.F. Bottjer 555
R. Bourquin 359
A. Brahim-Bounab 247
H. Brandenberger 213
M. Brissaud 355
B. Brottier 609
X. Buisson 577
G. Busca 3, 157, 195, 713
O. Buzek 639

S. Calisti 395
O. Cambon 601
B. Capelle 577, 589, 675
H. Carru 589
W.E. Carter 699
St.D. Casey 205
A. Cenci 639, 715
P. Cerez 373
J. Cermak 639
A. Clairon 273
F. Cordara 639
N. Cyr 163

B. D'Albaret 259
J. Dalibard 273
J.F. Darces 667
K. Davies 327
J.A. Davis 643
H. De Boer 523
E. De Clercq 161
Ph. Defranould 395
R. Delaite 487
A. De Marchi 517
U. Denskat 231
M. Desaintfuscien 387
J. Détaint 577, 589, 601, 675
F. Deyzac 255, 349
N. Dimarcq 373
M.L. Dingus 277
J.L. Dion 355
E. Donzier 351
J. Dowsett 683
B. Dulmet 359
P. Durrant 411

G. Engel 567
L. Eyraud 355

S.J. Feltham 429
B. Fischer 523
G. Fréon 549

J.J. Gagnepain 121, 215, 419
S. Galliou 487
L. Garramone 715
J. Gary 197
S.L. Gilbert 267
V. Giordano 169
E. Girardet 495
A. Goiffon 577, 601
J. Groslambert 145, 215
S. Guellati 273
B. Guinot 307
W. Gurtner 335

A. Hamel 169
A.M. Harrison 501
Ph. Hartl 91, 221, 225, 229, 641
D. Hauden 419
J. Heighway 411
T. Heindorff 523
D.J. Heinzen 267
J. Hermann 301
M.O. Hongler 719
C.K. Hruska 295
J. Huet 495
J.P. Hurrell 555

A. Ibanez 577
M. Imae 327
W.M. Itano 267
N.E. Ivanov 25

C. Jacques 163
D. Janiaud 349
M. Jardino 387
R. Jefferies 643
W.A. Johnson 555
C. Joly 589

K. Kalliomäki 177, 219
S.K. Karuza 555
G.V. Kinal 313
D. Kirchner 631
H. Klapper 659
P. Klöckler 239
D.J. Knight 643
P. Krempl 567
A.O. Kuprijanov 25

J.P. Lacroix 249
W. Lechner 225
O. Lefort 351
R. Lefèvre 589
P.D. Lett 273
W. Lewandowski 307
J.T. Liu 157
M. Lombardi 231
M. Lopriore 13

E. Maccioni 381
D. Maeusli 231
J. Mannermaa 177, 237
C.H. Manney 267

T. Mansten 219
D.W. Martin 451
Y. Matsumoto 463
E.M. Mattison 277
J. Mc Nabb 175
P. Menary 643
H. Mérigoux 667
D.P. Morgan 411
E. Morikawa 463
P. Morley 683
M. Mourey 487
S. Muller 349

K. Nakagawa 367
H. Nau 313
B. Neubig 481

J.M. Oberto 231
A. O'Gallagher 197
S. Ohmori 463
M. Ohtsu 367

P. Palleau 723
A.F. Paradon 233
P.R. Pearce 643
L. Penavaire 395
B. Pernice 715
P. Petit 169
B.W. Petley 17
V. Pettiti 639
E. Philippot 577, 601
W. Phillips 273
M. Planat 509
F. Plumelle 387
P.A. Probst 439
L. Prost 213

C.V. Quach 719

M.G. Raizen 267
N.F. Ramsey 9
F. Raymond 395
H. Reder 221
Ch. Reigber 225, 229
P. Renoult 495
D.S. Robertson 699
S.L. Rolston 273
C. Rosetti 81
M. Rothacher 335
G.D. Rovera 161
E. Rubiola 243
J. Rutman 145

V.A. Salicshev 25
C. Salomon 273
W. Schäfer 225
E. Schanda 469
Th. Schildknecht 239, 335
W. Schlüter 187
F. Schmidt 221
R. Schröder 523
T. Schwammlein 231
J. Schwartzel 577, 589, 675
W. Schwitz 5
M.T. Sebastian 659
G. Setti 703
G. Sillioc 259
I.F. Silvera 277
R.E. Simons 451
V.R. Singh 245
R.C. Smythe 107
R.B. Spencer 683
S. Spirkovitch 351
J. Stadler 567
S. Starker 313
F. Strumia 381
B. Studer 653
J.J. Suter 695
B.R. Swabey 643
L. Sweet 197
R. Sweet 197
R.L. Sydnor 689
G. Sylos Labini 715

C.E. Tanner 273
P. Tavella 527
C.R. Tellier 247
K.H. Thiel 221, 321
P. Thomann 157
C. Thomas 307, 527
G. Théobald 373
L. Tivon Sing 543
Y. Toudic 577
R. Tourde 233
P. Tremblay 163
K. Truempy 285

P. Uhrich 233
J.P. Valentin 487
E. Valtaoja 707
J. Vanier 143
F. Vernotte 215
R.F.C. Vessot 277
J. Viennet 543
F.J. Voit 555
J. Vondrak 639

M. Wakao 463
F.L. Walls 197
R.L. Walsworth 277
R.N. Watts 273
C.S. Weimer 267
M. Weiss 327
T. Weissert 327
C.I. Westbrook 273
R.J. Williamson 501
J. Willner 313
D.J. Wineland 267

J. Yu 387

A. Zarka 577, 589, 675
L. Zhang 157
Y. Zhang 419
Y. Zheng 589, 675
W. Zingg 653